Essentials of

Genomic and Personalized Medicine

Edited by
Geoffrey S. Ginsburg
& Huntington F. Willard

Essentials of Genomic and Personalized Medicine

This page intentionally left blank

Essentials of Genomic and Personalized Medicine

Edited by

Geoffrey S. Ginsburg, M.D., Ph.D.
Center Director, Center for Genomic Medicine
Duke Institute for Genome Sciences & Policy
Professor of Medicine
Duke University Medical Center
Durham, North Carolina 27708

and

Huntington F. Willard, Ph.D.
Director
Duke Institute for Genome Sciences & Policy
Nanaline H. Duke Professor of Genome Sciences
Howard Hughes Medical Institute Professor
Duke University
Durham, North Carolina 27708

AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD • PARIS
SAN DIEGO • SAN FRANCISCO • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier

Academic Press is an imprint of Elsevier
525 B Street, Suite 1900, San Diego, CA 92101-4495, USA
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA
32 Jamestown Road, London NW1 7BY, UK

First edition 2010

Copyright © 2010 Elsevier Inc, except Chapter 54 which is in the public domain. All rights reserved

No part of this publication may be reproduced, stored in a retrieval system
or transmitted in any form or by any means electronic, mechanical, photocopying,
recording or otherwise without the prior written permission of the publisher

Permissions may be sought directly from Elsevier's Science & Technology Rights
Department in Oxford, UK: phone (+44) (0) 1865 843830; fax (+44) (0) 1865 853333;
email: permissions@elsevier.com. Alternatively visit the Science and Technology Books
website at www.elsevierdirect.com/rights for further information

Notice
No responsibility is assumed by the publisher for any injury and/or damage to persons
or property as a matter of products liability, negligence or otherwise, or from any use
or operation of any methods, products, instructions or ideas contained in the material
herein. Because of rapid advances in the medical sciences, in particular, independent
verification of diagnoses and drug dosages should be made

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

ISBN: 978-0-12-374934-5

For information on all Academic Press publications
visit our website at www.books.elsevier.com

Typeset by Macmillan Publishing Solutions
(www.macmillansolutions.com)

Printed and bound in Canada

10 11 12 13 14 15 10 9 8 7 6 5 4 3 2 1

Working together to grow
libraries in developing countries
www.elsevier.com | www.bookaid.org | www.sabre.org
ELSEVIER BOOK AID International Sabre Foundation

Table of Contents

Preface ix

Abbreviations xi

Contributors xix

1 The Foundations of Genomic and Personalized Medicine 1
Geoffrey S. Ginsburg and Huntington F. Willard

SECTION 1 BASICS

2 Organization, Variation and Expression of the Human Genome 13
Huntington F. Willard

3 DNA Sequencing for the Detection of Human Genome Variation 27
Samuel Levy and Yu-Hui Rogers

4 Genome-Wide Association Studies and Genotyping Technologies 38
Kevin V. Shianna

5 Copy Number Variation and Human Health 46
Charles Lee, Courtney Hyland, Arthur S. Lee, Shona Hislop and Chunhwa Ihm

6 DNA Methylation Analysis: Providing New Insight into Human Disease 60
Susan Cottrell, Theo de Vos, Juergen Distler, Carolina Haefliger, Ralf Lesche, Achim Plum and Matthias Schuster

7 DNA Microarrays in Biological Discovery and Patient Care 73
Andrew J. Yee and Sridhar Ramaswamy

8 Proteomics: The Deciphering of the Functional Genome 89
Li-Rong Yu, Nicolas A. Stewart and Timothy D. Veenstra

9 Comprehensive Metabolic Analysis for Understanding of Disease 97
Christopher B. Newgard, Robert D. Stevens, Brett R. Wenner, Shawn C. Burgess, Olga Ilkayeva, Michael J. Muehlbauer, A. Dean Sherry and James R. Bain

SECTION 2 INFORMATICS

10 Bioinformatic and Computational Analysis for Genomic Medicine 111
Atul J. Butte

11 Systems Biology and Systems Medicine 131
Nathan D. Price, Lucas B. Edelman, Inyoul Lee, Hyuntae Yoo, Daehee Hwang, George Carlson, David J. Galas, James R. Heath and Leroy Hood

12 Electronic Medical Records in Genomic Medicine Practice and Research 142
Glenn S. Gerhard, Robert D. Langer, David J. Carey and Walter F. Stewart

13 Online Health Information Retrieval by Consumers 151
Mark S. Boguski

SECTION 3 TRANSLATIONAL

14 Translational Genomics: From Discovery to Clinical Practice 163
Geoffrey S. Ginsburg

15 Pharmacogenetics and Pharmacogenomics 175
Iris Grossman and David B. Goldstein

16 Clinical Implementation of Translational Genomics 191
Wendy K. Chung

17 The Role of Genomics in Enabling Prospective Health Care 201
Ralph Snyderman

18 Genome Policy Considerations for Genomic Medicine 209
Susanne B. Haga

19 Federal Regulation of Genomic Medicine 223
Janet Woodcock

20 Economic Issues and Genomic Medicine 233
David L. Veenstra, Louis P. Garrison and Scott D. Ramsey

21 Public Health Genomics 245
Alison Stewart, Marta Gwinn, Ron Zimmern and Muin Khoury

SECTION 4 CLINICAL (CARDIOLOGY)

22 The Genomics of Hypertension 259
Chana Yagil and Yoram Yagil

23 Lipoprotein Disorders 269
Sekar Kathiresan and Daniel J. Rader

24 Genomics of Myocardial Infarction 289
Carlos A. Hubbard and Eric J. Topol

25 Acute Coronary Syndromes 303
L. Kristin Newby

26 Heart Failure in the Era of Genomic Medicine 313
Ivor J. Benjamin and Jeetendra Patel

27 Genomic Assessment of Cardiac Transplant Rejection 326
Michael Pham, Mario C. Deng, Jay Wohlgemuth and Thomas Quertermous

28 Genetics and Genomics of Hypertrophic Cardiomyopathy 336
J. Martijn Bos, Steve R. Ommen and Michael J. Ackerman

29 Genetics and Genomics of Arrhythmias 350
Jeffrey A. Towbin and Matteo Vatta

30 Genetics and Genomics in the Management of Hemostasis and Thrombosis 374
Richard C. Becker and Felicita Andreotti

31 Genomics of Congenital Heart Disease 390
Jessie H. Conta and Roger E. Breitbart

SECTION 5 CLINICAL (ONCOLOGY)

32 Genomics in the Management of Lymphomas 407
Lisa Rimsza

33 Genomics in Leukemias 421
Lars Bullinger, Hartmut Dohner and Jonathan R. Pollack

34 Genomics in the Diagnosis and Management of Lung Cancer 432
Hasmeena Kathuria, Avrum Spira and Jerome Brody

35 Genomics in the Diagnosis and Management of Breast Cancer 446
Erich S. Huang and Andrew T. Huang

36 Colorectal Cancer 457
G.L. Wiesner, T.P. Slavin and J.S. Barnholtz-Sloan

37 Genomic Evaluation and Management of Prostate Cancer 477
Phillip G. Febbo and Philip W. Kantoff

38 Genomic Assessment of Ovarian Cancer 493
Tanja Pejovic, Matthew L. Anderson and Kunle Odunsi

39 Genomic Evaluation of Pancreatic Neoplasms 501
Asif Khalid and Kevin McGrath

40 Genomic Evaluation of Head and Neck Cancer 511
Giovana R. Thomas and Yelizaveta Shnayder

41 Genomic Evaluation of Brain Tumors and Gliomas 522
Sean E. Lawler and E. Antonio Chiocca

42 Targeted Therapies for Cancer 532
Jeffrey S. Ross

SECTION 6 CLINICAL (INFLAMMATORY DISEASE)

43 Genomics in the Evaluation and Management of Rheumatoid Arthritis 551
Robert M. Plenge and Michael E. Weinblatt

44 Genomic Evaluation of Multiple Sclerosis 566
Francisco J. Quintana and Howard L. Weiner

45 Genomic Assessment of Inflammatory Bowel Disease 575
Ad A. van Bodegraven and Cisca Wijmenga

46 Asthma Genomics 590
Scott T. Weiss, Benjamin A. Raby and Juan C. Celedón

47 Genomics in the Evaluation and Management of Chronic Obstructive Pulmonary Disease 603
Peter J. Barnes

48 Genetics and Genomics of Interstitial Lung Disease 616
Paul W. Noble and Mark P. Steele

49 Peptic Ulcer Disease 627
John Holton

SECTION 7 CLINICAL (METABOLIC DISEASE)

50 Genomics in Pathogenesis of Cirrhosis 645
N.A. Shackel, K. Patel and J. McHutchison

51 Genomic Medicine and Obesity 661
J. Alfredo Martínez

52 Diabetes 676
Maggie Ng and Nancy J. Cox

SECTION 8 CLINICAL (NEUROLOGICAL DISEASE)

53 Genetics and Genomics of Dementia 687
Robert L. Nussbaum

54 Genetics and Genomics of Parkinson's Disease 700
Shushant Jain and Andrew B. Singleton

55 Genomic Considerations in Ophthalmology 712
Janey L. Wiggs

56 Genomics in the Diagnosis and Management of Depression 722
Brigitta Bondy

SECTION 9 CLINICAL (INFECTIOUS DISEASE)

57 Genomic Approaches to the Host Response to Pathogens 733
M. Frances Shannon

58 Host Genomics and Bacterial Infections 744
Melissa D. Johnson and Mihai Netea

59 Genomics in the Evaluation and Management of Sepsis 760
Christopher W. Woods, Robert J. Feezor and Stephen F. Kingsmore

60 Genomics and the Management of Hepatitis 774
N.A. Shackel, K. Patel and J. McHutchison

Index 787

This page intentionally left blank

Preface

With the completion of the Human Genome Project at the beginning of this decade and the rapid development and application of new advances in our ability to understand and query the human genome and its gene set, it is time already to review the early sketches of a much-anticipated transformation of the practice of medicine. We have seen the very first signs of a fundamental shift in how we behold human physiology and pathology, how we view the concept of what is "normal", how we consider individuals and their prospects for lifelong health, and how we design health care systems that are equally adaptable to the demands of population-wide epidemics and the opportunities for personalized care that utilizes genome-based information to consider individual susceptibility to disease and therapeutic options.

Genome-based data, information, knowledge, and eventually wisdom will make possible the kind of health care that has been dreamed of since the advent of disease-based medicine early in the 20th century. A system of health care that harnesses the data intensity of the genome and its derivatives, along with imaging, clinical, and environmental information, will empower physicians and other health care providers to do what they have always aspired to do – make medical care as individualized as possible. But this newfound information and knowledge will also allow each of us as consumers of health care to take more control of our futures and to develop a more strategic and prospective approach to health. We stand at the dawn of a profound change in science and medicine's predictive nature and in our understanding of the biological underpinnings of health and disease. Even in this early light, we can see the outlines of a coming ability to:

- predict individual susceptibility to disease, based on genetic, genomic, and other factors;
- provide more useful tools and individualized programs for disease prevention, based on knowledge of one's susceptibility;
- detect the onset of disease earlier and before it is clinically evident, based on newly discovered biological markers that arise from changes at the molecular level;
- preempt disease progression, as a result of early detection; and
- target medicines and their dose more precisely and safely to each patient, on the basis of a deep understanding of disease mechanism and the role that genetic and genomic factors play in the individual response to drugs.

The time for this revolution in genomic and personalized medicine has come. As its name suggests, this book is intended to lay out the essentials of this new approach to medicine and to emphasize for particular conditions the clinical opportunities that present themselves today. It is designed to complement the more comprehensive two-volume book *Genomic and Personalized Medicine*, published last year, which presented in substantial depth the foundations of this new science, outlined the early opportunities for the practice of medicine to incorporate genome-based analysis into health care, and anticipated the many conditions to which genomic and personalized medicine will apply in the years ahead.

This volume on *Essentials of Genomic and Personalized Medicine* describes an emerging field that spans the breadth of clinical medicine, with many challenges for society at large and for health care systems in particular. Nonetheless, we are optimistic that the appropriate delivery models and economic incentives will be developed in a trustworthy framework that will be embraced by societies around the globe.

Our intended audience is clinically oriented but broad, ranging from medical students to residents and fellows to practitioners in any of the health care professions – physicians in any of the medical specialties, surgeons, nurses, genetic (and genomic) counselors, and laboratory directors.

In times of transformation, we are all students. We hope that this *Essentials* volume will help usher in this new era of genomic and personalized medicine and will provide a useful and thorough introduction to the science and practice of this new approach to human health.

Geoffrey S. Ginsburg, M.D., Ph.D.
Huntington F. Willard, Ph.D.
June 2009

This page intentionally left blank

Abbreviations

2D-DIGE	Two-dimensional difference gel electrophoresis
2D-PAGE	Two-dimensional polyacrylamide gel electrophoresis
5-HT	5-Hydroxytriptamine or serotonin
5-HTT	5-HT transporter
5-HTTLPR	5-HTT gene-linked polymorphic region
6-MMP	6-Methylmercaptopurine
6-MP	6-Mercaptopurine
6-MTIMP	6-Methyl-thioinosinemonophosphate
6-TGN	6-Thioguanine nucleosides
6-TIMP	6-Thioinosinemonophosphate
AAU	Acute anterior uveitis
ABCA1	Adenosine triphosphate-binding cassette protein A1
ABCB1	Adenosine triphosphate-binding cassette, sub-family B
ABI	Ankle brachial index
ACC	American College of Cardiology
ACE	Angiotensin converting enzyme
aCGH	Array comparative genomic hybridization
ACR	Acute cellular rejection
AD	Alzheimer's disease
AD	Aortic dilation
ADH	Autosomal dominant hypercholesterolemia
ADHD	Attention deficit hyperactivity disorder
ADIGE	Autosomal dominant idiopathic generalized epilepsy
ADJME	Autosomal dominant juvenile myoclonic epilepsy
ADNFLE	Autosomal dominant nocturnal frontal lobe epilepsy
ADP	Adenosine 5-diphosphate
ADR	Adverse drug reaction
AED	Antiepileptic drug
Ago2	Argonaute 2 mRNA endonuclease
AHA	American Heart Association
AIDS	Acquired immune deficiency syndrome
AIMs	Ancestry informative markers
AJCC	American Joint Committee on Cancer
ALL	Acute lymphoblastic leukemia
ALOX5AP	Arachidonate 5-lipooxygenase-activating protein
AMACR	Alpha methylacyl CoA reductase
AML	Acute myeloid leukemia
AMR	Antibody mediated rejection
AOO	Age of onset
Apo(a)	Apolipoprotein(a)
ApoE	Apolipoprotein E
APOE	Apolipoprotein E (gene name)
AR	Androgen receptor
AR	Autosomal recessive
ARACNe	Algorithm for the Reconstruction of Accurate Cellular Networks
ARE	Androgen response element
ARH	Autosomal recessive hypercholesterolemia
ARJP	Autosomal recessive juvenile Parkinsonism
ARVC	Arrhythmogenic right ventricular cardiomyopathy
AS	Ankylosing spondylitis
ASCA	Anti-*Saccharomyces cerevisiae* antibodies
ASCO	American Society for Clinical Oncology
ASCVD	Atherosclerotic cardiovascular disease
ASD	Atrial septal defect
ASO	Anti-sense oligonucleotides
ASPE	Allele-specific primer extension
ATM	Ataxia telangiectasia mutated
ATP	Adenosine triphosphate

ATP Adult Treatment Panel
ATR Ataxia-telangiectasia and RAD3-related
AZT Azidothymidine

B1-AR β1-adrenergic receptor
BBB Blood–brain barrier
BCAA Branched-chain amino acids
Bcl B-cell lymphoma protein
BCL-2 B-cell lymphoma 2
Bcl-Abl Breakpoint cluster region – Abelson kinase fusion protein
BDNF Brain-derived neurotrophic factor
BER Base excision repair
BEST β-Blocker Evaluation of Survival Trial
BH Bcl-2 homology domain
BLAST Basic local alignment and search tool
BLAT BLAST-like alignment tool
BMC Bone marrow cell
BMI Body mass index
BMP-7 Bone morphogenic protein-7
BOLD Blood oxygen level-dependent
BPDE Benzo (a) pyrene diol epoxide
BPH Benign prostatic hypertrophy
BRCA Breast cancer gene
BRCA1 Breast cancer 1, early onset, protein
BRCA2 Breast cancer type 2 susceptibility protein
BRD2 Bromodomain containing 2 (gene)
BSML Bioinformatic sequence markup language

CA Capsid (viral protein)
CAD Coronary artery disease
CADASIL Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy
CAE Childhood absence epilepsy
CALGB Cancer and leukemia group B
CAP College of American Pathologists
CARD Caspase recruitment domain (gene)
CARGO Cardiac Allograft Rejection Gene Expression Observational study
CAV Cardiac allograft vasculopathy
CBD Cortical basal degeneration
CBP CREB-binding protein
CBZ Carbamazepine
CCD Charge coupled device
CCL Chemokine (C-C motif) ligand
CCND1 Cyclin D1, a proto-oncogene
CD Cluster of differentiation
CD Crohn's disease
CDAI Crohn's disease activity index
CDC Centers for Disease Control
cDNA Complementary deoxyribonucleic acid
CDS Clinical decision support
CED-4 Cell death abnormality gene 4
CETP Cholesteryl ester transfer protein
CF Cystic fibrosis
CFTR Cystic fibrosis transmembrane conductance regulator
CGAP Cancer Genome Anatomy Project
CGH Comparative genomic hybridization
CGHa Comparative genomic hybridization array
CGP Cancer Genome Project
CHD Congenital heart disease
CHD Coronary heart disease
CHF Congestive heart failure
CHMP Committee for Medicinal Products for Human Use
CHRNA4 Cholinergic receptor, nicotinic, alpha 4 (gene)
CIBEX Center for Information Biology Gene Expression
CIC Citrate/isocitrate carrier
CICR Calcium-induced calcium release
CKD Chronic kidney disease
CLI Critical limb ischemia
CLIA Clinical Laboratory Improvement Amendments
CML Chronic myelogenous leukemia
CMT Charcot-Marie-Tooth disease
CMV Cytomegalovirus or human herpes virus 5
CNS Central nervous system
CNV Copy number variant
COPA Cancer outlier profile analysis
COR C-terminal of ROC domain
COX Cyclooxygenase
COX-2 Cyclooxygenase-2
CPOE Computerized provider order entry
CRC Colorectal cancer
CREB Cyclic AMP response binding protein
CRF Circulating recombined form
CRP C-reactive protein
CS Cell signaling factor
CSA Comparative sequence analysis
CSF Cerebrospinal fluid
CSF1 Colony stimulating factor 1
CTC Circulating tumor cells
CTGF Connective tissue growth factor
CTL Cytotoxic T lymphocytes
CTLA Connective tissue late antigen
CTSA Clinical and Translational Science Awards
CX32 Connexin-32
CX37 Connexin-37
CYP Cytochrome P450
CYP2C19 Cytochrome P450 2C19
CYP2D6 Cytochrome P450 2D6

D2R Dopamine-2 receptor
DAI Disease Activity Score (for Ulcerative Colitis)
dbSNP NCBI SNP database
DCM Dilated cardiomyopathy
DGGE Denaturing gradient gel electrophoresis
DGI Diabetes Genetics/Initiative
DHF Diastolic heart failure

DHLPC	Denaturing high performance liquid chromatography
DLB	Diffuse Lewy body
DLBD	Diffuse Lewy body disease
DLDA	Diagonalized linear discriminant analysis
DLG	Disc large homologue (gene)
DMEs	Drug-metabolizing enzymes
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DNMTs	DNA methyltransferases
DOPAC	Dihydroxyphenylacetic acid
ds	Double stranded
dsRNA	Double-stranded ribonucleic acid
DTC	Direct-to-consumer
DU	Duodenal ulcer
DZ	Dizygotic
EAE	Experimental autoimmune encephalomyelitis
EBI	European Bioinformatics Institute
EBV	Epstein–Barr virus
ECL	Eenterochromaffin-like cells
ECM	Extracellular matrix
EDG-1	Endothelial differentiation gene-1
EGAPP	Evaluation of Genomic Applications in Practice and Prevention
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
EHR	Electronic health record
EIA	Enzyme immunoassay
EL	Endothelial lipase
ELISA	Enzyme-linked immunosorbent assay
ELMO1	Engulfment and cell motility 1
EM	Extracellular matrix protein
EMB	Endomyocardial biopsy
EMBL-EBI	European Molecular Biology Laboratory–European Bioinformatics Institute
EMEA	European Medicines Evaluation Agency
ENCODE	Encyclopedia of DNA Elements
EOAD	Early-onset Alzheimer's disease
EPIYA	Single letter codes for amino acids
eQTL	Expression quantitative trait loci
ERK	Extracellular signal-regulated kinase
ESR	Erythrocyte sedimentation rate
ESRD	End-stage renal disease
EST	Expressed sequence tag
FA	Fanconi anemia
FACS	Fluorescent-activated cell sorting
FANCA	Fanconi anemia proteins A
FANCD2	Fanconi anemia protein D2
FAP	Familial adenomatous polyposis
FCHL	Familial combined hyperlipidemia
fCJD	Familial Creutzfeldt–Jakob disease
FCS	Familial hyperchylomicronemia syndrome
FD	Familial dysbetalipoproteinemia
FDA	Food and Drug Administration (USA)
FDB	Familial defective apoB-100
FFPE	Formalin-fixed, paraffin-embedded
FGF	Fibroblast growth factor
FGF-2	Fibroblast growth factor-2
FH	Familial hypercholesterolemia
FHTG	Familial hypertriglyceridemia
FISH	Fluorescence *in situ* hybridization
FKBP	FK506-binding protein
FL	Follicular lymphoma
FLT-3	FMS-like tyrosine kinase-3
fMRI	Functional MRI
FNA	Fine needle aspirate
FOV	Field of view
FSGS	Focal segmental glomerulosclerosis
FTD	Frontotemporal dementia
FUSION	Finland–United States Investigation of NIDDM genetics
GABBR1	GABA B receptor 1 (gene)
GABRG2	GABA A receptor, gamma 2 (gene)
GAIN	Genetic Association Information Network
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GBM	Glomerular basement membrane
GBS	Guillain–Barre Syndrome
GCKR	Glucokinase regulatory protein
GC-MS	Gas chromatography–mass spectrometry
GCSF	Granulocyte colony-stimulating factor
GDNF	Glial cell line-derived neurotrophic factor
GEFS+	Generalized epilepsy with febrile seizures plus
GEO	Gene Expression Omnibus
GEP	Gene expression profiling
GFAP	Glial fibrillary acidic protein
GFP	Green fluorescent protein
GLEPP1	Glomerular epithelial protein 1
GLP	Good laboratory practice
GM-CSF	Granulocye macrophage–colony stimulating factor
GO	Gene ontology
GPI	Globus pallidus interna
GR	Glucocorticoid receptor gene
GSEA	Gene set enrichment analysis
GSIS	Glucose-stimulated insulin secretion
GSTT	Glutathione *S*-transferase
GU	Gastric ulcer
GWAS	Genome-wide association study
GZM	Granzyme
H	Heterozygosity
HAART	Highly active anti-retroviral therapy
HAV	Hepatitis A
HBV	Hepatitis B
HCC	Hepatocellular carcinoma
HCM	Hypertrophic cardiomyopathy
HCV	Hepatitis C
HDAC	Histone deacetylase

HDL	High-density lipoproteins
HDM2	Human double minute 2
HDV	Hepatitis D
HeLa	Cervical cancer cell line
HEPS	Highly exposed persistently seronegative individuals
HER2	Human epidermal growth factor receptor 2
HEV	Hepatitis E
HF	Heart failure
HFPEF	Heart failure with preserved ejection fraction
HGF	Hepatocyte growth factor
HGPRT	Hypoxanthine phosphoribosyl transferase
HHS	Health and Human Services (U.S.)
HHV-6	Human herpes virus 6
HITSP	Healthcare Information Technology Standards Panel
HIV	Human immunodeficiency virus
HKMT	Histone lysine methyltransferases
HL	Hepatic lipase
HL7	Health Level 7
HLA/*HLA*	Human leukocyte antigen/(gene)
HLI	Hind limb ischemia
HMT	Histone methyltransferase
HNPCC	Hereditary nonpolyposis colorectal cancer
HNSCC	Head and neck squamous cell carcinoma
HP-1	Heterochromatin-associated protein-1
HPV	Human papilloma virus
HPV-16	Human papilloma virus-16
HR	Homologous recombination
HRPC	Hormone refractory prostate cancer
HS	Hyper-spectral
HSC	Hepatic stellate cell
HSP	Heat shock protein
HSV-1	Human herpes simplex virus type 1
HSV1-TK	Herpes simplex virus type 1-thymidine kinase
HSV-2	Human herpes simplex virus type 2
hTERT	Human telomerase reverse transcriptase
HTLV-I	Human T-cell leukemia/lymphoma virus type I
HTLV-III	Human T-cells lymphotropic virus type 3
HTS	High-throughput screening
HVJ	Hemagglutinating virus of Japan
IBD	Inflammatory bowel disease
IBDQ	Inflammatory bowel disease questionnaire
IC	Intermittent claudication
IC	Interstitial cystitis
ICAM	Intracellular adhesion molecules
ICD	Implantable cardioverter defibrillator
ICD	International Classification of Diseases
ICDc	Cytosolic, NADP-dependent isocitrate dehydrogenase
iDC	Immature dendritic cells
IDCM	Idiopathic dilated cardiomyopathy
IDDM	Insulin-dependent diabetes mellitus
IDL	Intermediate-density lipoproteins
IDO	Indolemaine 2,3-dioxygenase
IFN	Interferon
IFN-β	Interferon-beta
IFN-γ	Interferon-gamma
Ig	Immunoglobulin
IgA	Immunoglobulin-A
IGF-I	Insulin-like growth factor-I
IHC	Immunohistochemistry
IHE	Integrating the Healthcare Enterprise
IKBL	Inhibitor of κB-like (gene)
IL	Interleukin
IL (4/13)	Interleukin (4/13)
IL-1β	Interleukin-1 beta (gene)
IL-1ra	Interleukin-1 receptor antagonist
IL-6	Interleukin-6
IL-10	Interleukin-10
IMAGE	Invasive Monitoring Attenuation through Gene Expression study
iMC	Immature myeloid cells
IN	Integrase (viral protein)
IOM	Institute of Medicine (U.S.)
IS	Insertion sequence
ISB	Institute for Systems Biology
ISG	Intreferon stimulated gene
ISHLT	International Society for Heart and Lung Transplantation
IT	Information technology
IVDMIA	*In vitro* diagnostic multivariate index assay
JCA	Juvenile chronic arthritis
JHDM1	JmjC domain-containing histone demethylase-1
JME	Juvenile myoclonic epilepsy
Kb	Kilobase
KEGG	Kyoto Encyclopedia of Genes and Genomes
LARGO	Lung Allograft Rejection Gene expression Observational study
LAV	Lymphadenopathy-associated virus
LB	Lewy body
LCAT	Lecithin cholesterol acyltransferase
LCM	Laser capture microdissection
LD	Linkage disequilibrium
LDL	Low-density lipoproteins
l-DOPA	Levodopa
LOAD	Late-onset Alzheimer's disease
LOD	Logarithm of the odds
LOH	Loss of heterozygosity
LOINC	Logical Observation Identifiers, Names, and Codes
LOR	Line of response
Lp(a)	Lipoprotein(a)
LPA	Lysophosphatidic acid
LPL	Lipoprotein lipase
LPS	Lipopolysaccharide
LRR	Leucine-rich repeat

LRRK2	Leucine rich repeat kinase 2
LSD1	Lysine-specific demethylase-1
LTA	Lymphotoxin alpha
LTA4H	Leukotriene A4 hydrolase
LTB4	Leukotriene B4
LTNP	Long-term non-progression
LV	Left ventricle
LVAD	Left ventricular assist device
LVEDP	Left ventricular end-diastolic pressure
LVEF	Left ventricular ejection fraction
LVH	Left ventricular hypertrophy
LVOT(O)	Left ventricular outflow tract (obstruction)
MA	Matrix (viral protein)
mAb	Monoclonal antibodies
MADH6	Mothers against decapentaplegic homolog 6
MAGE-ML	MicroArray and Gene Expression Markup Language
MAGUK	Membrane-associated guanylate kinase
MAM	Metastasis-averse microenvironment
MAO	Monoamine oxidase
MAOA	Monoamine oxidate A
MAP	Mitogen-activated protein
MAPK	Microtubule-associated protein kinase
MAPK	Mitogen-activated protein kinase
MAQC	Microarray quality control
Mb	Megabase
MBD	Methylated DNA binding domain protein
MCD	Malonyl CoA decarboxylase
MCD	Minimal change disease
MCP	Monocyte/macrophage chemoattractant protein
MCP-1	Monocyte chemoattractant protein 1
MCS	Multi-species conserved sequence
M-CSF	Macrophage colony-stimulating factor
MDM2	Mouse double minute 2
MDP	Muramyl dipeptide
MDR	Multidrug resistance (gene)
MedDRA	Medical Dictionary for Regulatory Activities Terminology
MEF2A	Myocyte-enhancing factor 2A
Megsin	Mesangial cell-specific gene with homology to serpin
MEN	Multiple endocrine neoplasm
MeSH	Medical Subject Headings
MeV	Multiexperiment viewer
MGED	Microarray Gene Expression Data society
MGMT	O6-Methylguanine-DNA methyltransferase
MHC	Major histocompatability complex
MI	Myocardial infarction
MIAME	Minimal Information About a Microarray Experiment
MIM	Metastasis-inclined microenvironment
miR-192	MicroRNA-192
miRNA	Micro-ribonucleic acid
MITF	Microphthalmia transcription factor
MLEM	Maximum likelihood expectation maximization
MLPA	Multiplex ligation-dependent probe amplification
MMP	Matrix metalloproteinase
MMP-3	Matrix metalloproteinase-3
MMR	Mismatch repair
MPSS	Massively parallel signature sequencing
MPTP	1-Methyl-4-phenyl-1, 2, 3, 6-tetrahydropyridine
mQTL	Metabolite quantitative trait loci
mRNA	Messenger ribonucleic acid
mRSS	Modified Rodnan Skin Score
MS	Multiple sclerosis
MS	Multispectral
MS/MS	Tandem mass spectrometry
MSA	Multiple system atrophy
MTDNA	Mitochondrial DNA
MTHFR	Methylenetetrahydrofolate reductase (gene)
mTOR	Mammalian target of rapamycin
MTP	Microsomal triglyceride transfer protein
MTX	Methotrexate
MudPIT	Multidimensional protein identification technology
MVP	Mitral valve prolapse
MZ	Monozygotic
NAAT	Nucleic acid amplification testing
NAD(P)H	Nicotinamide adenine dinucleotide phosphate reduced form
NAFLD	Non-alcoholic fatty liver disease
NASH	Non-alcoholic steatohepatitis
NAT2	*N*-Acetyltransferase 2
NAWM	Normal appearing white matter
NC	Nucleocapsid (viral protein)
N-CAM	Neural cell adhesion molecule
NCBI	National Center for Biotechnology Information
NCE	New chemical entity
NCEP	National Cholesterol Education Program
NCI60	Refers to a set of 60 cancer cell lines curated at the NCI (National Cancer Institute)
ncRNA	Non-coding RNA
NEF	NEgative regulatory Factor (viral protein)
NER	Nucleotide excision repair
NF1/2	Neurofibromatosis1/2
NHGRI	National Human Genome Research Institute
NHR	Non-homologous recombination
NIEHS	National Institute of Environmental Health Sciences (NIH)
NIH	National Institutes of Health
NK	Natural killer
NMR	Nuclear magnetic resonance
NNRTI	Non-nucleosidic reverse transcriptase inhibitor
NOD	Nucleotide oligomerization domain (gene)
NPV	Negative predictive value

NQO1	Nicotinamide adenine dinucleotide phosphate oxidoreductase
NRTI	Nucleosidic reverse transcriptase inhibitor
NSAID	Non-steroidal anti-inflammatory drug
nt	Nucleotides
NUD	Non-ulcer dyspepsia
NYHA	New York Heart Association
OBO	Open Biomedical Ontologies
OCTN	Organic cation transporter (gene)
OHCA	Out of hospital cardiac arrest
OMIM	Online Mendelian Inheritance in Man
OMP	Outer membrane protein
OmpC	Outer membrane porin C of E. coli
opn	Osteopontin (gene)
OPN	Osteopontin (protein)
OR	odds ratio
ORESTES	Open reading frame expressed sequence tags
ORF	Open reading frame
OSEM	Ordered subset expectation maximization
p53BP1	p53 Binding protein 1
PAD	Peripheral arterial disease
PAGE	Polyacrylamide gel electrophoresis
PAI	Pathogenicity island
PAM	Prediction analysis of microarrays
pANCA	Perinuclear antineutrophil cytoplasmatic antibodies
PAOD1	Peripheral arterial occlusive disease
Pap	Papanicolaou
Parvo B19	Human parvovirus B-19
PBC	Primary biliary cirrhosis
PBL	Peripheral blood lymphocytes
PBMC	Peripheral blood mononuclear cell
PC	Pyruvate carboxylase
PCA	Principal components analysis
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PCSK9	Proprotein convertase subtilisin/kexin type 9
PCV	Procarbazine/CCNU/vincristine
PD	Parkinson disease
PD	Pharmacodynamic
PDA	Patent ductus arteriosus
PDA	Personal digital assistant
PDD	Parkinson disease with dementia
PDGF	Platelet-derived growth factor
PDGFR	Platelet-derived growth factor receptor
PDH	Pyruvate dehydrogenase
PDTC	Pyrrolidine dithiocarbamate
PDYN	prodynorphin (gene)
PFGE	Pulse field gel electrophoresis
PGE2	Prostaglandin E2
PGx	Pharmacogenetics and pharmacogenomics
PHT	Phenytoin
PI	Protease inhibitor
PI3K	Phosphotidylinositol triphosphate kinase
PIN	Prostatic intraepithelial neoplasia
PK	Pharmacokinetics
PLTP	Phospholipid transfer protein
PMAGE	Polony Multiplex Analysis of Gene Expression
PMBC	Peripheral blood mononuclear cells
PML-RAR	Pro-myelocytic leukemia retinoic acid receptor oncoprotein
PON1	Paraoxonase
PPI	Percutaneous peripheral intervention
PPV	Positive predictive value
PR	Protease (viral)
PrEC	Prostate epithelial cells
PRIDE	Proteomics Identifications Database
PRNP	Prion protein (gene)
PS	Pulmonary stenosis
PSA	Prostate-specific antigen
PsA	Psoriatic arthritis
PSC	Primary sclerosing cholangitis
PSP	Progressive supranuclear palsy
PTEN	Phosphatase and tensin homolog
PTM	Pre-trans-splicing molecule
PUD	Peptic ulcer disease
PZ	Plasticity zone
QPCR	Quantitative PCR
qRT	Quantitative real-time PCR
qRT-PCR	Quantitative reverse-transcriptase polymerase chain reaction
QTL	Quantitative trait locus
QT-PCR	Quantitative real-time polymerase chain reaction
QV	Quality value
R&D	Research and Development
RA	Rheumatoid arthritis
RAAS	Renin-angiotensin-aldosterone system
RAPD-PCR	Random amplified polymorphic deoxyribonucleic acid–polymerase chain reaction
RB	Retinoblastoma
RB	Retinoblastoma (gene)
RCM	Restrictive cardiomyopathy
REPAIR-AMI	Reinfusion of enriched progenitor cells and infarct remodeling in acute myocardial infarction
RET	Ret proto-oncogene
REV	REgulator of Virion protein expression (viral protein)
RFLP	Restriction fragment length polymorphism
RISC	RNA interference silencing complex
RNA	Ribonucleic acid
RNAi	Ribonucleic acid interference
ROC	RAS in complex proteins
ROCO	ROC (RAS of complex proteins)/COR (C-terminal of ROC)

ROS	Reactive oxygen species
RP	Rapid progression
RPA	Ribonuclease protection assay
RSV	Respiratory syncytial virus
RT	Reverse transcriptase (viral protein)
RT-PCR	Real-time polymerase chain reaction
RT-PCR	Reverse transcription – polymerase chain reaction
RV	Right ventricle
RVHF	Right ventricular heart failure
SACGHS	Secretary's Advisory Committee on Genetics, Health and Society (U.S.)
SAGE	Serial Analysis of Gene Expression
SALSA	Scoring algorithm for spectral analysis
SAM	*S*-Adenosyl-l-methionine (AdoMet)
SAM	Significance analysis of microarrays
SAM	Systolic anterior motion of mitral valve
SAP/Exo I	Shrimp alkaline phosphatase/exonuclease I
SARS	Severe acute respiratory syndrome
SBE	Single-base extension
SCD	Sudden cardiac death
SCID	Severe combined immunodeficiency
SCN1A	Voltage-gated sodium channel, type 1, alpha subunit (gene)
SCOTS	Selective capture of transcribed sequences
SCX	Strong cation exchange
SD	Standard deviation
SDO	Standards development organization
SE	Shared epitope of HLA (human leukocyte antigen) genes
SE	Spongiform encephalopathies
SEBASTIAN	System for Evidence-Based Advice through Simultaneous Transaction with an Intelligent Agent across a Network
SEC	Sinusoidal endothelial cell
SERPA	Serological proteome analysis
SHF	Systolic heart failure
shRNA	Short hairpin ribonucleic acid
SIP1	Smad-interacting protein 1
siRNA	Small interfering ribonucleic acid
SIV	Simian immunodeficiency virus
SJS/TEN	Stevens–Johnson syndrome/toxic epidermal necrolysis
SLC12A3	Solute carrier family 12 member 3
SLE	Systemic lupus erythematosus
SLN	Sentinel lymph node
SMD	Stanford Microarray Database
SN	Substantia nigra
SNOMED CT	Systematized Nomenclature of Medicine, Clinical Terms
SNP	Single nucleotide polymorphism
SNpc	Substantia nigra pars compacta
SNR	Substantia nigra pars reticulata
SOX10	SRY-box containing gene 10
SpA	Spondyloarthropathy
SPARC	Secreted protein, acidic, and rich in cysteine, osteonectin
SPMS	Secondary progressive MS
SR-BI	Scavenger receptor, class BI
ss	Single-stranded
SSc	Systemic sclerosis
SSCP	Single strand conformation polymorphism
SSRP	Simple sequence repeat polymorphism
SSTR2	Somatostatin receptor type 2
STAT	Signal transducers and activators of transcription
STEMI	ST segment Elevation Myocardial Infarction
STK11 gene	Serine/threonine kinase 11
STN	Subthalmic nucleus
SU	Surface (viral protein)
SVAS	Supravalvar aortic stenosis
SVR	Sustained viral response
SWOG	Southwest Oncology Group
TAM	Tumor associated macrophages
TAT	TransActivator of Transcription (viral protein)
TBB	Transcriptomic-based biomarkers
TCF7L2	Transcription factor 7-like 2
TCR	T-cell receptor
TDM	Therapeutic drug monitoring
TF	Transcription factor
TFAP2B	Transcription factor activating protein 2β
TGF	Transforming growth factor
TGF-β	Transforming growth factor beta
TH	Tyrosine hydroxylase
Th1	T-helper cell, type 1
Th2	T-helper cell, type 2
TIGR	The Institute for Genomic Research
TIMP	Tissue inhibitor of metalloproteinase
TIMP-1	Tissue inhibitor of metalloproteinase-1
TK	Tyrosine kinase
TLE	Temporal lobe epilepsy
TLR	Toll-like receptor
TLS	Trans-lesion DNA synthesis
TM	Trans-Membrane (viral protein)
TNF	Tumor necrosis factor
TNF-α	Tumor necrosis factor-alpha
TNF-β	Tumor necrosis factor-beta
TNM	Tumor-node-metastasis stage
TOF	Tetralogy of Fallot
TPMT/*TPMT*	Thiopurine methyl transferase/(gene)
Treg	Regulatory T-cells
TSA	Trichostatin A
TSG	Tumor suppressor gene
TSP	Thrombospondin
UADT	Upper aerodigestive tract
UC	Ulcerative colitis

UCSC	University of California at Santa Cruz
UMLS	Unified Medical Language System
UNAIDS	United Nation mission on AIDS
US	Ultrasound
US	United States of America
USF1	Upstream transcription factor 1
USpA	Undifferentiated spondyloarthropathy
UTR	Untranslated region (three prime or five prime)
V-ATPase	Vacuolar adenosine triphosphatase
V-CAM	Vascular cell adhesion molecule
VEGF	Vascular endothelial growth factor
VF	Ventricular fibrillation
VIF	Virion infectivity factor
VLDL	Very low-density lipoproteins
VPR	Viral protein R
VPU	Viral protein U
VS	Ventricular septal defect
VT	Ventricular tachycardia
VZV	Varicella zoster virus
WGA	Whole-genome amplification
WGAS	Whole Genome Association Studies
WGS	Whole genome shotgun
WHO AIDS	World Health Organization mission on AIDS
WHO	World Health Organization
WNV	West Nile virus
WSS	Western-specific sequence
WTCCC	Welcome Trust Case Control Consortium
XMRV	Xenotropic murine leukemia virus-related virus
ZDV	Zidovudine
ZFP	Zinc finger protein

Contributors

Michael J. Ackerman Mayo Clinic, Windland Smith Rice Sudden Death Genomics Laboratory, 200 First Street, SW Rochester, MN 55905.

Peter J. Barnes National Heart & Lung Institute, Section of Airway Disease, Dovehouse Street, Scadding Building, London, SW3 6LY, UK.

Richard Becker Duke University Medical Center, Department of Medicine – Cardiovascular Thrombosis Center, 2400 Pratt Street, Durham, NC 27705.

Ivor J. Benjamin University of Utah Health Sciences Center, Div. Cardiology, Center for Cardiovascular Translational Biomedicine, 500 North Medical Drive, Salt Lake City, UT 84132.

Mark Boguski 40 Newbrook Circle, Chestnut Hill, MA 02467.

Brigitta Bondy Psychiatric Hospital of University Munich, Section Psychiatric Genetics and Neurochemistry, Nussbaumstrasse 7, 80336, Munich, Germany.

Roger E. Breitbart Children's Hospital Boston, Department of Cardiology, 300 Longwood Avenue, Boston, MA 02115.

Jerome S. Brody Boston University School of Medicine, Department of Medicine, 715 Albany Street R304 Boston, MA 02118.

Lars Bullinger University of Ulm Internal Medicine III, Robert-Koch-Str. 8, 89081 Ulm, Germany.

Atul Butte Stanford University School of Medicine, Stanford Ctr. For Biomedical Informatics Research, 251 Campus Drive, Rm. X-215 MS-5479, Stanford, CA 94305.

Wendy K. Chung Columbia University, Division of Molecular Genetics Department of Pediatrics, 1150 Street Nicholas Avenue, Room 620, New York, NY 10032.

Susan E. Cottrell Amgen, Inc., 1201 Amgen Court West, Seatle, WA 98119.

Nancy J. Cox Professor and Chief, Section of Genetic Medicine, Deptartments of Medicine and Human Genetics, 5841 S. Maryland Avenue – AMB612, MC6091, Chicago, IL 60637.

Phillip G. Febbo Duke University Medical Center, Department of Medicine – Oncology, Duke Institute for Genome Sciences & Policy, Box 3382, Durham, NC 27710.

Glenn S. Gerhard Weis Center for Research/Geisinger Clinic, 100 North Academy Avenue, Danville, PA 17822-2607.

Geoffrey S. Ginsburg Duke Institute for Genome Sciences & Policy, Center for Genomic Medicine, Box 3382, Durham, NC 27710.

David B. Goldstein Duke Institute for Genome Sciences & Policy, Center for Human Genome Variation, Box 3382, Durham, NC 27710.

Iris Grossman Pharmacogenetics, Research and Development, GlaxoSmithKline, 3030 E. Cornwallis Road, Research Triangle Park, NC 27709.

Susanne Haga Duke Institute for Genome Sciences & Policy, Box 3382, Durham, NC 27710.

John Holton Royal Free & University College London Medical School, Centre for Infectious Diseases & International Health, Windeyer Institute of Medical Sciences, 46 Cleveland Street, London, W1T 4JF, UK.

Leroy Hood The Institute for Systems Biology, 1141 North 34th Street, Seattle, WA 98103-8904.

Andrew T. Huang Koo Foundation Sun Yat-Sen Cancer Center, 125 Lih-Der Rd, Pei-Tou District, Taipei, 112 Taiwan ROC.

Melissa Johnson Duke University Medical Center, Department of Medicine – Infectious Diseases, Box 3306, Durham, NC 27710.

Muin Khoury Centers for Disease Control & Prevention, Nat'l Office of Public Health Genomics, 4770 Buford Highway, MS K89, Atlanta, GA 30341.

Sean Lawler The Ohio State University Medical Center, Department of Neurological Surgery, 400 W. 12th Avenue; Room 385C, Wiseman Hall, Columbus, OH 43210.

Charles Lee Brigham & Women's Hospital, Harvard Medical School, Department of Pathology, 221 Longwood Avenue, EBRC 404A, Boston, MA 02115.

Samuel Levy J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850.

J. Alfredo Martínez University of Navarra, Department of Physiology and Nutrition, C/Irunlarrea, 1, 31008 Pamplona, Spain.

Kevin McGrath University of Pittsburgh, Division of Gastroenterology, Hepatology & Nutrition, 200 Lothrop Street, PUH, M2 – C Wing, Pittsburgh, PA 15213.

John McHutchison Duke University Clinic Research Institute, Department of Gastroenterology, Duke University Medical Center, Durham, NC.

L. Kristin Newby Duke University Medical Center, Department of Medicine – Cardiology, Box 3213, Durham, NC 27710.

Christopher Newgard Duke University Medical Center, Duke Independence Park Facility, 4321 Medical Park Drive, Suite 200, Durham, NC 27704.

Paul Noble Duke University, Div. of Pulmonary, Allergy & Critical Care Medicine, Box 3171, Durham, NC 27710.

Robert L. Nussbaum Division of Medical Genetics, Department of Medicine & UCSF Institute for Human Genetics, University of California, San Francisco, CA 94143.

Tanya Pejovic Oregon Health Sciences University, Department of Obstetrics and Gynecology, 3181 S.W. Sam Jackson Park Road, Portland, OR 97239.

Robert Plenge Harvard Medical School, Brigham and Women's Hospital, Division of Rheumatology, 75 Francis Street, Boston, MA 02115.

Thomas Quertermous Stanford University, Division of Cardiovascular Medicine, 300 Pasteur Drive, Falk CVRC, Stanford, CA 94305.

Daniel J. Rader University of Pennsylvania School of Medicine, 654 BRBII/III Labs, 421 Curie Blvd., Philadelphia, PA 19104-6160.

Sridhar Ramaswamy Massachusetts General Hospital, 185 Cambridge Street, Boston, MA 02114.

Lisa Rimsza Department of Pathology, University of Arizona, Tuscon, AZ 85724-5043, USA.

Jeffrey Ross Albany Medical College MC-80, Department of Pathology & Laboratory Medicine, 47 New Scotland Avenue, Albany, NY 12208.

M. Frances Shannon John Curtin School of Medical Research, Australian Nat'l University, Division of Molecular Bioscience, GPO Box 334, Canberra, ACT 2601 Australia.

Kevin Shianna Duke Institute for Genome Sciences and Policy, Center for Applied Genomics and Technology, Box 3568, Durham, NC 27710.

Andrew Singleton NIH, Nat'l Institute on Aging, Laboratory of Neurogenetics, Molecular Genetics Unit, Bldg. 35, Rm 1A-1000, MSC 3707, 9000 Rockville Pike, Bethesda, MD 20892.

Ralph Snyderman Chancellor Emeritus, Duke University, DUMC 3059, Durham, NC 27710.

Alison Stewart Strangeways Research Laboratory, Public Health Genetics Unit, Worts Causeway, Cambridge, CB1 8RN, UK.

Giovana Thomas University of Miami School of Medicine, Department of Otolaryngology – Head & Neck Surgery, Miami, FL 33136.

Eric J. Topol Scripps Clinic Division of Cardiovascular Disease, 10666 North Torrey Pines Rd., Mail Drop SW206, La Jolla, CA 92037.

Jeffrey A. Towbin Baylor College of Medicine, Pediatric Cardiology, One Baylor Plaza, Mail Stop BCM320, Houston, TX 77030.

Timothy D. Veenstra National Cancer Institute, SAIC Frederick, Inc., PO Box B, Frederick, MD 21702-1201.

David L. Veenstra University of Washington, Department of Pharmacy, Box 357630, Health Sciences Bldg, Rm H-275J, Seattle, WA 98195.

Howard L. Weiner Harvard Medical School, Center for Neurologic Diseases, 77 Avenue Louis Pasteur, HIM 730, Boston, MA 02115.

Scott T. Weiss Harvard Medical School, Department of Medicine, 181 Longwood Avenue, Boston, MA 02115.

Georgia Wiesner Case Western Reserve University, School of Medicine – Department of Genetics, 2109 Adelbert Road – Biomedical Research Bldg 630, Cleveland, Ohio 44106-4955.

Janey L. Wiggs Harvard Medical School and The Massachusetts Eye and Ear Infirmary, Department of Opthalmology, 243 Charles Street, Boston, MA 02114.

Cisca Wijmenga University Medical Center Groningen, Head, Department of Genetics, Room E2.030 – PO Box 30000 9700 RB Groningen, The Netherlands.

Huntington F. Willard Duke Institute for Genome Sciences & Policy, Box 3382, Durham, NC 27710.

Janet Woodcock Food and Drug Administration, Acting Deputy Commissioner for Operations, Parklawn Bldg., Room 14-71, 5600 Fishers Ln, Rockville, MD 20857.

Christopher Woods Duke University Medical Center, Department of Medicine – Infectious Diseases, Service 113, VA Med Center, 508 Fulton Street, Durham, NC 27705.

Yoram Yagil Ben-Gurion University, Department of Nephrology and Hypertension, Barzilai Medical Center Campus, Ashkelon 78306, Israel.

CHAPTER

The Foundations of Genomic and Personalized Medicine

Geoffrey S. Ginsburg and Huntington F. Willard

GENOMIC AND PERSONALIZED MEDICINE

The Human Genome Project, completed in 2003, has provided scientists and clinicians with a diverse set of novel molecular tools that can be used to understand health and manage disease. Variation in the human genome has long been the cornerstone of the field of human genetics (see Box 1.1), and its study led to the establishment of the medical specialty of medical genetics (Nussbaum et al., 2007). Now genome sequencing, copy number variation, transcriptional readouts, and comprehensive measurements of micro RNA, protein, and metabolite levels provide a "systems approach" to probe and predict human health and disease states that has greatly broadened the impact of principles of genetics and genomics on clinical medicine. These advances have provided both a conceptual and technological underpinning for the development of the field of genomic medicine as a driver of personalized health care. For the first time in the history of medicine, health care providers as well as patients can use predictive tools to develop a new model for health care based on health planning that is *proactive* and preventive, as opposed to the current model in health care that is *reactive*, episodic, and geared toward acute crisis intervention once disease is already manifest and largely irreversible.

The growing transformation of clinical practice in the era of genomic and personalized medicine is perhaps best exemplified today in the field of cancer care, as illustrated in some detail in subsequent chapters of this volume. Oncologists now practice with a suite of genomic testing opportunities that include *BRAC1/BRAC2* testing in familial syndromes of breast and ovarian cancers. In colorectal cancer, Hereditary Non-polyposis Colon Cancer (HNPCC) or Lynch syndrome and familial adenomatous polyposis (FAP) coli are conditions for which there is testing for mismatch repair gene mutations (for HNPCC) or *APC* mutations (FAP) that has been widely adopted. The paradigm for oncology is largely based on the principle that accurate prognosis and proper therapy can be matched to the molecular characteristics of the individual patient's tumor. Thus, at the time of diagnosis, whole-genome expression data are now being used routinely to identify subtypes of cancer not previously recognized by traditional methods of analysis (Bullinger and Valk, 2005; Dave et al., 2006; Potti et al., 2006a; Staudt, 2003; Valk et al., 2004).

There is now compelling evidence of clinical adoption of genomic testing by oncologists: in 2008, for example, RNA expression signatures were used for risk stratification and prognosis in breast cancer for more than 39,000 "treat" versus "no-treat" decisions (Securities and Exchange Commission, 2009). However, there continues to be a need to develop the

Copyright © 2010, Elsevier Inc.
All rights reserved.

BOX 1.1 Genetics and Genomics

Throughout this and the many other chapters in this *Essentials* volume, the terms "genetics" and "genomics" are used repeatedly, both as nouns and in their adjectival forms. While these terms seem similar, they in fact describe quite distinct (though frequently overlapping) approaches in biology and in medicine. Having said that, there are inconsistencies in the way the terms are used, even by those who work in the field. To some, genetics is a subfield of genomics; to others, genomics is a subfield of genetics. Arguably, depending on the perspective one has in mind, both may be right!

Here, we provide operational definitions to distinguish the various terms and the subfields of medicine to which they contribute.

The field of *genetics* is the scientific study of heredity and of the genes that provide the physical, biological, and conceptual bases for heredity and inheritance. To say that something – a trait, a disease, a code, or information – is genetic refers to its basis in genes and in DNA.

Heredity refers to the familial phenomenon whereby traits (including clinical traits) are transmitted from generation to generation, due to the transmission of genes from parent to child. A disease that is said to be inherited or hereditary is certainly genetic; however, not all genetic diseases are hereditary (witness cancer, which is always a genetic disease, but is only occasionally an inherited disease).

Genomics is the scientific study of a genome or genomes. A *genome* is the complete DNA sequence, containing the entire genetic information of a gamete, an individual, a population, or a species. As such, it is a subfield of genetics when describing an approach taken to study genes. The word "genome" originated as an analogy with the earlier term "chromosome," referring to the physical entities (visible under the microscope) that carry genes from one cell to its daughter cells or from one generation to the next. Genomics gave birth to a series of other "-omics" that refer to the comprehensive study of the full complement of genome products – for example, proteins (hence, *proteomics*), transcripts (*transcriptomics*), or metabolites (*metabolomics*). The essential feature of the "omes" is that they refer to the complete collection of genes or their derivative proteins, transcripts, or metabolites, not just to the study of individual entities. While formally the field of genomics refers to the study of genomes (and hence, DNA) only, it sometimes takes on the broader meaning of referring to any large-scale approach; the less specific term "genome sciences" is also sometimes used to refer to all of the -omics to connote global and comprehensive approaches to the study of biology and medicine.

By analogy with genetics and genomics, *epigenetics* and *epigenomics* refer to the study of factors that affect gene (or, more globally, genome) function, but without an accompanying change in genes or the genome. Some typical epigenetic factors involve changes in DNA methylation or modifications to chromatin that change genome structure and hence influence gene expression even in the absence of changes in the DNA sequence. The *epigenome* is the comprehensive set of epigenetic changes in a given individual, tissue, tumor, or population. It is the paired combination of the genome and the epigenome that best characterize and determine one's phenotype.

Medical genetics is the application of genetics to medicine with a particular emphasis on inherited disease. Medical genetics is a broad and varied field, encompassing many different subfields, including clinical genetics, biochemical genetics, cytogenetics, molecular genetics, the genetics of common diseases, and genetic counseling. Medical genetics is one of 24 medical specialties recognized by The American Board of Medical Specialties, the preeminent medical organization overseeing physician certification in the United States. As of 2007, there were approximately 2300 board-certified medical geneticists in the United States.

Genetic medicine is a term sometimes used to refer to the application of genetic principles to the practice of medicine and thus overlaps medical genetics. However, genetic medicine is somewhat broader, as it is not limited to the specialty of Medical Genetics but is relevant to health professionals in many, if not all, specialties and subspecialties. Both medical genetics and genetic medicine approach clinical care largely through consideration of individual genes and their effects on patients and their families.

By contrast, *genomic medicine* refers to the use of large-scale genomic information and to the consideration of the full extent of an individual's genome, proteome, transcriptome, metabolome, and/or epigenome in the practice of medicine and medical decision-making. The principles and approaches of genomic medicine are relevant well beyond the traditional purview of medical genetics and include, as examples, gene expression profiling to characterize tumors or to define prognosis in cancer, genotyping variants in the set of genes involved in drug metabolism or action to determine an individual's correct therapeutic dosage, scanning the entire genome for millions of variants that influence one's susceptibility to disease, or analyzing multiple protein biomarkers to monitor therapy and to provide predictive information in presymptomatic individuals.

Finally, *personalized medicine* refers to a rapidly advancing field of health care that is informed by each person's unique clinical, genetic, genomic, and environmental information. The goals of personalized medicine are to take advantage of a molecular understanding of disease to optimize preventive health care strategies and drug therapies while people are still well or at the earliest stages of disease. Because these factors are different for every person, the nature of disease, its onset, its course, and how it might respond to drug or other interventions are as individual as the people who have them. In order for personalized medicine to be used by health care providers and their patients, these findings must be translated into precision diagnostic tests and targeted therapies. Since the overarching goal is to optimize medical care and outcomes for each individual, treatments, medication types and dosages, and/or prevention strategies may differ from person to person – resulting in unprecedented customization of patient care.

The principles underlying genomic and personalized medicine and their applications to the practice of clinical medicine are presented throughout the chapters that comprise this volume.

---

evidence that these signatures will enhance outcomes across a broad population of women with breast cancer. A prospective cooperative group clinical trial has been initiated in Europe (MINDACT) that aims to measure the effectiveness of a gene expression predictor of breast cancer prognosis to guiding adjuvant chemotherapy when compared to predictions based solely on the traditional clinical parameters for prognoses (Bogaerts et al., 2006). A study sponsored by the US National Cancer Institute (TAILORx) aims to use an RNA signature to identify low risk breast cancer patients unlikely to benefit from

chemotherapy (Sparano, 2006). Similar studies are getting started for prognostic signatures in lung cancer (Potti et al., 2006a). These are exemplars of the approach of genomic and personalized medicine and serve as clear examples where a genome-based approach has resulted in the opportunity to redefine disease phenotypes and at the same time redefine therapeutic strategies.

Of equal if not greater importance in achieving the goal of personalized treatment is an ability to predict response to specified therapies, particularly for those regimens that are part of routine clinical practice today. The selection of therapy for many cancer patients is still largely empiric and guided by large randomized clinical trials on populations of patients. Estimates of benefits from this approach for individuals are extrapolations from the effects seen in these large trials and do not necessarily apply to individual patients.

Genomic signatures that predict response and resistance to a spectrum of cytotoxic chemotherapies may now allow assignment of patients to effective treatment regimens best suited to the unique characteristics of their tumor (Potti et al., 2006b). Genomic predictors of chemotherapy response thus provide an opportunity to determine which drug would be optimal for an individual patient in clinical scenarios for which past studies have not shown a clear superiority for any of the currently available drugs.

Beyond cancer, other fields of medicine are also benefiting from whole-genome approaches that are defining both susceptibility to complex disease through genome-wide association studies (GWAS) or genome resequencing (see below), as well as signatures that define disease states and predictive outcomes based on analyses from both disease tissues and from blood (Altshuler et al., 2008; Aziz et al., 2007). Blood-based gene expression profiling is particularly important as it presents the opportunity to report on disease processes from remote and often inaccessible sites for direct analyses. Instead of analyzing single genes, global gene expression provides a "molecular signature" that may distinguish between one disease state and another. In addition to identifying signatures or patterns of gene expression that represent a disease state, analyses can be constructed to identify representative pathway genes that might point to novel pathophysiology relevant to the underlying disease state. Peripheral blood gene expression signatures have now been reported in a variety of conditions described in subsequent chapters, including rheumatoid arthritis (Lequerré et al., 2006; Shou et al., 2006), systemic lupus erythematosis (Rus et al., 2004), multiple sclerosis (Bomprezzi et al., 2003; Singh et al., 2007), asthma (Brutsche et al., 2002), malignancies (Alizadeh et al., 2000; Golub et al., 1999), solid organ transplantation (Baron et al., 2007; Horwitz et al., 2004), as well as environmental exposures (Dressman et al., 2007; Lodovici et al., 2007; Meadows et al., 2008; Wu et al., 2003). Many of these conditions have an inflammatory component and thus affect immune cells in the vascular compartment. It is hypothesized that these cellular changes are the basis for the differences in gene expression that is observed in RNA extracted from whole blood specimens or from specific circulating cell types. The greatest potential of this approach would be the enhancement in accurately classifying patients by the type and severity of their disease and to individualize the therapy based on the biology of the disease in an individual patient. Significant power can be anticipated from approaches that combine *both* assessment of an individual's constitutional genome and evaluation of gene expression signatures of one's health status (Figure 1.1) (Hardy and Singleton, 2009).

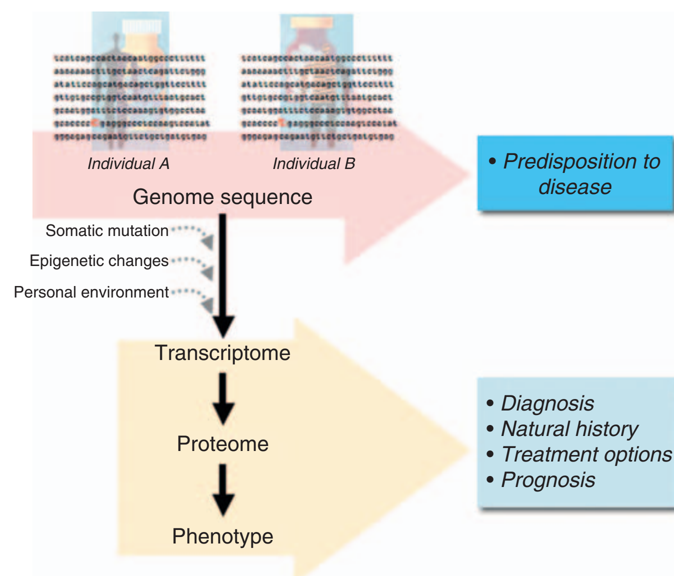


Figure 1.1 The promise of personal genomics in the era of genomic and personalized medicine.

GENES, GENOMES, AND DISEASE

In the context of genomic and personalized medicine, a key question is to what extent genetic variation influences the likelihood of disease onset; determines or signals the natural history of disease; and/or provides clues relevant to the management of disease. Variation in one's constitutional genome can have a number of different direct or indirect effects on gene expression, thus contributing to the likelihood of disease (Frazer et al., 2009). It is not, however, just the human genome whose variation is relevant to an individual's state of health, but there are thousands of microorganisms, both symbiotic and pathogenic, whose genomes are also relevant to human phenotypes, and sequence determination of their genomes is providing new insights and approaches for the diagnosis, study, and treatment of infectious disease (see Box 1.2).

Contemporary frameworks for considering the impact of variation in the human genome on disease build on decades of success in establishing the role of individual, typically rare, mutations as causal determinants of now more than 2000 simple Mendelian diseases (Online Mendelian Inheritance in Man, 2009). As a result of that success, much of which paralleled the development of technologies in the early stages of the Human Genome Project (Altshuler et al., 2008; Peltonen and McKusick, 2001), attention has now turned to the genes presumed to underlie susceptibility to common complex diseases, which are the subject of many of the subsequent chapters in this book.

There are two distinct, but nonexclusive models for thinking about human genetic variation and disease (Altshuler and Clark, 2005; Fearnhead et al., 2005; Florez et al., 2003; Pritchard, 2001). One – the "common allele, common disease" hypothesis – posits that variation common in the population accounts for the relatively higher or lower risk that some individuals (and their families) have for a particular condition. Under this model, the collection of 10–15 million common variants in the genome (see Chapter 2) underlies the range of susceptibility that one finds in the general population, modulated by the particular and often variable environmental inputs and factors that are present in that population and that may, in fact, shift or even obscure the relative impact of inherited factors. An alternative model – the rare variant hypothesis – argues that genetic susceptibility to disease is due to the accumulated risk conferred by multiple rare variants in an individual's genome, variants therefore not likely to be captured by study of the common variants identified by studies to date (Goldstein, 2009).

These two hypotheses suggest different approaches that will likely be informative for delineating the genetic contribution to disease, both for designing research studies and for eventual clinical surveillance. It is worth emphasizing that these two hypotheses are not mutually exclusive and are each likely to be correct in some cases; indeed, there is already evidence supporting each for different diseases (Frazer et al., 2009).

Genome-Wide Association Studies

The common allele, common disease hypothesis has been explored with notable success in a number of conditions, utilizing large cohorts of well-phenotyped patients and high-throughput methods to genotype up to 500,000 or a million variants (so-called single nucleotide polymorphisms [SNPs], see Chapter 2) in the genome (Manolio et al., 2008). These

BOX 1.2 The Genomes Within

The human genome is not the only genome relevant to the practice of medicine.

Both in states of health and disease, our own genome is vastly outnumbered by the genomes of a host of microorganisms, many living peacefully and continuously on various body surfaces, especially throughout the gastrointestinal tract, others wreaking havoc as adventitious viral, bacterial, or fungal pathogens.

The genomes of thousands of microorganisms have been determined and are being utilized to provide rapid diagnostic tests in clinical settings, to predict antibiotic or antifungal efficacy, to identify the source of airborne, water, or soil contaminants, to monitor hospital or community environments, and to better understand the contribution of microbial ecosystems and various environmental exposures to diverse human phenotypes.

The human colon contains more than 400 bacterial species comprising some 10^{13}–10^{14} microorganisms. Each adult's gut provides a unique environment – the *microbiome* – whose origins and impact on human disease are just being explored. The microbiotic gene set is significantly different from that of the human genome and thus has the capacity to alter the metabolic profile of different individuals or different populations, with clinically meaningful effects on drug metabolism, toxicity, and efficacy (Gill et al., 2006; Li et al., 2008; Palmer et al., 2007). The applications to microbiomes of approaches in genomics (as well as proteomics and transcriptomics) are revolutionizing clinical diagnostics, for example, to identify unknown viral infections (Delwart, 2007; Long et al., 2004; Wang et al., 2003) or to diagnose antibiotic-resistance infections such as methicillin-resistant *Staphylococcus aureus* (MRSA) (Francois et al., 2007).

The field of *metagenomics* explores this heterogeneous ecosystem by comprehensive sequence analysis of the collected genomes from biological specimens (such as stool, urine, sputum, water sources, and air), followed by both taxonomic and bioinformatic analysis to deconvolute the many genomes contained in such specimens and to define the different organisms, their genes, and genome variants. This approach is particularly informative for characterizing organisms that cannot be cultured in standard microbiology labs. A number of diseases have been associated with large-scale imbalances in the gut microbiome, including Crohn's disease, ulcerative colitis, antibiotic-resistant diarrhea, and obesity (Frank and Pace, 2008; Ley et al., 2006; Turbaugh et al., 2006).

Undoubtedly, the states of health and disease are determined in part by the balance of genomes both within us and external to us. The full complement of genomic information from both of these sources of genomes will provide insights into defining the states of health and disease and the basis for unsurpassed precision in both the prevention of disease and its treatment.

GWAS report the statistical association of one or more variants in a narrow genomic region (which may or may not contain an annotated gene) with the presence or absence of the clinical condition. The reported SNPs define immediately accessible risk factors for that condition, at least in the population(s) under study and can provide novel insights into the biology of the disease. It should be stressed, however, that in most instances, causality of the reported SNP(s) and the increased risk has not been proved; it may be that the actual causal variant is not the SNP itself but is a currently undetected variant that lies in linkage disequilibrium with the SNP (Frazer et al., 2009).

In the most favorable cases, the associated SNP may be a nonsynonymous variant, leading to a pathological amino acid change in the relevant gene (Thorleifsson et al., 2007), or may be a variant in an RNA splice site, leading to a clinically meaningful change in the production of the gene's transcript(s) (Heinzen et al., 2007). But in most instances, the functional impact of the associated SNP is obscure, notwithstanding very clear genetic evidence of a role of genome variation in susceptibility to the particular condition. SNPs are but one of several types of genome variation that can influence gene expression and/or disease (see Chapter 2). Copy number variants have also been associated with some disorders (reviewed in Estivill and Armengol, 2007), and it will require integrated genomic, genetic and functional studies to elucidate the precise basis for the role(s) of genome variation in different diseases.

Medical Resequencing in Search of Rare Variants

While GWAS certainly establish that common alleles do indeed increase susceptibility to common disease in some instances, they do not allow one to conclude that this is always the case. Indeed, in most (but not all) cases to date, the common SNPs found to be associated with disease only explain a small fraction of the total genetic variation, implicating an as yet undiscovered (and presumably rarer) basis for most genetic variation underlying a given condition. While such association is sufficient to conclude that a particular variant does indeed contribute to disease in the population under study, such a finding is insufficient to say anything declarative about the cause of our likelihood of disease in a particular case. This conclusion is, of course, highly relevant to the prospects of genomic and personalized medicine.

An alternative or complementary approach to genome-wide genotyping of common variants is to resequence specific genes in a cohort of affected individuals, in an effort to uncover rare variants responsible for (or at least statistically associated with) the disease in question. To date, most efforts have focused on one or several genes that were believed to be strong candidates for the phenotype understudy; the notable exception thus far is exemplified by the whole-genome resequencing of the HuRef genome and correlation of novel variants detected in that study with his family and personal medical histories (Levy et al., 2007). Rare variants, including nonsynonymous variants, in relevant candidate genes have been detected at a statistically significant higher frequency in the genomes of patients with colorectal adenomas (Fearnhead et al., 2004), with low-plasma high-density lipoprotein cholesterol (HDL-C) (Cohen et al., 2004), with triglyceride levels in the lowest quartile (Romeo et al., 2007), and with familial cases of X-linked mental retardation (Tarpey et al., 2009).

These successes point to a strategy of resequencing relevant genes in individuals at the extremes of the population distribution for measurable traits (Topol and Frazer, 2007) in which the only limiting parameters are the cost of sequencing and the quality of the phenotypic or quantitative data. In a proof-of-principle for this case-control approach, a recent study sequenced coding exons and splice junctions of 58 genes in nearly 400 obese and lean individuals, at the >95th or <10th percentile of body mass index (Ahituv et al., 2007). Of the ~1000 variants detected, most were rare variants, including over 270 nonsynonymous mutations; many were found only in the obese cohort and thus become strong functional candidates for a role in obesity.

Searching for Somatic Mutations

While GWAS are restricted to inherited variation, medical resequencing studies can target either inherited or somatic variants. In cancer especially, it is of interest to use medical resequencing to search for somatic mutations in tumor tissue in order to identify genes potentially relevant to cancer development and/or progression (Stratton et al., 2009). Two important points emerge from such studies (Greenman et al., 2007; Jones et al., 2009; Sjoblom et al., 2006). First, the genes implicated by virtue of discovering rare somatic variants in multiple cases of a particular cancer tend to be different from those identified in previous genetic studies as inherited risk factors. This provides novel insights into the biology of human cancer and suggests candidates for further exploring mechanisms of tumorigenesis or metastasis or for developing therapeutic approaches. Second, however, the large number of mutations uncovered by this approach introduces the need for caution, as many will be "bystander" or "passenger" mutations only associated with cancer, not genes involved directly in cancer.

FROM THE GENOME TO PERSONALIZED MEDICINE

Of all the promises of the current scientific and social revolution stemming from advances in our understanding of the human genome and its variation, genomic and personalized medicine has been the most eagerly awaited. The prospect of examining an individual's entire genome (or at least a significant fraction of it) in order to make individualized risk predictions and treatment decisions is an attractive, albeit challenging, one (Bentley, 2004; Kraft and Hunter, 2009; Willard et al., 2005).

Having access to the reference human genome sequence has been transformational for the fields of human genetics and

genome biology but by itself is an insufficient prerequisite for genomic medicine. As alluded to above and discussed in subsequent chapters, equally important are the various complementary technologies to reliably capture information on individual genomes, their epigenetic modification, and their derivatives in the transcriptome, proteome, and metabolome for health and disease status (Figure 1.1). Each of these technologies provides information that, in combination with clinical data and evaluation of environmental triggers, will in time contribute to assessment of individual risks and guide clinical management and decision-making. Critical enablers of this new approach have been innovations in laboratory technology (to address biologically and medically relevant questions on a scale and with a throughput hardly imaginable just a few decades ago), paired with equally transforming developments in informatics and information systems to handle the onslaught of genomic data (West et al., 2006).

Personal Genomics

At the heart of the genomic approach to personalized medicine will be information from individual genomes, a fast-moving area of technological development that is spawning a social and information revolution among consumers. Dramatic improvements in sequencing technology (Bentley, 2006; Mardis, 2008) have reduced the cost and time of resequencing projects to a level that invites conjecture about the long awaited "$1000 genome" (Dalton, 2006; Service, 2006; Wolinsky, 2007). The much-ballyhooed release of the genome sequences of Craig Venter ("HuRef") (Levy et al., 2007) and James Watson (Wheeler et al., 2008) has stirred up concerns about "celebrity genomics" (Check, 2007); however, while the Venter and Watson sequences may have been the first, they have been rapidly followed by several others (see Chapter 2), and numerous additional genomes are already in various sequencing pipelines.

What remains unsettled for now is what degree of genome surveillance will be most useful, either for research or for clinical practice, a topic that is discusssed frequently in subsequent chapters. Some have argued that, while whole-genome sequencing is increasingly possible, it is unlikely to provide more information about established disease associations than would high-density, genome-wide genotyping to detect both SNP and copy number variation. Targeted resequencing of, for example, exons and known regulatory regions would allow detection of rare variants in portions of the genome most relevant to disease at a fraction of the cost of whole-genome sequencing (Hodges et al., 2007).

The availability of associated clinical data is variable among the studies announced to date, but there exists, at least among some participants, a strong sense of "health-information altruism" to contribute to the much needed large-scale correlation of genotype and phenotype (Kohane and Altman, 2005). Notwithstanding individual's willingness to make genome sequence data (much less medical information) available more-or-less publicly, substantial concerns have been raised about privacy, since a surprisingly (to some) small number of SNPs or other genome variants are sufficient to allow identification of individuals (Lin et al., 2004; McGuire and Gibbs, 2006; McGuire et al., 2008). Absolute privacy and anonymity may be an impossible standard.

A Consumer Revolution

While the genome revolution has without doubt been driven by technological improvements and by an explosion in the availability of genome data, the push for incorporation of genome information into clinical practice may come as much or more from consumers as from professionals. A half dozen or more companies are already offering genome-wide SNP profiles to the public, some with associated risk estimates for relevant clinical conditions. At least a few companies will also sequence individual genomes for a cost that, while high currently, is not out of reach for some individuals. There is a clear and important research agenda that needs to be developed in concert with these technological breakthroughs that allows health providers and the public to understand the information and more so to believe that is accurate, informative, and actionable (Feero et al., 2008; Hunter et al., 2008; McGuire et al., 2007, 2008). With the vast amount of information contained in the human genome sequence, the stakes are high for patients, physicians, and the public to ensure proper reading, interpretation, and communication of the information are carried out (Janssens et al., 2008). This may be a "disruptive technology" in health care delivery with the provision of health and disease risk information to consumers without physician intervention and guidance. It will not be long before a patient will bring a report of a whole genome to a physician's office and ask for guidance. What will the physicians of today, armed with a paucity of genomic training, tell them?

CHALLENGES IN THE TRANSLATION OF GENOMICS TO HUMAN HEALTH

Despite the clear advances in technology to bring genomic information closer to physicians, patients, and the public, looming ever closer are issues that are outside of the sphere of the scientists that have been involved in the discovery and early translational activities. In the United States, the Institute of Medicine has convened a Roundtable on Translating Genomics to Health (Institute of Medicine, 2007), the Centers for Disease Control has independently developed a pilot project on the Evaluation of Genomic Applications in Practice and Prevention (EGAPP, 2004), and the Genomic and Personalized Medicine Act of 2007 (Personalized Medicine, 2007) was introduced in Congress to also address similar issues.

Recently, Scheuner et al. (2008) carried out an extensive meta-analysis of studies using genomics toward clinical application in chronic disease. This study aimed to understand the current state of translation focusing on the following questions: "What are the outcomes of genomic medicine? What is the current level of consumer understanding about genomic medicine and what information do consumers need before they seek services? How is genomic medicine best delivered? What are the

challenges and barriers to integrating genomic medicine into clinical practice?"

Using a total of 68 articles in their analysis, the authors synthesized information on the delivery of genomic medicine for common adult-onset conditions (Scheuner et al., 2008). The major findings of the study were not surprising in terms of the genome policy issues that have been previously summarized (Haga and Willard, 2006).

- Education – The primary care workforce feels woefully unprepared to integrate genomics into regular practice. Consumers are also unclear about genetics and genetic testing for common diseases.
- Privacy – Consumers are worried about the possible adverse consequences of genetic testing, particularly the privacy issues and discrimination against receiving employment and health insurance.
- Evidence – There needs to be outcome data for genetic testing and chronic disease to assess whether patients who receive the test do better clinically.
- Cost – Cost uncertainty (both in terms of delivery and reimbursement) is an important issue to many of the stakeholders of genomic issues.

Health Professional and Public Education

Education of health professionals and the public will be essential to advancing the use of genomics into health care (Frueh et al., 2005). With all of the rapid advancements in genomics research and technologies, it will be challenging to keep health professionals informed about the benefits, risks, and limitations of new tools as they become available. In addition, the public and health care workforce will need to understand the appropriate applications of genomic tools, including their benefits, risks, and limitations; how they may improve clinical management; inherited versus acquired genomic variations (e.g., implications for family members); and privacy and confidentiality. Although several surveys have documented the below average physician knowledge of genetics (Metcalfe et al., 2002), none has assessed knowledge of the newer field of genomics. But, several papers have been published recognizing the importance of pharmacogenetics (Frueh and Gurwitz, 2004), and steps are underway to develop more educational materials in this area.

Privacy Fears

There has been ongoing debate about the uniqueness of genetic information and whether it warrants special protections beyond those in place for standard medical information (Haga and Willard, 2006). In the United States, fear of discrimination by employers and health insurers is the main concern, whereas in the United Kingdom, use of genetic information by life insurers is the major concern (Apse et al., 2004; Hall et al., 2005). Despite the outcome of these debates, the attention paid to genetics by the popular press and public has raised concern about genetic information. The potential for genetic discrimination has been a major concern for researchers, health professionals, patients, and the public. In order for genomic biomarkers to be integrated into routine clinical practice, associated fears with this type of testing must be put to rest. In 2008, the United States enacted legislation to protect against genetic discrimination by employers and health insurers. This is an important step in allaying public fears that may otherwise hinder a comprehensive genomic and personalized approach to medicine.

Building the Evidence for Clinical Utility

Perhaps the most important factor hindering the appropriate integration of genomics into clinical practice is the lack of evidence for its clinical utility (i.e., evidence that use of a genomic technology actually improves health outcomes). Evidence generation needs to be more practicable and practical. There is also a need for greater collaboration among stakeholder groups and for innovation in both study design and analysis methods. Clinical outcome studies are needed that demonstrate the clinical utility of genomic interventions that are linked to specific, actionable clinical recommendations, or practice guidelines. Public–private partnerships are likely to be required to generate the evidence base for genomic medicine. These collaborations are desirable because no signal stakeholder group is likely to have sufficient resources or expertize to conduct the necessary studies.

Cost Issues

As with any new innovation, genomic testing must be demonstrated to be clinically useful and cost effective and of value. But, because genomic technologies inherently involve diagnostic or prognostic testing, in addition to the complexities of incomplete gene penetrance and multiple gene × environment interactions, their assessment can be more challenging. In addition, perhaps more than in any other area of medicine, questions have arisen in regard to the economic incentives to develop these technologies. Formal health economics frameworks can be used to gain insights into these issues and provide guidance for research and development (Carlson et al., 2005; Flowers and Veenstra, 2004; Stallings et al., 2006). It is important to examine the drivers of cost effectiveness of genomic technologies and to consider approaches that include value-based reimbursement for genomic testing technologies. A particularly challenging area is pharmacogenomics, in which the economic incentives for developing diagnostics linked to therapeutics in the pharmaceutical industry are unclear. An integrated business model is needed that will be favorable for the effective delivery of genomic information to patients and clinicians.

THE FUTURE OF PERSONALIZED MEDICAL CARE

While the human genome sequence is now available, it is important to acknowledge that our knowledge of the genome and its biological complexity is nowhere near complete, and the use of genomic protocols in standard clinical care faces many challenges. There are a host of clinical, economic, insurance, privacy, and

commercialization concerns that will need to be addressed and that vary substantially among different countries. As a field, we can confront those with the certainty that the science behind genomic medicine is sound and the practice of medicine that it informs is evidence based. These issues are being dealt with systematically, and the prospects of using genomic information to offer patients health care that is truly personalized in nature – as will clearly see in the subsequent chapters – is finally within our reach.

REFERENCES

Ahituv, N., Kavaslar, N., Schackwitz, W., Ustaszewska, A., Martin, J., Hebert, S. Doelle, H., Ersoy, B., Kryukov, G., Schmidt, S. et al. (2007). Medical sequencing at the extremes of human body mass. *Am J Hum Genet* 80, 779–791.

Altshuler, D. and Clark, A. (2005). Harvesting medical information from the human family tree. *Science* 307, 1052–1053.

Altshuler, D., Daly, M. and Lander, E. (2008). Genetic mapping in human disease. *Science* 322, 881–888.

Alizadeh, A., Eisen, M., Davis, R., Ma, C., Lossos, I., Rosenwald, A., Boldrick, J., Sabet, H., Tran, T., Yu, X. et al. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403, 503–511.

Apse, K., Biesecker, B., Giardiello, F., Fuller, B. and Bernhardt, B. (2004). Perceptions of genetic discrimination among at-risk relatives of colorectal cancer patients. *Genet Med* 6, 510–516.

Aziz, H., Zaas, A. and Ginsburg, G.S. (2007). Peripheral blood gene expression profiling for cardiovascular disease assessment. *Genomic Med* 1, 105–112.

Baron, C., Somogyi, R., Greller, L., Rineau, V., Wilkinson, P., Cho, C., Cameron, M., Kelvin, D., Chagnon, R., Roy, D. et al. (2007). Prediction of graft-versus-host disease in humans by donor gene-expression profiling. *PLoS Med* 4(1), e23.

Bentley, D.R. (2004). Genomes for medicine. *Nature* 429, 440–445.

Bentley, D.R. (2006). Whole-genome resequencing. *Curr Opin Genet Devel* 16, 545–552.

Brutsche, M., Joos, L., Carlen, B., Bissinger, R., Tamm, M., Custovic, A. and Woodcock, A. (2002). Array-based diagnostic gene-expression score for atopy and asthma. *J Allergy Clin Immunol* 109, 271–273.

Bomprezzi, R., Ringner, M., Kim, S., Bittner, M., Khan, J., Chan, Y., Elkahloun, A., Yu, A., Bielekova, B., Meltzer, P. et al. (2003). Gene expression profile in multiple sclerosis patients and healthy controls: Identifying pathways relevant to disease. *Hum Mol Genet* 12, 2191–2199.

Bogaerts, J., Cardoso, F., Buyse, M., Braga, S., Loi, S., Harrison, J., Bines, J., Mook, S., Decker, N., Ravdin, P. et al. (2006). TRANSBIG consortium. Gene signature evaluation as a prognostic tool: challenges in the design of the MINDACT trial. *Nat Clin Pract Oncol* 10, 540–551.

Bullinger, L. and Valk, P.J. (2005). Gene expression profiling in acute myeloid leukemia. *J Clin Oncol* 23, 6296–6305.

Carlson, J., Henrikson, N., Veenstra, D. and Ramsey, S. (2005). Economic analyses of human genetics services: A systematic review. *Genet Med* 7, 519–523.

Check, E. (2007). Celebrity genomes alarm researchers. *Nature* 447, 358–359.

Cohen, J., Kiss, R., Pertsemlidis, A., Marcel, Y., McPherson, R. and Hobbs, H. (2004). Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* 305, 869–871.

Dalton, R. (2006). Sequences step up to the speed challenge. *Nature* 443, 258–259.

Dave, S.S., Fu, K., Wright, G.W., Lam, L., Kluin, P., Boerma, E., Greiner, T., Weisenburger, D., Rosenwald, A., Ott, G. et al. (2006). Lymphoma/Leukemia Molecular Profiling Project. Molecular diagnosis of Burkitt's lymphoma. *N Engl J Med* 354, 2431–2442.

Delwart, E.L. (2007). Viral Metagenomics. *Rev Med Virol* 17, 115–131.

Dressman, H., Muramoto, G., Chao, N., Meadows, S., Marshall, D., Ginsburg, G., Nevins, J. and Chute, J. (2007). Gene expression signatures that predict radiation exposure in mice and humans. *PLoS Med* 4, e106.

EGAPP. (2004). http://www.egappreviews.org/about.htm

Estivill, X. and Armengol, L. (2007). Copy number variants and common disorders: Filling the gaps and exploring complexity in genome-wide association studies. *PLoS Genet* 3, e190.

Fearnhead, N.S., Wilding, J., Winney, B., Tonks, S., Bartlett, S., Bicknell, D., Tomlinson, I., Montensen, N.J. and Bodmer, W.F. (2004). Multiple rare variants in different genes account for multifactorial inherited susceptibility to colorectal adenomas. *Proc Natl Acad Sci USA* 101, 15992–15997.

Fearnhead, N.S., Winney, B. and Bodmer, W.F. (2005). Rare variant hypothesis for multifactorial inheritance: Susceptibility to colorectal adenomas as a model. *Cell Cycle* 4, 521–525.

Feero, W., Guttmacher, A. and Collins, F. (2008). The genome gets personal – Almost. *J Am Med Assoc* 299, 1351–1352.

Florez, J., Hirschhorn, J. and Altshuler, D. (2003). The inherited basis of diabetes mellitus: Implications for the genetic analysis of complex traits. *Annu Rev Genomics Hum. Genet* 4, 257–291.

Flowers, C.R. and Veenstra, D. (2004). The role of cost-effectiveness analysis in the era of pharmacogenomics. *Pharmacoeconomics* 22, 481–493.

Francois, P., Scherl, A., Hochstrasser, D. and Schrenzel, J. (2007). Proteomic approach to investigate MRSA. *Methods Mol Biol* 391, 179–199.

Frank, D.N. and Pace, N.R. (2008). Gastrointestinal microbiology enters the metagenomics era. *Curr Opin Gastroenterol* 24, 4–10.

Frazer, K., Murray, S., Schork, N. and Topol, E. (2009). Human genetic variation and its contribution to complex traits. *Nat Rev Genet* 10, 241–251.

Frueh, F.W. and Gurwitz, D. (2004). From pharmacogenetics to personalized medicine: A vital need for educating health professionals and the community. *Pharmacogenomics* 5, 571–579.

Frueh, F.W., Goodsaid, F., Rudman, A., Huang, S.M. and Lesko, L.J. (2005). The need for education in pharmacogenomics: A regulatory perspective. *Pharmacogenomics J* 5, 218–220.

Gill, S.R., Pop, M., DeBoy, R., Eckburg, P.B., Turnbaugh, P.J., Samuel, B.S., Gordon, J.I., Relman, D.A., Fraser-Liggett, C.M. and Nelson K.E. (2006). Metagenomic analysis of the human distal gut microbiome. *Science* 312, 1355–1359.

Goldstein, D.B. (2009). Common genetic variation and human traits. *N Engl J Med* 360, 1696–1698.

Golub, T., Slonim, D., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J., Coller, H., Loh, M., Downing, J., Caliguri, M. et al. (1999). Molecular

classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* 286, 531–537.

Greenman, C., Stephens, P., Smith, R., Dalgliesh, G., Hunter, C., Bignell, G., Davies, H., Teague, J., Stevens, C., Edkins, S. et al. (2007). Patterns of somatic mutation in human cancer genomes. *Nature* 446, 153–158.

Haga, S. and Willard, H. (2006). Defining the spectrum of genome policy. *Nat Rev Genet* 7, 966–972.

Hall, M., McEwen, J., Barton, J., Walker, A., Howe, E., Reis, J., Power, T., Ellis, S., Tucker, D., Harrison, B. et al. (2005). Concerns in a primary care population about genetic discrimination by insurers. *Genet Med* 7, 311–316.

Hardy, J. and Singleton, A. (2009). Genomewide association studies and human disease. *N Engl J Med* 360, 1759–1768.

Heinzen, E., Yoon, W., Tate, S., Sen, A., Wood, N., Sisodiya, S. and Goldstein, D. (2007). Nova2 interacts with a *cis*-acting polymorphism to influence the proportions of drug-responsive slice variants of SCN1A. *Am J Hum Genet* 80, 876–883.

Hodges, E., Xuan, Z., Balija, V., Kramer, M., Molla, M., Smith, S., Middle, C., Rosesch, M., Albert, T., Hannon, G. et al. (2007). Genome-wide *in situ* exon capture for selective resequencing. *Nat Genet* 39, 1522–1527.

Horwitz, P., Tsai, E., Putt, M., Gilmore, J., Lepore, J., Parmacek, M., Kao, A., Desai, S., Goldberg, L., Brozena, S. et al. (2004). Detection of cardiac allograft rejection and response to immunosuppressive therapy with peripheral blood gene expression. *Circulation* 110, 3815.

Hunter, D., Khoury, M. and Drazen, J. (2008). Letting the genome out of the bottle – Will we get our wish? *N Engl J Med* 358, 105–107.

Institute of Medicine. (2007). http://www.iom.edu/CMS/3740/44443.aspx

Janssens, A., Gwinn, M., Bradley, L., Oostra, B., van Duijn, C. and Khoury, M. (2008). A critical appraisal of the scientific basis of commercial genomic profiles used to assess health risks and personalize health interventions. *Am J Hum Genet* 82, 593–599.

Jones, S., Hruban, R.H., Kamiyama, M., Borges, M., Zhang, X., Parsons, D., Lin, J., Palmisano, E., Brune, K., Jaffee, E. et al. (2009). Exonic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. *Science* 324, 217.

Kohane, I. and Altman, R. (2005). Health-information altruists – A potentially critical resource. *N Engl J Med* 353, 2074–2077.

Kraft, P. and Hunter, D. (2009). Genetic risk prédiction – Are we there yet? *N Engl J Med* 360, 1701–1703.

Lequerré, T., Gauthier-Jauneau, A.-C., Bansard, C., Derambure, C., Hiron, M., Vittecoq, O., Daveau, M., Mejjad, O., Daragon, A., Tron, F. et al. (2006). Gene profiling in white blood cells predicts infliximab responsiveness in rheumatoid arthritis. *Arthritis Res Ther* 8, R105.

Levy, S., Sutton, G., Ng, P., Feuk, L., Halpern, A., Walenz, B., Axelrod, N., Huang, J., Kirkness, E., Denisov, G. et al. (2007). The diploid genome sequence of an individual human. *PLoS Biol* 5, e254.

Ley, R.E., Turnbaugh, P.J., Klein, S. and Gordon, J.I. (2006). Human gut microbes associated with obesity. *Nature* 444, 1022–1023.

Li, M., Wang, B., Zhang, M., Rantalainen, M., Wang, S., Zhou, H., Zhang, Y., Shen, J., Pang, X., Zhang, M. et al. (2008). Symbiotic gut microbes modulate human metabolic phenotypes. *Proc Natl Acad Sci USA* 105, 2117–2122.

Lin, Z., Owen, A. and Altman, R. (2004). Genomic research and human subject privacy. *Science* 305, 183.

Lodovici, M., Luceri, C., Filippo, C., Romualdi, C., Bambi, F. and Dolara, P. (2007). Smokers and passive smokers gene expression profiles: Correlation with DNA oxidative damage. *Free Rad Biol Med* 43, 415–422.

Long, W.H., Ziao, H., Gu, X., Zhang, Q., Yang, H., Zhao, G. and Liu, J.H. (2004). A universal microarray for detection of SARS coronavirus. *J Virol Meth* 121, 57–63.

Manolio, T.A., Brooks, L.D. and Collins, F.S. (2008). A HapMap harvest of insights into the genetics of common disease. *J Clin Invest* 118, 1590–1605.

Mardis, E.R. (2008). Next-generation DNA sequencing methods. *Ann Rev Genomics Hum Genet* 9, 387–402.

McGuire, A. and Gibbs, R. (2006). No longer de-identified. *Science* 312, 370–371.

McGuire, A., Cho, M., McGuire, S. and Caulfield, T. (2007). The future of personal genomics. *Science* 317, 1687.

McGuire, A., Caulfield, T. and Cho, M.K. (2008). Research ethics and the challenge of whole-genome sequencing. *Nat Rev Genet* 9, 152–156.

Meadows, S.K., Dressman, H.K., Muramoto, G.G., Himburg, H., Salter, A., Wei, Z., Ginsburg, G.S., Chao, N.J., Nevins, J.R. and Chute, J.P. (2008). Gene expression signatures of radiation response are specific, durable and accurate in mice and humans. *PLoS ONE* 3, e1912.

Metcalfe, S., Hurworth, R., Newstead, J. and Robins, R. (2002). Needs assessment study of genetics education for general practitioners in Australia. *Genet Med* 4, 71–77.

Nussbaum, R., McInnes, R. and Willard, H.F. (2007). *Genetics in Medicine*, *7th edition*. WB Saunders Co., Philadelphia.

Online Mendelian Inheritance in Man. (2009). http://www.ncbi.nlm.nih.gov/sites/entrez?db=OMIM&itool=toolbar

Palmer, C., Bik, E.M., DiGiulio, D., Rleman, D.A., and Brown, P.O. (2007). Development of the human infant intestinal microbiota. *PLoS Biol* 5, e177.

Peltonen, L. and McKusick, V. (2001). Dissecting human disease in the postgenomic era. *Science* 291, 1224–1229.

Potti, A., Mukherjee, S., Petersen, R., Dressman, H., Bild, A., Koontz, J., Kratzke, R., Watson, M., Kelley, M., Ginsburg, G. et al. (2006a). A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. *N Engl J Med* 355, 570–580.

Potti, A., Dressman, H.K., Bild, A., Riedel, R.F., Chan, G., Sayer, R., Cragun, J., Cottrill, H., Kelley, M.J., Petersen, R. et al. (2006b). Genomic signatures to guide the use of chemotherapeutics. *Nat Med* 12, 1294–1300.

Personalized Medicine. (2007). http://olpa.od.nih.gov/tracking/110/senate_bills/session1/s-976.asp

Pritchard, J. (2001). Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet* 69, 124–137.

Romeo, S., Pennacchio, L., Fu, Y., Boerwinkle, E., Tybjaerg-Hansen, A., Hobbs, H. and Cohen, J. (2007). Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. *Nat Genet* 39, 513–516.

Rus, V., Chenb, H., Zernetkinaa, V., Magderb, L.S., Mathaia, S., Hochberg, M. and Via, C. (2004). Gene expression profiling in peripheral blood mononuclear cells from lupus patients with active and inactive disease. *Clin Immunol* 112, 231–234.

Securities and Exchange Commission. (2009). http://www.sec.gov/Archives/edgar/data/1131324/000095013409005338/f51197e10vk.htm

Scheuner, M., Sieverding, P. and Shekelle. P. (2008). Delivery of genomic medicine for common chronic adult diseases: a systematic review. *J Am Med Assoc* 299, 1320–1334.

Service, R. (2006). The race for the $1000 genome. *Science* 311, 1544–1546.

Shou, J., Bull, C., Li, L., Qian, H., Wei, T., Luo, S., Perkins, D., Solenberg, P., Tan, S., Chen, X. et al. (2006). Identification of blood biomarkers of rheumatoid arthritis by transcript profiling of peripheral blood mononuclear cells from the rat collagen-induced arthritis model. *Arthritis Res Ther* 8, R28.

Singh, M.K., Scott, T.F., LaFramboise, W.A., Hu, F.Z., Post, J.C. and Ehrlich, G. (2007). Gene expression changes in peripheral blood mononuclear cells from multiple sclerosis patients undergoing α-interferon therapy. *J Neurological Sciences* 258, 52–59.

Sjoblom, T., Jones, S., Wood, L., Parsons, D., Lin, J., Barber, T., Mandelker, D., Leary, R., Ptak, J., Silliman, N. et al. (2006). The consensus coding sequences of human breast and colorectal cancers. *Science* 314, 268–274.

Sparano, J.A. (2006). TAILORx: Trial assigning individualized options for treatment (Rx). *Clin Breast Cancer* 4, 347–350.

Stallings, S., Witt, W., Finkelstein, S., Hiller, A., Sinskey, A. and Ginsburg, G. (2006). A framework to evaluate the economic impact of pharmacogenomics. *Pharmacogenomics* 7, 853–862.

Staudt, L.M. (2003). Molecular diagnosis of the hematologic cancers. *N Engl J Med* 348, 1777–1785.

Stratton, M.R., Campbell, P. and Futreal, P.A. (2009). The cancer genome. *Nature* 458, 719–724.

Tarpey, P., Smith, R., Pleasance, E., Whibley, A., Edkins, S., Hardy, C., O'Meara, S., Latimer, C., Dicks, E., Menzies, A. et al. (2009). A systematic, large-scale resequencing screen of X chromosome coding exons in mental retardation. *Nat genet* 41, 535–543.

Thorleifsson, G., Magnusson, K., Sulem, P., Bragi Walters, G., Gudbjartsson, D., Stefansson, H., Jonsson, T., Jonasdottir, A., Jonasdottir, A. and Stefansdottir, G. (2007). Common sequence variants in the LOXL1 gene confer susceptibility to exfoliation glaucoma. *Science* 317, 1397–1400.

Topol, E. and Frazer, K. (2007). The resequencing imperative. *Nature Genet* 39, 439–440.

Turnbaugh, P.J., Ley, R.E., Mahowald, M.A, Magrini, V., Mardis, E. and Gordon J.I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 444, 1027–1031.

Valk, P., Verhaak, R., Beijen, M., Erpelinck, C., Barjesteh vWvD-K.S., Boer, J., Beverloo, H., Moorhouse, M., van der Spek, P., Lowenberg, B. et al. (2004). Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med* 350, 1617–1628.

Wang, D., Urisman, A., Liu, Y.T., Springer, M., Ksiazek, T.G., Erdman, D., Mardis, E., Hickenbotham, M., Magrini, V. and Eldred, J. (2003). Viral discovery and sequence recovery using DNA microarrays. *PLoS Biol* 1, e2.

West, M., Ginsburg, G., Huang, A. and Nevins, J. (2006). Embracing the complexity of genomic data for personalized medicine. *Genome Res* 16, 559–566.

Wheeler, D., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y., Makhijani, V., Roth, G. et al. (2008). The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452, 872–876.

Willard, H.F., Angrist, M. and Ginsburg, G. (2005). Genomic medicine: Genetic variation and its impact on the future of health care. *Philos Trans R Soc Lond B Biol Sci* 360, 1543–1550.

Wolinsky, H. (2007). The thousand-dollar genome: Genetic brinksmanship or personalized medicine? *EMBO Reports* 8, 900–903.

Wu, M., Chiou, H., Ho, I., Chen, C. and Lee, T. (2003). Gene expression of inflammatory molecules in circulating lymphocytes from arsenic-exposed human subjects. *Environ Health Perspect* 111, 1429–1438.

Basics

Section 1

2. Organization, Variation and Expression of the Human Genome

3. DNA Sequencing for the Detection of Human Genome Variation

4. Genome-Wide Association Studies and Genotyping Technologies

5. Copy Number Variation and Human Health

6. DNA Methylation Analysis: Providing New Insight into Human Disease

7. DNA Microarrays in Biological Discovery and Patient Care

8. Proteomics: The Deciphering of the Functional Genome

9. Comprehensive Metabolic Analysis for Understanding of Disease

This page intentionally left blank

CHAPTER

Organization, Variation and Expression of the Human Genome

Huntington F. Willard

INTRODUCTION

The genetic variation that can influence health and disease has been a central, if not widely practiced, principle of medicine for over a hundred years, since the prescient observations of the British physician Sir Archibald Garrod established the concept of "chemical individuality" over a century ago (Garrod, 1902). What has limited broad application of this principle until now has been the generally presumed rarity or special nature of clinical circumstances or conditions to which genetic variation was relevant – rare disorders such as Garrod's alkaptonuria, inherited conditions limited to specific populations such as sickle cell anemia, or specialized situations such as the role of ABO incompatibility in blood transfusion. Now, however, with the availability of a "reference sequence" of the human genome, with emerging appreciation of the extent of genome variation among different individuals, and with a growing understanding of the role of common, not just rare, variation in disease, we are poised to begin to exploit the impact of that variation on human health on a broad scale, in the context of genomic and personalized medicine.

Variation in the human genome has long been the cornerstone of the field of human genetics, and its study led to the establishment of the medical specialty of medical genetics (Nussbaum et al., 2007). A crucial set of connections joining Mendel's principles of heredity, Garrod's concept of chemical individuality, the practice of medicine and the sequence of the genome came with Pauling's discovery of the molecular basis of sickle cell anemia and the direct correspondence between an individual's genetic make-up and the type of hemoglobins present in that individual's red cells (Pauling et al., 1949). The general nature and frequency of gene variants in the human genome became apparent with the classic work in the 1960s on the incidence of polymorphic protein variants in populations of healthy individuals (Harris and Hopkinson, 1972; Lewontin, 1967; reviewed in Harris, 1980).

Calculations based on those protein polymorphism data – now extended in a robust and comprehensive manner with the analysis of variation on a genome scale – lead to the inescapable conclusion that virtually every individual should be found to have his or her own unique constitution of gene products, the implications of which provide a conceptual foundation for what today we call "personalized medicine" as a new-age rediscovery of Garrod's "chemical individuality." Thus, with the availability of the human genome sequence (International Human Genome Sequencing Consortium, 2001, 2004; Venter et al., 2001) and determination of the extent of human genome variation, both within and among populations (International HapMap Consortium, 2003, 2007) and within individual genomes (Levy et al., 2007; Wheeler et al., 2008), awareness of widespread human variation can begin to be applied generally to the exploration of common human disease.

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

In this chapter, the organization, variation and expression of the human genome is presented as a foundation for the many chapters to follow on human genomics, on genome technology and informatics, on approaches in translational genomics and, finally, on the principles of genomic and personalized medicine as applied to specific diseases.

THE HUMAN GENOME

The typical human genome consists of approximately 3 billion (3×10^9) bp of DNA, divided among the 24 types of nuclear chromosomes (22 autosomes, plus the sex chromosomes, X and Y) and the much smaller mitochondrial chromosome (Tables 2.1, 2.2). The genome can be represented and evaluated in different ways, with different levels of resolution and degrees of sensitivity, depending on the clinical need (Figure 2.1).

Individual chromosomes can best be studied at metaphase in dividing cells, and karyotyping of patient chromosomes has been a valuable and routine clinical laboratory procedure for decades (Trask, 2002); various staining or hybridization-based analytical techniques have the ability to detect chromosome abnormalities ranging from an extra or missing whole chromosome (aneuploidy), to translocations or rearrangements involving just a portion of a chromosome(s), to deletions or duplications involving as little as perhaps a megabase (10^6 bp; Mb) of DNA.

More recent technologies involving overlapping sets (called "tiling paths") of isolated segments of the genome arrayed on microscope slides have provided vastly improved resolution and precision capable of evaluating in a rapid and comprehensive way the proper dosage (and in some cases the organization) of the corresponding DNA segments within an individual's genome (Figure 2.1b) (see Chapter 5). The ultimate resolution, of course, comes from direct sequence analysis, and a number of new technologies have reduced the cost and improved the throughput of sequencing individual genomes, facilitating comparisons with the reference human genome sequence and enabling medical resequencing of patient samples (see later section in this chapter) to search for novel variants or mutations that might be of clinical importance (Bentley, 2006) (Figure 2.1c, d) (see Chapter 3).

Genes in the Human Genome

While the human genome contains a currently estimated 20,000–25,000 genes (Clamp et al., 2007; International Human Genome Sequencing Consortium, 2004), the coding segments of those genes comprise less than 2% of the genome; as represented in Figure 2.1c, most of the genome, therefore, consists of DNA that lies between genes, far from genes or in vast areas spanning several Mb that appear to contain no genes at all ("gene deserts"). A *caveat* for this statement is that the process of gene identification and genome annotation remains very much a work-in-progress; despite the apparent robustness of recent estimates (Clamp et al., 2007), it is conceivable that

TABLE 2.1 Characteristics of the human genome[a]

Length of the human genome (basepairs)	3,253,037,807
Number of known protein-coding genes	21,541
Average gene density (genes/Mb)[b]	6.6
Number of non-coding RNA genes	4421
Number of SNPs[b]	13,022,900

[a] From Ensembl v. 48 (accessed February 2008).
[b] Mb = megabasepairs; SNP = single nucleotide polymorphism.

TABLE 2.2 Variation among human chromosomes[a]

Chromosome	Mb	Protein-coding genes	Genes/Mb	miRNA genes
1	247.25	2153	8.7	68
2	242.95	1315	5.4	60
3	199.50	1105	5.5	57
4	191.27	786	4.1	42
5	180.86	894	4.9	46
6	170.90	1109	6.5	36
7	158.82	1008	6.3	43
8	146.27	743	5.1	38
9	140.27	904	6.4	40
10	135.37	819	6.1	35
11	134.45	1368	10.2	37
12	132.35	1069	8.1	43
13	114.14	356	3.1	23
14	106.37	662	6.2	62
15	100.34	634	6.3	21
16	88.83	902	10.2	20
17	78.77	1217	15.5	40
18	76.12	289	3.8	15
19	63.81	1427	22.4	82
20	62.44	603	9.7	28
21	46.94	283	6.0	10
22	49.69	508	10.2	18
X	154.91	874	5.6	97
Y	57.77	80	1.4	3
Mitochondrial	0.016	22	–	–

[a]From Ensembl v. 48.

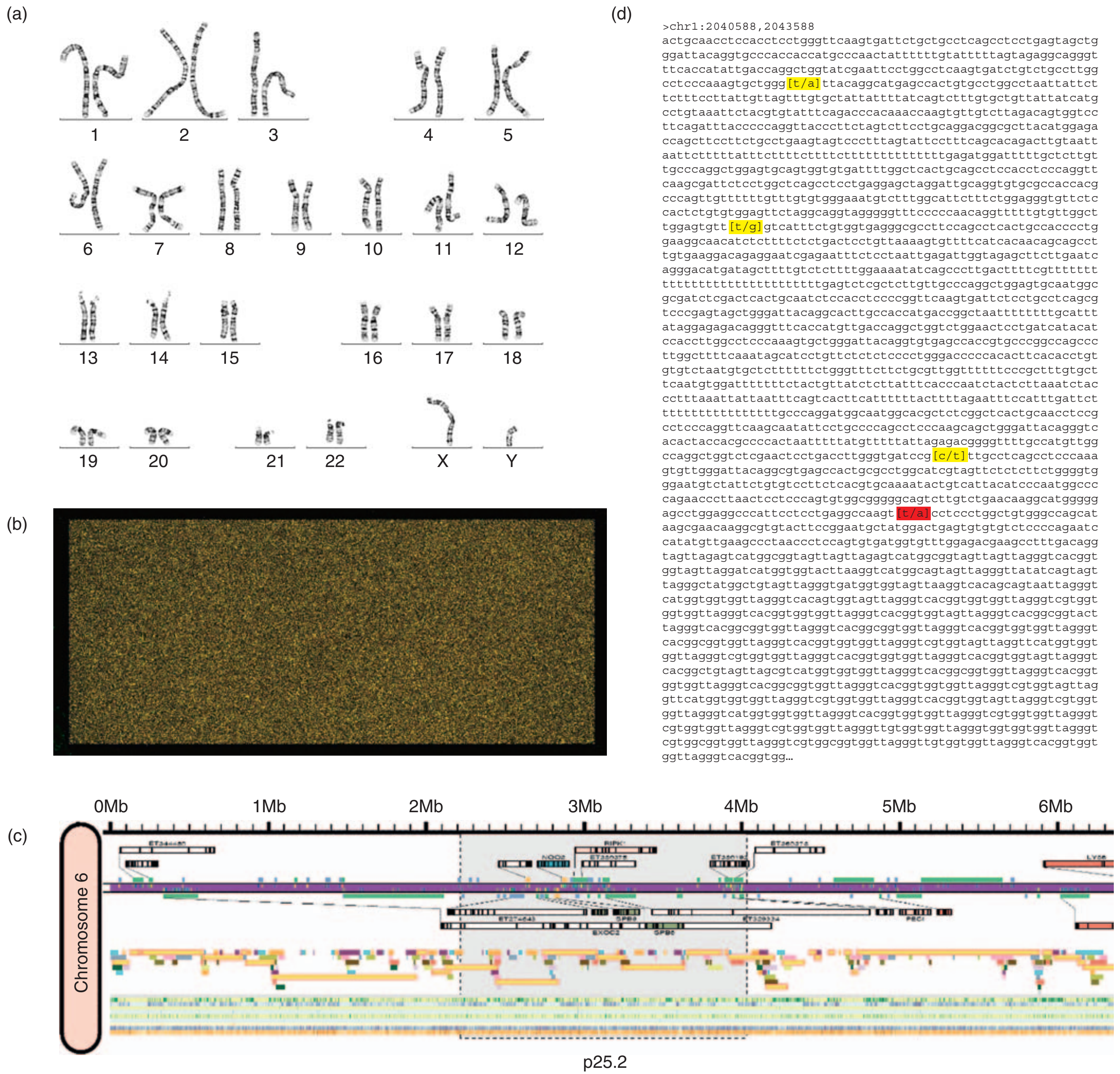


Figure 2.1 Four views of the human genome. (a) Karyotype of a normal male donor, HuRef, whose genome was the first individual diploid genome to be sequenced (Levy et al., 2007). The 24 types of human chromosome are shown after conventional G-banding – 22 pairs of autosomes and the two sex chromosomes, X and Y. (b) An array of genomic segments, showing 244,000 genomic elements hybridized to DNA from HuRef. (c) Schematic representation of the content of ~6 Mb from the short arm of chromosome 6, including the location of various genes and other features from the HuRef genome. (d) DNA sequence from the genome of James D. Watson, showing 3000 bp from chromosome 1. Watson's sequence is heterozygous at four positions, three (in yellow) that are known polymorphisms in various populations and one (in red) that is a novel variant. Figures in (a) and (b) were provided courtesy of Steve Scherer, Hospital for Sick Children, Toronto, Canada. (c) is part of a large poster representing the complete diploid HuRef genome (Levy et al., 2007).

there are some genes, including clinically relevant genes, that are currently undetected or that display characteristics that we do not currently recognize as being associated with genes. A maximum of 5% of the genome consists of DNA that has been quite well conserved through evolution, one indication of an important function. These and other considerations have led to the estimate that at most 20% of the genome is of functional importance (Pheasant and Mattick, 2007). Nonetheless, the statement that the vast majority of the genome consists of spans of DNA that are non-genic, of no obvious function, and of uncertain clinical relevance remains true.

In addition to being relatively sparse in the genome, genes are distributed quite non-randomly along the different human chromosomes. Some chromosomes are relatively gene-rich, while others are quite gene-poor, ranging from a high of ~22 genes/Mb to a low of ~3 genes/Mb (excluding the Y chromosome and the mitochrondrial chromosome) (Table 2.2). And even within a chromosome, genes tend to cluster in certain regions or in particular bands, a point of clear clinical significance when evaluating genome integrity, dosage or arrangement in different patient samples.

Coding and Non-Coding Genes

There are a number of different types of gene in the human genome. Most genes are protein-coding and are transcribed into messenger RNAs (mRNAs) that are ultimately translated into their respective proteins; their products comprise the list of enzymes, structural proteins, receptors and regulatory proteins that are found in various human tissues and cell types. However, there are additional genes whose functional product appears to be the RNA itself. These so-called non-coding RNAs (ncRNAs) have a range of functions in the cell, and some do not as yet have any identified function. But the genes whose transcripts make up the collection of ncRNAs represent about a sixth of all identified human genes (Table 2.1).

Some of the types of ncRNA play largely generic roles in cellular infrastructure, including transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs) involved in translation of mRNAs on ribosomes, spliceosomal RNAs involved in control of RNA splicing, and small nucleolar RNAs (snoRNAs) involved in modifying rRNAs (Griffiths-Jones, 2007; Mattick and Makunini, 2006). Other ncRNAs play roles in gene regulation, for example in epigenetic gene silencing (Ogawa and Lee, 2002).

A class of small RNAs of growing importance are the microRNAs, ncRNAs of only ~22 bases in length that suppress translation of target genes by binding to their respective mRNAs and thus regulate protein production from the target transcript(s) (Filipowicz et al., 2008). Some 255 microRNA genes were identified in the human genome initially (Lim et al., 2003), although the total number of such genes is now thought to be closer to 1000 (Bentwich, 2005; Griffiths-Jones, 2007) (Table 2.2). Some are evolutionarily conserved, while others appear to be of quite recent origin during primate evolution, thus underscoring the difficulty of determining the precise number and identity of human genes (Clamp et al., 2007). MicroRNAs have been shown to downregulate hundreds of mRNAs each, with different combinations of target RNAs in different tissues (Lim et al., 2005); combined, the microRNAs are thus predicted to control the activity of as many as 30% of all protein-coding genes in the genome (Filipowicz et al., 2008). While this is a fast-moving area of genome biology, several microRNAs have already been implicated in various human diseases, including cancer, developmental disorders and heart disease (Chang and Mendell, 2007; van Rooij et al., 2008).

Genome Composition and Landscape

As observed earlier, the distribution of genes in the genome is non-random, both within and between chromosomes. This in part is a reflection of the distribution of different types of DNA sequence, as the genome is partitioned into domains spanning hundreds of kilobasepairs to megabases, reflecting large-scale variation in the G+C content of DNA. These so-called "isochores" have been known for decades and, at a very gross level, mimic the pattern of light- and dark-staining bands that one observes in metaphase chromosomes (e.g., Figure 2.1a, c) (Eyre-Walker and Hurst, 2001). While the driving force behind the evolution of isochores is not clear, they influence the G+C content of genes contained within them (and, by virtue of the genetic code, therefore, the amino acid composition of the encoded proteins), the patterns of mutation and polymorphism detected, and the nature of various families of repeated DNA that reside there. Further – and most strikingly – different isochore domains contain clusters of genes that are highly or weakly expressed in a coordinated manner in different tissues (Caron et al., 2001; Gierman et al., 2007; Hurst et al., 2004). Thus, isochores reflect both the functional as well as structural organization of the genome. (See later section on "Expression of the Human Genome" for further discussion.)

VARIATION IN THE HUMAN GENOME

With completion of the reference human genome sequence, attention turned to the discovery and cataloging of variation in that sequence among different individuals (including both healthy individuals and those with various diseases) and among different populations. It has been estimated that there are some 10–15 million common sequence variants that are of sufficient frequency (minor allele frequency >5%) in one or more populations to be considered polymorphic in our species. In addition, there are countless very rare variants, many of which probably exist in only a single or a few individuals. In fact, given the number of individuals in our species, essentially each and every base pair in the human genome is expected to vary in someone somewhere around the globe. It is for this reason that the original genome sequence is considered a "reference" sequence, derived as a consensus of the limited number of individual

Table 2.3 Common variation in the human genome

Type of variation	Size range (approx.)[a]	Effect(s) in biology and medicine
Single nucleotide polymorphisms (SNPs)	1 bp	Non-synonymous ⟶ functional change in encoded protein? Others ⟶ potential regulatory variants? Most ⟶ no effect? ("neutral")
Copy number variants (CNVs)	10 kb to 1 Mb	Gene dosage variation ⟶ functional consequences? Most ⟶ no effect or uncertain effect
Insertion/deletion polymorphisms (in/dels)	1 bp to 1 Mb	In coding sequence: frameshift mutation? ⟶ functional change Most ⟶ uncertain effects
Inversions	Few bp to 100 kb	? break in gene sequence ? long-range effect on gene expression ? indirect effects on reproductive fitness Most ⟶ no effect? ("neutral")
Segmental duplications	10 kb to >1 Mb	Hotspots for recombination ⟶ polymorphism (CNVs)

[a]Abbreviations: bp = basepair; kb = kilobasepair; Mb = megabasepair.

genomes whose sequencing was part of the Human Genome Project, but actually identical to no individual's genome.

Types of Variation

Early estimates were that any two randomly selected individuals have sequences that are 99.9% identical or, put another way, that an individual genome would be heterozygous at approximately 3–5 million positions, with different bases (i.e., a T or a G) at the maternally and paternally inherited copies of that particular sequence position. The majority of these differences involve simply a single unit in the DNA code and are referred to as single nucleotide polymorphisms (SNPs) (Table 2.1) (see Chapter 3). The remaining variation consists of insertions or deletions (in/dels) of (usually) short sequence stretches, variation in the number of copies of repeated elements or inversions in the order of sequences at a particular locus in the genome (Figure 2.2). The total amount of in/del variation is more than originally anticipated and approaches 0.5%, not 0.1%, between any two randomly selected individuals (Levy et al., 2007). Any and all of these types of variation can influence disease and thus must be accounted for in any attempt to understand the contribution of genetics to human health (Table 2.3).

While the overall estimate of SNP heterozygosity is approximately 1 in 1500 bp, there is much more variation in non-coding sequences than in the coding segments of genes, reflecting strong selective pressure during evolution against certain types of change in gene sequences. The combination of particular alleles along chromosomes is also non-random, with particular combinations (haplotypes) being more prevalent over short distances, due to the relative inefficiency of meiotic recombination to separate alleles at sites that are physically close together (International HapMap Consortium, 2007; Nussbaum et al., 2007). The resulting patterns of linkage disequilibrium are relevant for designing strategies to examine genetic variation genome-wide, both as a practical matter (i.e., reducing the number of SNPs that need to be tested to reveal the underlying patterns of variation) and for evaluating the potential functional importance of any particular SNP allele (see Chapters 4 and 15).

Copy Number Variation

Over the past few years, a number of important studies have identified a previously unanticipated prevalence of structural variants in the genome, which collectively account for more variation in genome sequence than do SNPs (e.g., Levy et al., 2007; Redon et al., 2006; Sebat et al., 2004; Tuzun et al., 2005). The most common type of structural variation involves changes in the local copy number of sequences (including genes) in the genome, and these are generally referred to as copy number variants (CNVs) (Figure 2.2) (see Chapter 5).

A number of different technology platforms are now being used to detect CNVs, including arrays and direct genome sequencing (Korbel et al., 2007; Levy et al., 2007; Wong et al., 2007). As many such CNVs encompass genes (including microRNA genes; Wong et al., 2007) and as a significant number of novel CNVs are uncovered with every new population studied, a dedicated effort is underway to cataloged CNVs in the human genome worldwide and to associate these with clinical phenotype (Feuk et al., 2006; Scherer et al., 2007; Sharp et al., 2006). While most variation of this type is inherited, some CNVs occur *de novo* or even in somatic cells; in these cases, an individual will have different repeat lengths than do either of his or her parents.

Array-based methods (see Chapter 5) have rapidly gained acceptance for evaluating the association of both inherited and *de novo* CNVs with mental retardation and other developmental disorders (de Vries et al., 2005; Friedman et al., 2006; Lee et al., 2007; Weiss et al., 2008). It is of considerable ongoing interest to

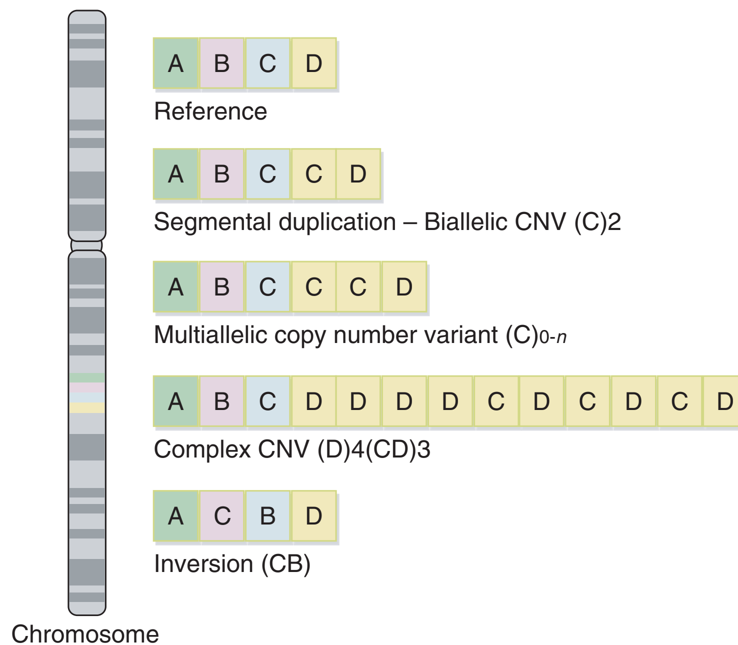


Figure 2.2 Schematic representation of different types of structural polymorphism in the human genome, leading to deletions, duplications, inversions and CNV changes relative to the reference arrangement. From Estivill and Armengol (2007), with permission.

evaluate the role of CNVs and other structural variants including deletions (Conrad et al., 2006) and inversions (Korbel et al., 2007; Stefansson et al., 2005; Tuzun et al., 2005) in the etiology of more common, complex diseases or traits of adulthood, including neurological and psychiatric conditions as well as pharmacogenetic traits (Beckmann et al., 2007; Buckland, 2003).

Variation in a Single Genome

The most extensive current inventory of the amount and type of variation to be expected in any given genome comes from the direct analysis of the diploid genome sequence of a single male individual, HuRef (Levy et al., 2007). Over 4 million variants were described, spanning some 12.3 Mb of DNA. About 20 Mb of "new" sequence was determined that was not previously available as part of the human reference sequence, reflecting in part the still unfinished nature of the human genome sequence and in part the particular patterns of inserted or deleted sequences that distinguish different genomes. Several hundred thousand in/dels were also found in this single genome. In addition, several hundred CNVs were detected, which overlapped at least 95 well-annotated genes. While most of these variants are identical to those found in other individuals in the population, others are likely to be what are termed "private" mutations, specific to HuRef and his family.

In the HuRef genome, at least 850 genes known to be involved in inherited disease contained at least one heterozygous variant, and over 300 of them contained at least one non-synonymous SNP (i.e., a SNP that, by virtue of the genetic code, is predicted to change the encoded amino acid). Of course, additional genes may also impact disease, and, overall, more than 4000 genes in the HuRef genome contained one or more non-synonymous SNP. Thus, at least 17% and perhaps as many as 44% of the genes in the HuRef genome were heterozygous and could encode proteins that differ in their amino acid sequence and/or are produced in different amounts (Levy et al., 2007). These estimates underscore the impact of gene and genome variation on human biology and on medicine. They also provide remarkable validation of the original estimates of Harris and Lewontin decades ago of the proportion of genes that are heterozygous in any given individual (Harris, 1980; Lewontin, 1967).

Table 2.3 and Figure 2.2 capture the general types of and characteristics of the most common variation in the human genome and in human genes. However, it is clear that we are still in a mode of discovery, as relatively few genomes or populations have been assessed to date; no doubt millions of additional SNPs remain to be uncovered, as well as many additional in/dels, inversions and CNVs, a portion of which will be expected to involve genes and other sequences of direct relevance to medicine. The issue of "what is normal?" – an essential concept in clinical medicine – remains very much an open question when it comes to the human genome (Shianna and Willard, 2006).

Variation in Populations

Most of the heterozygosity in the human genome is believed to be due to variants with a minor allele frequency of at least

1%. Taking advantage of major technological developments that have greatly increased the throughput of genotyping on a genome-wide scale, several large-scale projects have validated these estimates by gathering genotypic information on millions of SNPs worldwide (Hinds et al., 2005; International HapMap Consortium, 2003, 2007). Most of the studies to date, however, have been restricted to a small number of populations of Northern European, African and Asian origin used for SNP detection. From these and a large number of earlier studies that examined more populations but for many fewer variants, it has been concluded that some 85–90% of the variation found in our species is shared among different population groups; a relative minority of variants, therefore, are specific to or highly enriched/depleted in genomes from a particular population.

It is possible to use population-specific variants to obtain information on the geographic origin of a genome or of particular segments within a genome. Given the many millions of SNPs now available, there are at least hundreds of thousands of SNPs that are informative for such studies (so-called "ancestry informative markers," AIMs) (Kittles and Weiss, 2003; Paschou et al., 2007; Tian et al., 2008). This had led to two related, but distinct applications of such markers. First is the use of admixture mapping, tracing the location of particular SNPs associated with disease in populations of patients whose genomes are a mixture from at least two original populations, for example, African-Americans or Latinos (Price et al., 2007; Smith et al., 2004). Such an approach has already been used to map genes associated with several phenotypes whose frequency differs markedly between different population groups, including prostate cancer (Freedman et al., 2006), hypertension (Deo et al., 2007), skin pigmentation (McEvoy et al., 2006) and white blood cell count (Nalls et al., 2008).

The second use of AIMs is for ancestry testing unrelated to disease studies (Kittles and Weiss, 2003; Shriver and Kittles, 2004). While the motivations behind such testing and the potential uses (and, some fear, abuses) of biogeographic information are varied, the commercial availability and interpretation of genetic ancestry testing is controversial (Bolnick et al., 2007). Nonetheless, the availability of such information as an intentional or unwitting by-product of wide-scale genome analysis is inevitable, and both consumers/patients and health professionals need to be aware of this as genetic variation is explored in the context of individual genomes (see Chapter 18).

EXPRESSION OF THE HUMAN GENOME

A key question in exploring the origins, structure and function of the human genome is to understand how proper expression of our 20,000–25,000 genes is determined, how it can be influenced by either genetic variation or by environmental exposures or inputs, and by what mechanisms such alterations in gene expression can lead to pathology evident in the practice of clinical medicine. The control of gene activity – in development, in different tissues, during the cell cycle, and during the lifetime of an individual both in sickness and in health – is determined by a complex interplay of genetic and epigenetic features.

By "genetic" features, we here refer to those found in the genome sequence, which plays a role, of course, in determining the identity of each gene, its particular form (alleles), its level of expression (regulatory elements such as promoters, enhances, splice sites, etc.), and its particular genomic landscape (domains, isochores). By "epigenetic" features, here we mean packaging of the DNA into chromatin, in which it is complexed with a variety of histones as well as innumerable non-histone proteins that influence the accessibility and activity of genes and other genomic sequences. The structure of chromatin – unlike the genome sequence itself – is highly dynamic and underlies the control of gene expression that shapes in a profound way both cellular and organismal function (Felsenfeld and Groudine, 2003).

Genomic and Epigenomic Aspects of Gene Expression

Identifying the genomic sequences and features that direct spatial and temporal aspects of gene expression remains a formidable challenge in genome annotation. While several decades of work in molecular biology have defined critical regulatory elements such as promoters and enhancers for many individual genes, it is a tall order to perform such studies on a genome-wide scale.

At the genetic level, such studies have depended on identifying highly conserved elements, whose demonstrated sequence conservation over, in some cases, hundreds of millions of years provides *prima facie* evidence of their importance (Bejerano et al., 2004; Margulies et al., 2003). Several such ultraconserved elements have been validated *in vitro* as gene enhancers, providing some confidence that others identified in this way will also play a role in regulating gene expression (Pennacchio et al., 2006). However, by definition, such approaches will overlook regulatory elements that are not well conserved at the sequence level and/or that are newly evolved on the human lineage.

A complementary epigenetic approach has been to explore the characteristics of chromatin that are associated with active or repressed genes as a step towards identifying the transcriptional regulatory code for the human genome (Barrera and Ren, 2006; Bernstein et al., 2007). Such studies, largely employing the method of chromatin immunoprecipitation followed by array or sequence analysis (Hawkins and Ren, 2006), have uncovered predictive chromatin "signatures" for promoters and enhancers in the human genome (Heintzman et al., 2007; Kim et al., 2005). These analyses are part of a broad effort to explore epigenetic patterns in chromatin genome-wide to better understand control of gene expression in different tissues or disease states (ENCODE Project Consortium, 2007; Brena et al., 2006).

Increasing evidence points to a role for epigenetic changes in human disease in response to environmental influences

(Feinberg, 2007). The dynamic nature of epigenetic regulation – changes in DNA methylation or in histone modification over time, between different tissues, or in various disease states – allows for what has been called "phenotypic plasticity," relevant both to the origins and potential treatment of disease (Feinberg, 2007) (see Chapter 6). That monozygotic twins (who share identical genomes, but are frequently discordant with respect to clinical phenotypes) show epigenetic differences in DNA methylation demonstrates the potential for epigenetic effects to extend or modify information contained in the genome and thus to underlie at least some phenotypic differences (Fraga et al., 2005).

Allelic Imbalance

Other than well-known examples of monoallelic gene expression where only one of the two copies of a gene in a diploid cell is expressed (such as genomic imprinting and X chromosome inactivation; Nussbaum et al., 2007), it is commonly assumed that the vast majority of autosomal genes in the genome are expressed from both homologs at comparable levels. Recent evidence, however, using expressed SNPs to distinguish transcripts from the two alleles, has demonstrated widespread differential allelic expression for up to 20% of genes in the genome (Serre et al., 2008). This phenomenon, whose epigenetic or genetic basis is unknown, may be related to similar variability observed among X-linked genes in heterozygous females (Carrel and Willard, 2005). Further complexity is suggested by the finding of monoallelic gene expression, random with respect to parental origin, for approximately 5% of genes tested (Gimelbrant et al., 2007). This is an area of ongoing study and further implicates interactions between the genome and epigenetic control mechanisms as a determinant of gene expression in the human genome.

Genetic Control of Gene Expression Levels

It has been appreciated for decades that there is high variability in gene expression levels among individuals. A number of groups have begun to examine this variation as a complex quantitative trait, under genetic control (Cheung and Spielman, 2002). Several thousand loci exhibiting such variation have now been studied, and the factor(s) controlling the variation mapped around the genome (Cheung et al., 2005; Morley et al., 2004; Stranger et al., 2007a, b). A significant proportion of such effects map to the genes themselves, a result consistent with local sequence variation influencing the expression of such genes; many such local regulatory variants map to the promoter region of the gene or to the 3′ untranslated region (Cheung et al., 2005). Most of the mapped variation could be accounted for by SNP variation, but nearly 20% was due to CNVs (Stranger et al., 2007a, b), underscoring the functional importance of this type of genome variation. It is likely (though unproven in specific instances) that the discovered regulatory variants will correlate with the patterns of epigenetic modifications described earlier (Heintzman et al., 2007; Kim et al., 2005) and with the finding of widespread allelic imbalance around the genome (Serre et al., 2008).

Other determinants of gene expression variation map not to the variable locus itself, but to another locus elsewhere in the genome. This implies a network of regulatory interactions among different gene products in the cell, a finding that lends itself to a systems biology approach to begin to better define such networks.

Some of the differences in gene expression also vary among populations. Where examined, the phenotypes are attributable to common genetic variants that are more common in particular populations than in others (Spielman et al., 2007; Stranger et al., 2007a, b). These findings, therefore, may be relevant to complex genetic diseases whose prevalence differs among populations.

GENOME SEQUENCING

As described in the previous chapter, the cost of sequencing whole genomes is falling rapidly, with a number of new technologies that will soon invite practical consideration of if, when and how to introduce genome sequencing into medical practice (Mardis, 2008). The highly publicized release of the genome sequences of Craig Venter ("HuRef") (Levy et al., 2007) and James Watson (Wheeler et al., 2008) have been followed by the full sequences of at least four other genomes (*see 2009 Update*), and numerous additional genomes are already well underway (Table 2.4).

Table 2.4 Individual sequenced human genomes (as of May 2009)

Individuals	Nature of genome	Method	Year completed/ anticipated	References
"RP11"[a]	Partial haploid	Sanger sequencing	2003	Osoegawa et al., 2001; International Human Genome Sequencing Consortium, 2001
J. Craig Venter	Complete diploid	Sanger sequencing	2007	Levy et al., 2007
James D. Watson	Complete diploid	454 pyrosequencing	2007	Wheeler et al., 2008
YanHuang researcher, male	Complete diploid	Illumina sequencing by synthesis	2007	Qiu and Hayden, 2008; Wang et al., 2008
YanHuang paying customer	Complete diploid	Illumina sequencing by synthesis	2008	Qiu and Hayden, 2008
Yoruba HapMap sample, male	Complete diploid	Illumina sequencing by synthesis	2008	Bentley et al., 2008
Knome customers	Complete diploid	Various next-generation technologies	2008 and beyond	
Individual with acute myeloid leukemia, female	Complete diploid (skin) and tumor	454 pyrosequencing	2008	Ley et al., 2008
Personal Genome Project; 10 individuals	"Exomes" (coding exons only)	Polony sequencing	2008	www.personalgenomes.org; Church, 2006
Korean individual (SJK), male	Complete diploid	Illumina sequencing by synthesis	2008	Ahn et al., 2009
Venter Institute; 10–50 genomes	Complete diploid	Various next-generation technologies	2009 and beyond	
ClinSeq 1000	Exons of 200–400 cardiovascular genes	To be determined	2009 or beyond	www.genome.gov/25521304
YanHuang 99	Complete diploid	Various next-generation technologies	2010	Xin, 2007
GATC 100	Complete diploid	Various next-generation technologies	2010	http://www.gatc-biotech.com/en/discover_gatc/research/Human_Genome_Sequencing_Service.php
1000 Genomes Project	Various[b]	Various next-generation technologies	2010	www.1000genomes.org; www.genome.gov/26524516; Hayden, 2008; Kaiser, 2008
Personal Genome Project; 500–100,000	Exomes	Polony sequencing	?	www.personalgenomes.org

Source: Courtesy of Misha Angrist, Ph.D., Duke Institute for Genome Sciences & Policy.
[a]An anonymous male from Buffalo, New York, whose genome was overrepresented in the publicly funded Human Genome Project.
[b]6 people fully sequenced at 20× coverage; 180 at 2×; and exons of 1000 genes in 1000 people.

2009 UPDATE

The past year has seen enormous progress in our understanding of the human genome, especially in three areas: genetic variation among different individuals, magnitude and role of structural variation in genetic disease, and widespread transcription of the human genome.

With now six individual human genomes having been sequenced, from at least four geographically and/or ethnically distinct populations (Northern Europe, China, Korea and Nigeria) (see Table 2.4), it is possible to get a precise view of the amount of genetic variation to be found between any two individuals. Each of the sequenced genomes has revealed approximately 3–4 million SNPs, of which at least 15–30% are considered "novel," having not been previously documented. It is thus clear that much human genetic variation remains to be

discovered. Consistent with previous analyses involving smaller data sets, the average (autosomal) heterozygosity is about 1 SNP per 1300 bp, although this figure is significantly higher (1 SNP per 1000 bp) in a Yoruba genome from Nigeria, reflecting the higher level of genome diversity expected in African populations (Bentley et al., 2008; Tishkoff et al., 2009). Relevant to consideration of potential phenotypic effects of SNP variation, each of the sequenced genomes contained about 4000–5000 SNPs in coding sequences that would be predicted to result in a nonconservative amino acid change in the encoded protein, as well as over a hundred premature stop codons (Bentley et al., 2008; Levy et al., 2007; Wheeler et al., 2008). These become attractive candidates for functional studies.

SNPs represent only one type of human genetic variation. As initially tallied by Levy et al. (2007), far more variation between genomes can be accounted for by structural variants (insertions, deletions, inversions and copy number variants [CNVs]) (Kidd et al., 2008; Levy et al., 2007). While the exact amount of DNA involved is a matter of ongoing investigation, a comparison of two "typical" genomes appears to suggest about 99.5% identity, reflecting a higher level of sequence variation than widely popularized in the years immediately following the initial human genome sequence.

Eichler and colleagues (Itsara et al., 2009) examined the distribution of large CNVs in the general population, studying the genomes of some 2500 individuals. They found variants greater than 500 kb in size in 5–10% of individuals, with variants >1 Mb in 1–2% of the population. Given the large number of subjects examined, they could provide robust estimates of the excess of CNVs found in patients with autism, mental retardation and schizophrenia (Need et al., 2009; Stefansson et al., 2008; Sebat et al., 2007; Walsh et al., 2008), highlighting the phenotypic importance of rare copy number variation in the human genome. The role of such CNVs in human disease is particularly apparent in neuropsychiatric conditions (Cook and Scherer, 2008).

In addition to studies of genetic variation, recent work has greatly increased understanding of the genome itself and its transcription. Long-standing estimates had indicated that about 2% of the genome was transcribed into genic segments. It is clear now, however, that there is substantially more transcription that previously appreciated (see main chapter). Two particular classes of noncoding RNA (ncRNA) have emerged that greatly challenge preconceptions about what "genes" are and how one evaluates what is "important" in the genome. Both small and long ncRNAs, especially in promoter regions of genes, are predominant in transcriptome studies (Affymetrix ENCODE Transcriptome Project, 2009; Carninci, 2009). More generally, on the basis of telltale chromatin signatures in otherwise evolutionarily nonconserved portions of the genome, over a thousand large ncRNAs up to 20 kb in size were discovered (Guttman et al., 2009). This class of ncRNAs, initially uncovered nearly 20 years ago in the case of the *H19* and *XIST* noncoding genes, has been speculated to be involved in transcriptional control.

Application of new sequencing platforms to transcription profiling (so-called "RNA seq") has permitted global views of transcription and RNA processing (Wang et al., 2009). This approach has provided particular insight into genome-wide patterns of alternative splicing, its abundance and tissue-specificity (Castle et al., 2008; Pan et al., 2008; Sultan et al., 2008). Lastly, in a proof-of-concept study, RNA seq was used to detect novel gene fusions in a series of cancer cell lines and tumors (Maher et al., 2009).

It is clear that the availability of next-generation sequencing capability is revolutionizing the study of the human genome, its variation, its expression and its perturbation in disease.

2009 UPDATE REFERENCES

Affymetrix ENCODE Transcriptome Project, Cold Spring Harbor Laboratory ENCODE Transcriptome Project (2009). Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs. *Nature* 457, 1028–1032.

Ahn, S.-M., Kim, T., Lee, S., Kim, D., Ghang, H., Kim, D., Kim, B., Kim, S., Kim, W. and Kim, C. The first Korean genome sequence and analysis: Full genome sequencing for a socio-ethnic group. *Genome Res* 2009; [Epub ahead of print].

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R. et al. (2008). Accurate whole genome sequencing using reversible terminator chemistry. *Nature* 456, 53–59.

Carninci, P. (2009). The long and short of RNAs. *Nature* 457, 974–975.

Castle, J., Zhang, C., Shah, J., Kulkarni, A., Kalsotra, A., Cooper, T. and Johnson, J. (2008). Expression of 24,426 human alternative splicing events and predicted *cis* regulation in 48 tissues and cell lines. *Nat Genet* 40, 1416–1425.

Cook, E. and Scherer, S.W. (2008). Copy-number variations associated with neuropsychiatric conditions. *Nature* 455, 919–923.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M., Feldser, D., Huarte, M., Zuk, O., Carey, B., Cassady, J. et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458, 223–227.

Itsara, A., Cooper, G., Baker, C., Girirajan, S., Li, J., Absher, D., Krauss, R., Myers, R., Ridker, P., Chasman, D. et al. (2009). Population analysis of large copy number variants and hotspots of human genetic disease. *Am J Hum Genet* 84, 148–161.

Kidd, J.M., Cooper, G.M., Donahue, W.F., Hayden, H.S., Sampas, N., Graves, T., Hansen, N., Teague, B., Alkan, C., Antonacci, F. et al. (2008). Mapping and sequencing of structural variation from eight human genomes. *Nature* 453, 56–64.

Ley, T.J., Mardis, E.R., Ding, L., Fulton, B., McLellan, M.D., Chen, K., Dooling, D., Dunford-Shore, B.H., McGrath, S., Hickenbotham, M. et al. (2008). DNA sequencing of a cytogenetically normal acute myeloid leukemia genome. *Nature* 456, 66–72.

Maher, C., Kumar-Sinha, C., Cao, X., Kalyana-Sundaram, S., Han, B., Jing, X., Sam, L., Barrette, T, Palanisamy, N. and Chinnaiyan, A. (2009). Transcriptome sequencing to detect gene fusions in cancer. *Nature* 458, 97–101.

Need, A., Ge, D., Weale, M., Maia, J., Feng, S., Heinzen, E., Shianna, K., Yoon, W., Kasperaiciute, D., Gennarelli, M. et al. (2009). A genome-wide investigation of SNPs and CNVs in schizophrenia. *PLoS Genet* 5, e1000373.

Pan, Q., Shai, O., Lee, L., Frey, B. and Blencowe, B. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 40, 1413–1415.

Sebat, J., Lakshmi, B., Malhotra, D., Toge, J., Lese-Martin, C., Walsh, T., Yamron, B., Yoon, S., Krasnitz, A., Kendall, J. et al. (2007). Strong association of *de novo* copy number mutations with autism. *Science* 316, 445–449.

Stefansson, H., Rujescu, D., Cichon, S., Pietilainen, O., Ingason, A., Steinberg, S., Fossdal, R., Sigurdsson, E., Signmundsson, T., Buizer-Voskamp, J. et al. (2008). Large recurrent microdeletions associated with schizophrenia. *Nature* 455, 232–236.

Sultan, M., Schultz, M., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A., Parkhomshuk, D. et al., (2008). A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* 321, 956–960.

Tishkoff, S., Reed, F., Friedlaender, F., Ehret, C., Ranciaro, A., Froment, A., Hirbo, J., Awomoyi, A., Bodo, J., Doumbo, O. et al. (2009). The genetic structure and history of Africans and African Americans. *Science* 324, 1035–1044.

Walsh, T., McClellan, J., McCarthy, S., Addington, A., Pierce, S., Cooper, G., Nord, A., Kusenda, M., Malhotra, D., Bhandari, A. et al. (2008). Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 320, 539–543.

Wang, J., Wang, W., Li, R., Li, Y., Tian, G., Goodman, L., Fan, W., Zhang, J., Li, J., Zhang, J. et al. (2008). The diploid genome sequence of an Asian individual. *Nature* 456, 60–65.

Wang, Z., Gerstein, M. and Snyder, M. (2009). RNA-seq: A revolutionary tool for transcriptomics. *Nat Rev Genet* 10, 57–63.

REFERENCES

Ahn, S.-M., Kim, T., Lee, S., Kim, D., Ghang, H., Kim, D., Kim, B., Kim, S., Kim, W. and Kim, C. The first Korean genome sequence and analysis: Full genome sequencing for a socio-ethnic group. *Genome Res* 2009; [Epub ahead of print].

Barrera, L. and Ren, B. (2006). The transcriptional regulatory code of eukaryotic cells – insights from genome-wide analysis of chromatin organization and transcription factor binding. *Curr Opin Cell Biol* 18, 291–298.

Beckmann, J., Estivill, X. and Antonarakis, S. (2007). Copy number variants and genetic traits: Closer to the resolution of phenotypic variability. *Nat Rev Genet* 8, 639–646.

Bejerano, G., Pheasant, M., Makunini, I., Stephen, S., Kent, W., Mattick, J. and Haussler, D. (2004). Ultraconserved elements in the human genome. *Science* 304, 1321–1325.

Bentley, D.R. (2006). Whole-genome resequencing. *Curr Opin Genet Deve* 16, 545–552.

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R. et al. (2008). Accurate whole genome sequencing using reversible terminator chemistry. *Nature* 456, 53–59.

Bentwich, I. (2005). Identification of hundreds of conserved and nonconserved human microRNAs. *Nat Genet* 37, 766–770.

Bernstein, B., Meissner, A. and Lander, E. (2007). The mammalian epigenome. *Cell* 128, 669–681.

Bolnick, D., Fullwiley, D., Duster, T., Cooper, R., Fujimura, J., Kahn, J., Kaufman, J., Marks, J., Morning, A., Nelson, A. et al. (2007). The science and business of genetic ancestry training. *Science* 318, 399–400.

Brena, R., Huang, T. and Plass, C. (2006). Toward a human epigenome. *Nat Genet* 38, 1359–1360.

Buckland, P. (2003). Polymorphically duplicated genes: Their relevance to phenotypic variation in humans. *Ann Med* 35, 308–315.

Caron, H., van Schaik, B., van der Mee, M., Baas, F., Riggins, G., van Sluis, P., Hermus, M.C., van Asperen, R., Boon, K., Voute, P.A. et al. (2001). The human transcriptome map: Clustering of highly expressed genes in chromosomal domains. *Science* 291, 1289–1292.

Carrel, L. and Willard, H.F. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* 434, 400–404.

Chang, T. and Mendell, J. (2007). microRNAs in vertebrate physiology and human disease. *Annu Rev Genomics Hum Genet* 8, 215–239.

Cheung, V. and Spielman, R. (2002). The genetics of variation in gene expression. *Nat Genet* (Suppl 32), 522–525.

Cheung, V., Spielman, R., Ewens, K., Weber, T., Morley, M. and Burdick, J. (2005). Mapping determinants of human gene expression by regional and genome-wide association. *Nature* 437, 1365–1369.

Church, G. (2006). Genomes for all. *Sci Am* 294, 46–54.

Clamp, M., Fry, B., Kamal, M., Xie, X., Cuff, J., Lin, M., Kellis, M., Lindblad-Toh, K. and Lander, E. (2007). Distinguishing protein-coding and noncoding genes in the human genome. *Proc Natl Acad Sci USA* 104, 19428–19433.

Conrad, D., Andrews, T., Carter, N., Hurles, M. and Pritchard, J. (2006). A high-resolution survey of deletion polymorphism in the human genome. *Nat Genet* 38, 75–81.

Deo, R., Patterson, N., Tandon, A., McDonald, G., Haiman, C., Ardlie, K., Henderson, B., Henderson, S. and Reich, D. (2007). A high-density admixture scan in 1,670 African Americans with hypertension. *PLoS Genet* 3, e196.

De Vries, B., Pfundt, R., Leisink, M., Koolen, D., Vissers, L., Janssen, I., van Reijmersdal, S., Nillesen, W., Huys, E., de Leeuw, N. et al. (2005). Diagnostic genome profiling in mental retardation. *Am J Hum Genet* 77, 606–616.

ENCODE Project Consortium (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799–816.

Estivill, X. and Armengol, L. (2007). Copy number variants and common disorders: Filling the gaps and exploring complexity in genome-wide association studies. *PLoS Genet* 3, e190.

Eyre-Walker, A. and Hurst, L. (2001). The evolution of isochores. *Nat Rev Genet* 2, 549–555.

Feinberg, A. (2007). Phenotype plasticity and the epigenetics of human disease. *Nature* 447, 433–440.

Felsenfeld, G. and Groudine, M. (2003). Controlling the double helix. *Nature* 421, 448–453.

Feuk, L., Marshall, C., Wintle, R. and Scherer, S. (2006). Structural variants: Changing the landscape of chromosomes and design of disease studies. *Hum Mol Genet* 15(Review Issue 1), 15–66.

Filipowicz, W., Bhattacharyya, S. and Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: Are the answers in sight?. *Nat Rev Genet* 9, 102–114.

Fraga, M., Ballestar, E., Paz, M., Ropero, S., Setien, F., Ballestar, M., Heine-Suner, D., Cigudosa, J., Urioste, M., Benitez, J. et al. (2005). Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci USA* 102, 10604–10609.

Freedman, M., Haiman, C., Patterson, N., McDonald, G., Tandon, A., Waliszewska, A., Penney, K., Steen, R., Ardlie, K., John, E. et al. (2006). Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proc Natl Acad Sci USA* 103, 14068–14073.

Friedman, J., Baross, A., Delaney, A., Ally, A., Arbour, L., Asano, J., Bailey, D., Barber, S., Birch, P., Brown-John, M. et al. (2006). Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation. *Am J Hum Genet* 79, 500–513.

Garrod, A. (1902). The incidence of alkaptonuria: A study in chemical individuality. *Mol Med* 2, 274–282. Reprint (1996)

Gierman, H., Indemans, M., Koster, J., Goetze, S., Seppen, J., Geerts, D., van Driel, R. and Versteeg, R. (2007). Domain-wide regulation of gene expression in the human genome. *Genome Res* 17, 1286–1295.

Gimelbrant, A., Hutchinson, J., Thompson, B. and Chess, A. (2007). Widespread monoallelic expression on human autosomes. *Science* 318, 1136–1140.

Griffiths-Jones, S. (2007). Annotating noncoding RNA genes. *Annu Rev Genomics Hum Genet* 8, 279–298.

Harris, H. (1980). *The Principles of Human Biochemical Genetics, 3rd edition*. Elsevier, Amsterdam.

Harris, H. and Hopkinson, D. (1972). Average heterozygosity per locus in man: An estimate based on the incidence of enzyme polymorphisms. *Ann Hum Genet* 36, 9–20.

Hawkins, D. and Ren, B. (2006). Genome-wide location analysis: Insights on transcriptional regulation. *Hum Mol Genet* 15(Review Issue 1), 1–7.

Hayden, E. (2008). International genome project launched. *Nature* 451, 378–379.

Heintzman, N., Stuart, R., Hon, G., Fu, Y., Ching, C., Hawkins, R., Barrera, L., Calcar, S., Qu, C., Ching, K. et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 39, 311–318.

Hinds, D., Stuve, L., Nilsen, G., Halperin, E., Eskin, E., Ballinger, D., Frazer, K. and Cox, D. (2005). Whole-genome patterns of common DNA variation in three human populations. *Science* 307, 1072–1079.

Hurst, L., Pal, C. and Lercher, M. (2004). The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet* 5, 299–310.

International HapMap Consortium (2003). The International HapMap Project. *Nature* 426, 789–796.

International HapMap Consortium (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449, 851–862.

International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.

International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. *Nature* 431, 931–945.

Kaiser, J. (2008). A plan to capture human diversity in 1000 genomes. *Science* 319, 395.

Kim, T., Barrera, L., Zheng, M., Qu, C., Singer, M., Richmond, T., Wu, Y., Green, R. and Ren, B. (2005). A high-resolution map of active promoters in the human genome. *Nature* 436, 876–880.

Kittles, R. and Weiss, K. (2003). Race, ancestry, and genes: Implications for defining disease risk. *Annu Rev Genomics Hum Genet* 4, 33–67.

Korbel, J., Urban, A., Affourtit, J., Godwin, B., Grubert, F., Simons, J., Kim, P., Palejev, D., Carriero, N., Du, L. et al. (2007). Paired-end mapping reveals extensive structural variation in the human genome. *Science* 318, 420–426.

Lee, C., Iafrate, A. and Brothman, A. (2007). Copy number variations and clinical cytogenetic diagnosis of constitutional disorders. *Nat Genet* Suppl. 39, 48–54.

Levy, S., Sutton, G., Ng, P., Feuk, L., Halpern, A., Walenz, B., Axelrod, N., Huang, J., Kirkness, E., Denisov, G. et al. (2007). The diploid genome sequence of an individual human. *PLoS Biol* 5(10), e254.

Lewontin, R. (1967). An estimate of average heterozygosity in man. *Am J Hum Genet* 19, 681–685.

Ley, T.J., Mardis, E.R., Ding, L., Fulton, B., McLellan, M.D., Chen, K., Dooling, D., Dunford-Shore, B.H., McGrath, S., Hickenbotham, M. et al. (2008). DNA sequencing of a cytogenetically normal acute myeloid leukemia genome. *Nature* 456, 66–72.

Lim, L., Glasner, M., Yekta, S., Burge, C. and Bartel, D. (2003). Vertebrate microRNA genes. *Science* 299, 1540.

Lim, L., Lau, N., Garrett-Engele, P., Grimson, A., Schelter, J., Castle, J., Bartel, D., Linsley, P. and Johnson, J. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 433, 769–773.

Mardis, E.R. (2008). Next-generation DNA sequencing methods. *Ann Rev Genomics Hum Genet* 9, 387–402.

Margulies, E., Blanchette, M., Haussler, D. and Green, E.NISC Comparative Sequencing Program (2003). Identification and characterization of multi-species conserved sequences. *Genome Res* 13, 2507–2518.

Mattick, J. and Makunini, I. (2006). Non-coding RNA. *Hum Mol Genet* 15(Review Issue 1), 17–29.

McEvoy, B., Beleza, S. and Shriver, M. (2006). The genetic architecture of normal variation in human pigmentation: An evolutionary perspective and model. *Hum Mol Genet* Suppl. 15, 176–181.

Morley, M., Molony, C., Weber, T., Devlin, J., Ewins, K., Spielman, R. and Cheung, V. (2004). Genetic analysis of genome-wide variation in human gene expression. *Nature* 230, 743–747.

Nalls, M., Wilson, J., Patterson, N., Tandon, A., Zmuda, J., Huntsman, S., Garcia, M., Hu, D., Li, R., Beamer, B. et al. (2008). Admixture mapping of white cell count: Genetic locus responsible for lower white blood cell count in the health ABC and Jackson Heart Studies. *Am J Hum Genet* 82, 81–87.

Nussbaum, R., McInnes, R. and Willard, H.F. (2007). *Genetics in Medicine, 7th edition.* WB Saunders Co, Philadelphia.

Ogawa, Y. and Lee, J. (2002). Antisense regulation in X inactiviation and autosomal imprinting. *Cytogenet genome res* 99, 59–65.

Osoegawa, K., Mammoser, A., Wu, C., Frengen, E., Zeng, C., Catanese, J. and de Jong, P. (2001). A bacterial artificial chromosome library for sequencing the complete human genome. *Genome Res* 11, 483–496.

Paschou, P., Ziv, E., Burchard, E., Choudhry, S., Rodriguez-Cintron, W., Mahoney, M. and Drineas, P. (2007). PCA-correlated SNPs for structure identification in worldwide human populations. *PLoS Genet* 3, e160.

Pauling, L., Itano, H., Singer, S.J. and Wells, I.G. (1949). Sickle cell anemia a molecular disease. *Science* 110(2865), 543–548.

Pennacchio, L., Ahituv, N., Moses, A., Prabhakar, S., Nobrega, M., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K. et al. (2006). In vivo enhancer analysis of human conserved non-coding sequences. *Nature* 444, 499–502.

Pheasant, M. and Mattick, J. (2007). Raising the estimate of functional human sequences. *Genome Res* 17, 1245–1253.

Price, A., Patterson, N., Yu, F., Cox, D., Waliszewska, A., McDonald, G., Tandon, A., Schirmer, C., Neubauer, J., Bedoya, et al. (2007). A genomewide admixture map for Latino populations. *Am J Hum Genet* 80, 1024–1036.

Qiu, J. and Hayden, E. (2008). Genomics sizes up. *Nature* 451, 234.

Redon, R., Ishikawa, S., Fitch, K., Feuk, L., Perry, G., Andrews, T., Fiegler, H., Shapero, M., Carson, A., Chen, W. et al. (2006). Global variation in copy number in the human genome. *Nature* 444, 444–454.

Scherer, S., Lee, C., Birney, E., Altshuler, D., Eichler, E., Carter, N., Hurles, M. and Feuk, L. (2007). Challenges and standards in integrating surveys of structural variation. *Nat Genet* Suppl. 39, 7–15.

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S., Massa, H., Walker, M., Chi, M. et al. (2004). Large-scale copy number polymorphism in the human genome. *Science* 305, 525–528.

Serre, D., Gurd, S., Ge, B., Sladek, R., Sinnett, D., Harmsen, E., Bibikova, M., Chudin, E., Barker, D., Dickinson, T. et al. (2008). Differential allelic expression in the human genome: A robust approach to identify genetic and epigenetic *cis*-acting mechanisms regulating gene expression. *PLoS Genet* 4, e1000006.

Sharp, A., Cheng, Z. and Eichler, E. (2006). Structural variation of the human genome. *Annu Rev Genomics Hum Genet* 7, 407–442.

Shianna, K. and Willard, H.F. (2006). Human genomics: In search of normalcy. *Nature* 444, 428–429.

Shriver, M. and Kittles, R. (2004). Genetic ancestry and the search for personalized genetic histories. *Nat Rev Genet* 5, 611–618.

Smith, M., Patterson, N., Lautenberger, J., Truelove, A., McDonald, G., Waliszewska, A., Kessing, B., Malasky, M., Scafe, C., Le, E. et al. (2004). A high-density admixture map for disease gene discovery in African Americans. *Am J Hum Genet* 74, 1001–1013.

Spielman, R., Bastone, L., Burdick, J., Morley, M., Ewens, W. and Cheung, V. (2007). Common genetic variants account for differences in gene expression among ethnic groups. *Nat Genet* 39, 226–231.

Stefansson, H., Helgason, A., Thorleifsson, G., Steinthorsdottir, V., Masson, G., Barnard, J., Baker, A., Jonasdottir, A., Ingason, A., Gudnadottir, V. et al. (2005). A common inversion under selection in Europeans. *Nat Genet* 37, 129–137.

Stranger, B., Forrest, M., Dunning, M., Ingle, C., Beazley, C., Thorne, N., Redon, R., Bird, C., de Grassi, A.., Lee, C. et al. (2007a). Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 315, 848–853.

Stranger, B., Nica, A., Forrest, M., Dimas, A., Bird, C., Beazley, C., Ingle, C., Dunning, M., Flicek, P., Koller, D. et al. (2007b). Population genomics of human gene expression. *Nature Genet* 39, 1217–1224.

Tian, C., Plenge, R., Ransom, M., Lee, A., Villoslada, P., Selmi, C., Klareskog, L., Pulver, A., Qi, L., Gregersen, P. et al. (2008). Analysis and application of European genetic substructure using 300K SNP information. *PLoS Genet* 4, e4.

Trask, B. (2002). Human cytogenetics: 46 chromosomes, 46 years and counting. *Nat Rev Genet* 3, 769–778.

Tuzun, E., Sharp, AJ., Bailey, JA., Kaul, R., Morrison, VA., Pertz, LM., Haugen, E., Hayden, H., Albertson, D., Pinkel, D. et al. (2005). Fine-scale structural variation of the human genome. *Nature Genet* 37(7), 727–732.

van Rooij, E., Liu, N. and Olson, E.N. (2008). MicroRNAs flex their muscles. *Trends Genet*, 24, 159–166.

Venter, J., Adams, M., Myers, E., Li, P., Mural, R., Sutton, G., Smith, H., Yandell, M., Evans, C., Holt, R. et al. (2001). The sequence of the human genome. *Science* 291, 1304–1351.

Wang, J., Wang, W., Li, R., Li, Y., Tian, G., Goodman, L., Fan, W., Zhang, J., Li, J., Zhang, J. et al. (2008). The diploid genome sequence of an Asian individual. *Nature* 456, 60–65.

Weiss, L.A., Shen, Y., Korn, J., Arking, D., Miller, D., Fossdal, R., Saemundsen, E., Stefansson, H., Ferreira, M., Green, T. et al. (2008). Association between microdeletion and microduplication at 16p11.2 and autism. *New Engl J Med* 358, 667–675.

Wheeler, D.A., Srinivasan M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.-J., Makhijani, V., Roth, G.T. et al. (2008). The complete genome of an individual by massively parallel DNA Sequencing. *Nature* 452, 872–876.

Wong, K., deLeeuw, R., Dosanjh, N., Kim, L., Cheng, Z., Horsman, D., MacAulay, C., Ng, R., Brown, C., Eichler, E. et al. (2007). A comprehensive analysis of common copy-number variataions in the human genome. *Am J Hum Genet* 80, 91–104.

Xin, H. (2007). The million-dollar genome. *Science* 318, 550.

RECOMMENDED RESOURCES

Brown, T.A. (2006). *Genomes 3, 3rd edition*. Garland Science Publ., New York.

Nussbaum, R., McInnes, R. and Willard, H.F. (2007). *Genetics in Medicine, 7th edition*. WB Saunders Co., Philadelphia.

Strachan, T. and Read, A.P. (2004). *Human Molecular Genetics 3, 3rd edition*. Garland Science Publ., New York.

NCBI Human Genome Resources
http://www.ncbi.nlm.nih.gov/Genome/guide/human/

Online Mendelian Inheritance in Man
http://www.ncbi.nlm.nih.gov/sites/entrez?db=OMIM&itool=toolbar

UCSC Genome Bioinformatics
http://genome.ucsc.edu/

Ensembl Human Genome Browser
http://www.ensembl.org/Homo_sapiens/index.html

Genome Sequence of J. Craig Venter ("HuRef")
http://www.jcvi.org/cms/research/projects/huref/overview/

Genome Sequence of James D. Watson
http://jimwatsonsequence.cshl.edu/about.html

CHAPTER

DNA Sequencing for the Detection of Human Genome Variation

Samuel Levy and Yu-Hui Rogers

INTRODUCTION

The discovery and characterization of DNA polymorphisms in human populations is an important step toward understanding the contribution of genome sequence variants to predisposition, onset and progression of disease phenotypes. By way of example, molecular epidemiology studies have been able to establish the importance of the wide range of DNA variants that cause monogenic disorders such as cystic fibrosis, where over 1000 mutations in the CFTR gene have been identified and implicated in the disease phenotype (Rowntree and Harris, 2003). The impact of the mutational landscape of the CFTR gene on biological function is complex, and emerging evidence implicates additional modifier genes contributing to the range of disease phenotypes observed (Davies et al., 2005). The situation is likely to be compounded further when identifying and characterizing the contribution of multiple genetic loci to phenotypes found in polygenic diseases such as coronary heart disease, diabetes and hypertension (see Chapter 2). While our understanding of how polymorphisms can contribute either singly or collectively to either disease onset or progression is still actively being developed, so too are the technologies that enable their detection and ascertainment in human populations.

Human DNA polymorphism detection has been greatly empowered in the last decade by the production of a high quality, mostly finished, DNA sequence of the human genome. The production of this sequence, and the fact that several distinct DNA sources were employed, enabled the direct identification of 2.1 million single nucleotide polymorphisms (SNPs) (Lander et al., 2001; Venter et al., 2001) to supplement those already found in the public variant databases (dbSNP). The discovery of these SNPs in a small number of humans set the stage for an ambitious experiment, the HapMap project, to define common haplotypes through the use of 4 million SNP types in 269 distinct humans (Consortium, 2005). It is anticipated that these datasets will enable the identification of a set of informative SNPs with which disease association studies can be accomplished with improved success. Indeed, early successes highlighting the utility of the HapMap data in identifying alleles implicated in complex diseases such as asthma (Laitinen et al., 2004), age related macular degeneration (Klein et al., 2005) and type II diabetes (Sladek et al., 2007), amongst others, have clearly set the stage for more studies of this kind (see Chapter 4). Essentially, one potential study design would likely involve the use of SNPs as markers for a disease in an affected DNA population compared to a control DNA population. Subsequently, any marker statistically associated with the disease group would inform the genomic loci requiring further sequence analysis to detect the causative variant. There are currently over 10 million SNPs and other polymorphisms such as insertion/deletion variants (indels) in public databases that potentially provide a marker set for efficient disease-gene association studies. It is important to note that even this large variant

Copyright © 2010, Elsevier Inc.
All rights reserved.

set still might not represent the variants causative of diseases. This is likely due to the fact that significant polymorphism discovery efforts to date have been performed in a random fashion across the genomes of only a limited number of individuals. Therefore, we still need to discover genetic variation in regions of functional DNA sequence in the genomes of individuals with disease, in order to provide an accurate understanding of disease etiology.

In this chapter, we outline the sequencing protocols and the attendant computational approaches employed for the discovery and initial characterization of DNA variants, with an emphasis on the most commonly used Sanger-based sequencing (Sanger et al., 1977) methodologies established in our laboratories. These polymorphisms range from the substitution of single bases (i.e., SNPs) to the insertion or deletion of nucleotides in size ranges from 1 basepair (bp) to greater than 20,000 bp. Recent reports on the unexpected number and scope of larger scale copy number variants or structural variants ranging from 1–3 million bases (Mb) in size (Redon et al., 2006; Sebat et al., 2004) will not be addressed here since current methods used for their analysis involve either genotyping or microarray-based comparative genome hybridization, as presented elsewhere (see Chapter 5).

We also highlight, where possible, how the application of these methods and advances has impacted our understanding of human biology. Several recent reviews addressed the methodologies and development of advanced sequencing methods and recent commercial platforms that exploit their use (Fan et al., 2006; Shendure et al., 2004). These advances in genome sequencing for variant detection are of particular importance as we have begun to see genome sequences and complete variants sets of individual humans, including as of this writing James D. Watson by 454 Life Sciences (Wheeler et al., 2008) and J. Craig Venter (Sanger-based sequencing (Levy et al., 2007)). The discussion in this chapter will focus primarily on mature approaches for use in high-throughput experimental environments, as well as on the types of experimental results that can and have been generated. It will be evident that comprehensive experimental designs and economies of scale are possible when applying these techniques in a production environment.

DNA SEQUENCING

DNA sequencing methods applied to either single or multiple DNA samples provide the definitive approach to discovering polymorphisms. These polymorphisms can be detected at the genomic and/or the mRNA level depending on the source of the material employed. For example, genomic DNA subcloned into BACs or in small to medium insert size bacterial libraries were created to produce a finished human genome sequence and genomic variants therein (Osoegawa et al., 2001). Other sequencing projects have employed expression sequence tag (EST) or cDNA sequencing of mRNA libraries to characterize variants primarily at the transcript, and by inference, at the protein level (Strausberg et al., 2003). All these approaches benefit from the random sampling of DNA regions by selecting bacterial colonies containing DNA inserts, DNA sequence determination using capillary-based electrophoresis of incorporated dye terminator nucleotide bases and subsequent sequence alignment of reads to construct reliable assemblies of genomic or mRNA molecules. Other approaches to DNA sequence elucidation that will be covered in this section involve the targeted amplification of defined genomic loci typically for study in large populations, the use of microfabricated high-throughput bead or surface-based sequencing devices and high-density oligonucleotide arrays. A comparison of these distinct methodologies is provided in Table 3.1.

Whole Genome Shotgun Sequencing

Polymorphism discovery on a genome-wide level can be accomplished by applying the Sanger sequencing technique (Sanger et al., 1977) and the whole genome shotgun (WGS) strategy (Fleischmann et al., 1995; Venter et al., 2001; Venter et al., 1996) on a small pool of DNA samples.

Owing to its high accuracy, versatility and ability to generate complete DNA sequence information, the Sanger sequencing method has been considered the definitive approach for DNA variant discovery. The WGS methodology relies upon generating a random sampling of genomic regions at a sufficiently deep coverage level to produce enough sequence data that can then be assembled by sequence alignment methods. First, multiple copies of the genome are randomly shredded into pieces of approximately 2–40 kilobasepairs (Kb) in size and subsequently cloned into a vector or plasmid. These constructs are then replicated in bacteria and sequenced from both ends to produce pairs of linked sequences, termed "mate pairs", representing 500–800 bp at the end of each insert. These sequence reads are then assembled computationally to generate a set of contiguous high-quality regions with potential information regarding any associated sequence gaps of a known size.

Currently, the Sanger-based WGS sequencing process, excluding the library construction step, is typically fully supported by equipment automation, as outlined in Figure 3.1, and managed as a production "pipeline" with varying degrees of integration between different steps employing a laboratory information management system.

Library Construction

Genomic DNA is randomly sheared via nebulization to produce fragments with a distribution of approximately 2–40 Kb. The DNA fragments are ligated to adapters after end-polishing and size selection. After several rounds of purification, the fragments are inserted into plasmid vectors. The resulting library is electroporated into *E. coli* cells.

Colony Plating and Picking

The *E. coli* cells containing the plasmid vectors with inserted DNA fragments are then spread onto agar plates and incubated to form *E. coli* colonies. The colonies are picked into liquid media and incubated for further growth to generate sufficient

Table 3.1 Comparison of distinct approaches in DNA sequence acquisition and the detection of variants

	Sanger based WGS	**Targeted PCR**	**Sequence by synthesis**	**Sequence by hybridization**
Current platforms	ABI 3730 xl	ABI 3730 xl	Roche-454, Illumina/ Solexa, ABI SOLiD (strictly hybridization-ligation chemistry)	Affymetrix, Perlegen, Nimblegen
Read length (bp)	500–800	500–800	25–400	60–100
Detection SNP/Indel	Y/Y	Y/Y (up to amplicon size)	Y/Y	Y/Y (primarily deletions in sample)
Construction of long-range haplotypes	Y	Limited to targeted regions	Uncertain	Limited to targeted regions
Current cost per sample < $2 million	N	Uncertain	Y	Y

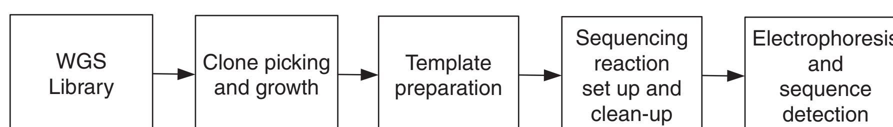


Figure 3.1 Whole Genome Shotgun sequencing process flow.

numbers of plasmid template copies. Subsequently, these template copies are isolated and purified from the host cells.

Template Preparation

The alkaline lysis plasmid isolation approach (Sambrook et al., 1989) is the most widely utilized method for plasmid template isolation due to its robustness. Bacterial cells are lysed, cell debris removed by centrifugation and plasmid DNA recovered from the cleared lysate by isopropanol precipitation.

Sequencing Reactions

For DNA sequence determination, the most widely used technique is the Sanger dideoxy sequencing method (Sanger et al., 1977). It mimics the natural process of DNA replication in vivo with the presence of target DNA molecules, universal M13 sequencing primers, polymerase, deoxynucleotides (dNTPs) as well as chain terminating dideoxynucleotides which interrupt chain elongation upon incorporation generating DNA fragments of different lengths. Several improvements have been made to the original method to greatly enhance its throughput, robustness and detection sensitivity. These include the development of multicolor fluorescent detection (Smith et al., 1986), fluorescently tagged chain-terminators (Prober et al., 1987), a cycle sequencing protocol (McCombie et al., 1992) and fluorescence resonance energy transfer (FRET) dyes (Ju et al., 1995; Lee et al., 1997). Currently, most sequencing reactions are carried out using FRET dye terminators, modified *Taq* polymerase and a cycle sequencing protocol.

Sequence Detection

The technology for the electrophoresis-based separation of Sanger sequenced DNA fragments and the detection of individual bases have advanced substantially in the past two decades. As a result, the throughput, detection sensitivity and costs of sequencing, as well as the accuracy and quality of the data generated by these sequencers, have improved significantly. The invention of multicolor fluorescent sequencing in 1986 (Smith et al., 1986) replaced laborious and hazardous radioisotope labeling techniques (Sanger et al., 1982) with automated signal detection, high-speed computational recording and signal processing. In the mid-1990s, the successful integration of the capillary-based electrophoresis and multicolor fluorescent detection (Bashkin et al., 1996a; Bashkin et al., 1996b; Behr et al., 1999) further enabled the creation of the true high-throughput sequencing machines. Today, state-of-the-art capillary-based sequencers have the ability to read over 1 million bp of sequence per 24-hour period with long read lengths (an average of >800 bp per read) and an average quality value (QV) of greater than 30 (99.9% accuracy).

Whole Genome Assembly

The whole genome assembly process utilizes several aspects of experimental design to ensure that unambiguous construction of long contiguous sequence can be generated. These include the generation of sequence reads from each end of clone inserts using universal primers pairs to ensure that a majority of inserts will have their ends sequenced. This enables both the generation

	T	I	R	O
GGAGGACGTGCGCGGCCGCCTGGTGCAGTACCGCGGCGAGGTGCAGGCCATGCTCGGCCAGAGCACCGAG	Contig	1100189367757	6835 - 6905	-
..........T...	read	1098415676515	316 - 386	-
..........T...	read	1098537170481	392 - 462	-
....	read	1099790795496	352 - 422	-
....	read	1099890242771	248 - 318	-
....	read	1099896833584	448 - 518	-
..........T...	read	1099806280429	377 - 447	+
....	read	1099791382666	311 - 381	-
..........T...	read	1099882077584	331 - 401	-
....	read	1099891898631	553 - 623	-
....	read	1099901226215	374 - 444	+
....	read	1099914202318	417 - 487	+
..........T...	read	1099905252229	619 - 689	-

Figure 3.2 Display and detection of a single nucleotide polymorphism in a human genome assembly of a single DNA donor. Sequence reads overlap and can be assembled to form the consensus sequence of the contig where 5 of the 12 reads show a T variation at a C consensus base (variant position is highlighted with an arrow). T = type of sequence read, either assembled contig or individual read, I = unique identifier, R = co-ordinates of alignment, O = strand orientation that aligns.

of contiguous sequence and the potential for ordering and orienting larger sequence segments essentially created by sequence alignment of the individual reads whose mate pair relationship is known. A variety of genome assembler software packages have been designed with this basic rationale at their core (Celera Assembly (Myers et al., 2000), Phusion (Mullikin and Ning, 2003)). The Atlas assembler uses as input both BAC-based clone sequences and reads generated via the WGS strategy (Havlak et al., 2004), an approach that was used to assemble the rat genome (Gibbs et al., 2004).

There are two basic steps to assembly: (i) the creation of unique regions of contiguous (contigs) assemblies from the sequence overlaps between reads and (ii) the sequential end-to-end organization of contigs by employing the mate pair information contained in the reads constituting each contig. The successful application of this strategy was a significant challenge for the sequencing and assembly of human DNA since greater than 45% of the genome sequence is repetitive in nature (Lander et al., 2001; Venter et al., 2001). The repetitive nature of the human and other mammalian genomes thus confounds the ability to determine accurate read sequence overlaps and the placement of contigs into the correct order and orientation. Our solution is to employ clone insert libraries of a size larger than the corresponding repeat regions, typically either fosmid or BAC libraries (40 kb and >100 kb respectively), thus enabling the spanning of repeats by assembling adjacent unique sequence (Venter et al., 2001). The selection of a range of insert sizes libraries is typically built into the experimental design when sequencing a new genome.

Variant Detection

Variants can be reliably detected from assembled sequence reads from either a single or multiple DNA donors in the region of unique genome sequence assembly. The example in Figure 3.2 show the reads assembled from a single human identifying a known SNP found in a coding region of the APOE gene and, as expected, displays a close to equiprobable occurrence of both allelic forms (five reads with the T allele and seven reads with C allele). This approach has been employed with success to identify 974,400 heterozygous variants in a single canine individual (Kirkness et al., 2003). The inclusion of random shotgun sequence reads from distinct DNA donors is another approach to employing genome assembly and read placement to identify polymorphic site. This approach was employed using five DNA donors in the initial shotgun sequence assembly of the human genome (Venter et al., 2001). An important factor in using a WGS assembly approach to detecting polymorphic sites is that enough sequence coverage from a single donor should be available. In this manner it is possible to provide a good quality assembly with associated consensus sequence and associated allelic variants. This means that additional donors can be included either post-assembly or as part of the assembly process if the species in question does not possess extensive interindividual polymorphism. Recent work attempting to detect both SNP and indels in an individual human sample suggests that detection of heterozygous variants, with 99% probability using a Sanger chemistry-based WGS approach, necessitates a variant being reported by 20 reads or more (Levy et al., 2007).

Another challenge to polymorphism detection during genome assembly is the occurrence of many small heterozygous indel polymorphisms in human DNA, typically between 1–20 bp length (Ball et al., 2005). These indel events result in the occurrence of long and short alleles, requiring a methodology to accurately separate the reads from each allele, thereby providing an accurate representation of the indel. Subsequently, a practical decision needs to be made regarding which form will be

represented in the consensus sequence. Typically this can be the "major" allele even though in a single donor this will be present effectively at equal probability.

Lastly, a strategy needs to be implemented to best catalog the occurrence and type of the other allele or alleles, in the case of sequencing multiple DNA donor samples. An example of the detection and display of an indel from a genome assembly of a single DNA donor can be seen in Figure 3.3. The figure displays the consensus sequence (the contig sequence) containing the short, major, allele since gaps are included in the alignment, whilst the long form (inclusion of ATTCT) is found in two of the eight assembled reads. We recently applied a modified sequence assembly algorithm to detect heterozygous indels variants from the diploid genome sequence of an individual human (Levy et al., 2007). Using this approach we detected over 345,000 heterozygous indels events of lengths 1–571 bp with at least 200 indels in protein-coding regions of genes.

PCR Amplicon Re-Sequencing

The goal of directed sequencing or re-sequencing is to use the existing sequence data from a particular genomic locus or loci to determine sequence variations of these same loci in different DNA samples. This provides the nature and the frequency of DNA polymorphisms and can be used to assess whether these variants could be implicated in, for example, either disease onset or progression.

PCR re-sequencing methodology is widely accepted as the gold standard (Kwok and Chen, 2003) for the discovery of polymorphisms (both SNPs and indels) in targeted regions since it provides the most complete information including the genotype, the location and sequence context. This approach has been applied to understand sequence variation in populations (Crawford et al., 2004) and for the identification of variants and haplotypes that could potentially explain disease phenotypes in cancer (Cox et al., 2005). It combines the powerful PCR technique (Mullis et al., 1986) with the informative Sanger sequencing method (Sanger et al., 1977). Rapid advancement of the technologies and dramatic reduction in costs for direct sequencing of PCR products, as well as the completion of the reference human genomes in 2001 (Lander et al., 2001; Venter et al., 2001) has enabled the systematic large-scale discovery of polymorphisms in targeted regions. The approach generally involves PCR primer design, the amplification of the targeted region from whole genomic DNA using PCR, followed by Sanger sequencing and the analysis of the targeted sequences (Figure 3.4).

Primer Design

The goal of directed sequencing is to provide reliable amplification of a single genomic region, and this is readily achieved via the design of unique pairs of oligomers for priming of the PCR. The resulting amplicon encompassing the targeted region can vary in size; however reproducible and robust assays range from 350 bp to 1000 bp in length per amplicon. In order to provide coverage of a larger genomic locus with this amplicon size, it becomes necessary to tile PCR products with some overlap (typically 100 bp) between amplicons. The primer design process attempts to generate oligomers for products with similar melting temperatures whilst avoiding primer-dimerization forming interfering secondary structures. Other factors to permit specific amplification would be the G+C content of the region, the proximity

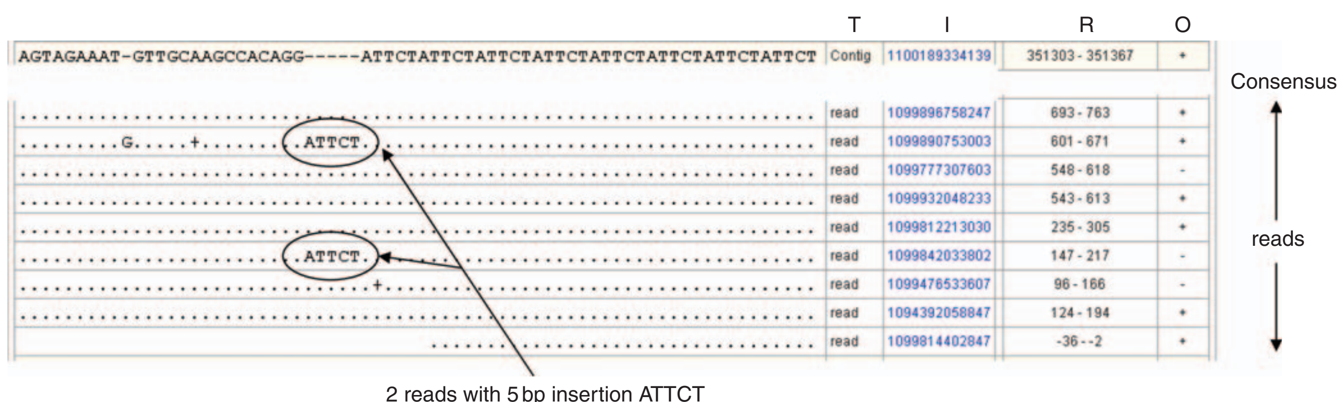


Figure 3.3 Display and detection of a 5 bp insertion/deletion polymorphism in a human genome assembly of a single DNA donor. The consensus sequence displays the "short" allelic form whilst the "long" form (ATTCT) is found in 2 of the 8 sequence assembled reads. T = type of sequence read, either assemblyed contig or individual read, I = unique identifer, R = co-ordinates of alignment, O = strand orientation that aligns. "+" = gap inserted into read to permit correct alignment.

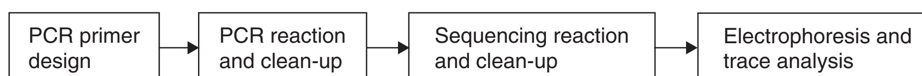


Figure 3.4 PCR directed sequencing process flow.

of repeat regions, the uniqueness of the oligomer sequences in the genome and the absence of known DNA variants in the primer binding sites. These considerations can be embodied in a computational pipeline as illustrated in Figure 3.5. The sequence context, in this case percent G+C quantity, are considered before identifying a candidate genomic region for primer design. In a high-throughput directed sequencing pipeline, it is frequently desirable to employ few, or preferable one, PCR protocol, and in our experience G+C content allows a distinction between at least two desired PCR conditions (high and low G+C content). Once the PRIMER3 software (Rozen and Skaletsky, 2000) generates candidate primer pairs for a desired target region, these are further assessed for their product uniqueness (low and high copy repeat regions) and to ensure that resulting amplicons contains only one type of complex sequence, i.e. homopolymer or known indel polymorphism. The last requirement is important since sequence detection is performed using capillary-based detection methods and chromatograms potentially contain a mixture of alternate alleles at any one mobility position. This renders the process of allelic definition potentially difficult in the case of multiple indel events in a single amplicon.

PCR Set-Up and Clean-Up

As described in the primer design section, due to the variable G+C content within the genome, application of different PCR primer design criteria, as well as different PCR protocols might be necessary when targeting different regions. In our laboratory, we have two fully validated high-throughput amplification protocols (Rand et al., 2005), one for regions of normal G+C content and another for amplicons with melting temperatures (T_m) >82°C. AmpliTaq Gold® DNA polymerase (Applied Biosystems) is used to perform hot-start PCR to minimize the formation of primer–dimers and spurious amplicons and ensure efficient, consistent and specific amplification results in a high-throughput environment.

A Shrimp Alkaline Phosphatase/Exonuclease I (SAP/Exo I) mix is used to digest excess dNTPs and amplification primers. (While SAP is commonly used to remove 5′ phosphate groups from DNA, it also degrades dNTPs.) The SAP/Exo I method provides an effective, single tube approach to PCR clean-up that is easily automated and scalable.

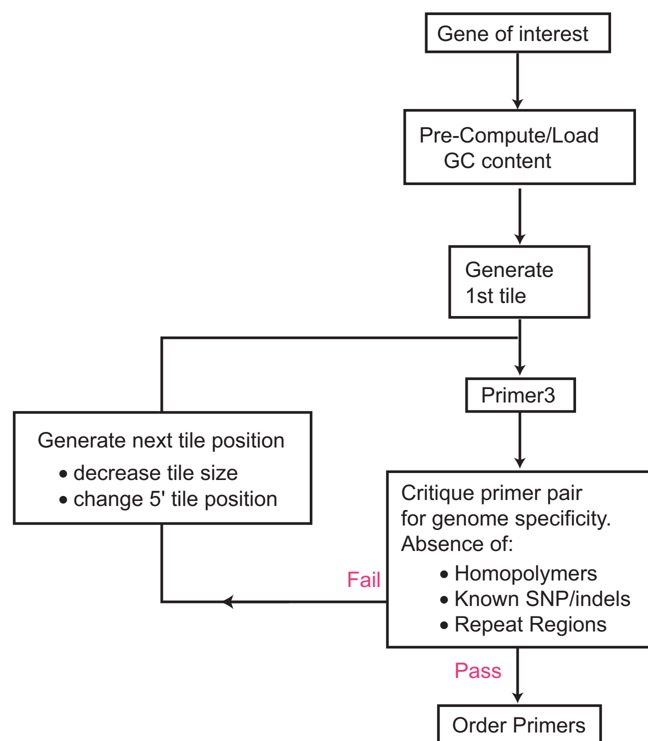


Figure 3.5 Primer design pipeline for PCR based directed sequencing.

Sequencing Reaction Set-Up and Clean-Up

In order to simplify the downstream sequencing process, the PCR primers are frequently tagged with M13 forward and reverse sequences on their 5′ ends. The resulting amplicons can then be sequenced using universal M13 sequencing primer sets and a standard sequencing protocol (see previous section). Reaction products can be purified with sodium acetate and ethanol prior to analysis on the sequencers.

Electrophoresis and Trace Analysis

The state-of-the-art technology for the electrophoresis-based separation of Sanger sequenced DNA fragments and the detection of individual bases can be accomplished by a combination of trace filtering, trace separation and mixed base calling. Electropherogram peaks are described as "mixed" in that a primary peak is detected with a secondary but significantly present sub-peak due to two distinct and different allelic forms contributing to the locus, We have developed a trace analysis pipeline that applies digital signal filters to reduce the occurrence of secondary, typically noisy, sequence peaks (Figure 3.6), thus enabling the accurate distinction of a single position with mixed peaks. This approach has enabled a substantial improvement of common artifacts observed in diploid genomic regions sequenced after PCR amplification that include dye-blob, PCR slippage and primer synthesis length errors, all of which confound mixed peak detection. After generating either a raw (unfiltered) or digital signal filtered chromatogram, it is then possible to identify mixed peaks that vary significantly in height, area and width to characterize departures from the reference sequence. This general approach can be performed with many different software tools in the public domain (Polyphred, Mutation Surveyor, novoSNP and SNPdetector, InSNP), the most extensively tested of which is Polyphred. In the most recent version of this software, detailed parameterization was performed using sequence reads from PCR-amplified diploid samples and a statistical framework was developed to report evidence of genotypes being called (Stephens et al., 2006).

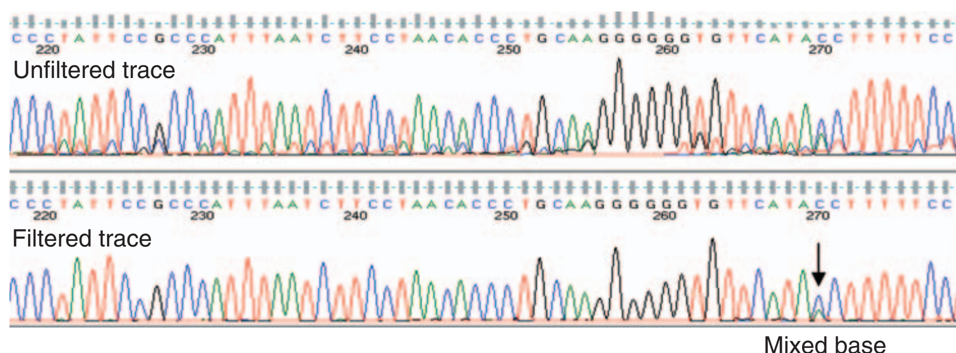


Figure 3.6 Sequence chromatograms of a locus revealing a SNP before and after digital signal filtering were applied to remove PCR slippage artifacts due to the presence of proximal (not seen) homopolymeric DNA.

Heterozygous indel polymorphisms result in significant changes in the chromatograms from diploid DNA samples where typically traces contain mixed peaks from the point of the insertion/deletion onwards. One can develop computational methods that essentially separate the traces to enable the definition of the long and short alleles (Bhangale et al., 2005; Bhangale et al., 2006). Polyphred, InSNP (Manaster et al., 2005) and Mutation Surveyor are all capable of detecting heterozygous indels, but as of this writing an unbiased detailed comparison of performance has not yet been attempted.

Sequencing by Synthesis

Recently implemented sequencing by synthesis technologies can employ PCR (Mullis et al., 1986) to achieve clonal amplification of single DNA molecules. In addition, the sequencing by synthesis principle in a massively parallel format has been developed by Rothberg and colleagues (Margulies et al., 2005) and by Church and colleagues (Shendure et al., 2005). In each of these new technologies, different chemistries are applied to interrogate, tag and detect the sequence bases.

For example, utilizing nanotechnology and pyrosequencing chemistry (Nyren et al., 1993), coupled with an enzymatic luminometric inorganic pyrophosphate detection technique, Rothberg and colleagues at 454 Life Sciences recently developed a novel and highly parallel system capable of sequencing approximately 100 million bp in a 7-hour period by sampling over 400,000 sequence clones. They were able to illustrate this technological advance by sequencing and assembling the *Mycoplasma genitalium* and *Streptococcus pneumoniae* genomes (0.6 and 2.1 Mb respectively in size). Compared to the current state-of-the-art Sanger DNA sequencing and capillary-based electrophoresis platform, their system generates "raw" data with approximately 100 times higher throughput. However, compared to the Sanger/electrophoresis-based sequencing approach, it still has several limitations, such as the short sequence read lengths which, averaging 200–250 bases per read, are only a quarter of Sanger read lengths. Another problem is that the accuracy of base calling on individual reads is low especially for genomic regions containing homopolymers. This leads to the increased likelihood of spurious polymorphism detection. In addition, the 454 sequencing methodology currently does not incorporate the use of mate pairs that come from large insert libraries (>10 kb), which for the successful assembly of repeat rich genomes, like those of most eukaryotes, are important in order to span large repeat regions. Nonetheless, this technology has been employed to generate some sequence data from DNA extracted from fossil remains (Green et al., 2006) and has been employed most recently to generate sixfold coverage of an individual genome (Wheeler et al., 2008).

Some of these problems can be circumvented by adopting a hybrid sequencing approach, that is one that employs, in addition to the sequencing-by-synthesis approach (454), a small number of large insert clones (typically fosmid libraries) sequenced by Sanger technology to provide order and orientation in the final genome assembly. While this approach has been successfully employed for a number of bacterial genomes (Goldberg et al., 2006), establishing the most efficient mix of the sequence-by-synthesis contribution and fosmid libraries will be the challenge for the completion of larger mammalian genomes. Each of the sequencing-by-synthesis approaches mentioned above share similar advantages as well as limitations. However, they do show great promise for the discovery of polymorphisms in a rapid, cost effective manner.

The massively parallel sequencing-by-synthesis approach enabled by the 454 picoliter plate technology can also be utilized for targeted polymorphism discovery. In this manner, PCR amplicons that contain the targeted genomic regions are extensively sampled since it is possible to generate currently up to 400,000 clonal amplifications of the PCR product per experiment. This approach has potentially great value for the discovery and detection of DNA variations present in low frequency in the samples. This is especially applicable for mutation detection in tumor tissue, in which potentially only a few cells carry activating mutations and the surrounding cell population contributes to the "contamination" of the tumor DNA. The lowest detection level thus far obtained has been for alleles present in the sample as low as 9% for indel mutations in the EGFR gene in lung adenocarcinoma specimens (Thomas et al., 2006) and at a similar frequency for non-synonymous point mutations in the FGFR1 gene from glioblastoma samples (Strausberg et al., personal communication). In both of these cases the corresponding PCR-amplified,

Sanger sequence chromatograms lack the necessary secondary peak signal to detect a variant above the background noise level, where the limit of detection of a secondary allele is thought to be between 15–25%.

Sequencing by Hybridization

The sequencing by hybridization on microarray approach has also become a widely utilized tool for SNP detection and analysis (Kukita et al., 2005; Patil et al., 2001; Maraganore et al., 2005; Kozal, 1996; Cronin et al., 1996; Chee et al., 1996; Hacia et al., 1996; Kapranov et al., 2002) due to the technological advances that make the manufacturing of high-density DNA arrays possible. The genomic DNA is generally prepared by PCR to amplify the regions of interest or by using a whole genome amplification methodology (Dean et al., 2002) and tagged with fluorescent dyes for detection prior to hybridizing to the oligonucleotide array.

This approach allows the detection of sequence present in a test sample via direct hybridization and was able to detect the unique (non-repeat) contiguous sequence from human DNA chromosome 21 conserved with mouse and dog (Frazer et al., 2001) and the haplotype block structure of chromosome 21 in 24 human DNA samples (Patil et al., 2001). In both cases, this was accomplished by creating a tiling array where each position was interrogated for all possible bases using a set of eight distinct oligonucleotide probes comprising the position in question and 12 bp flanking to provide for binding specificity. The interspecies hybridization experiments revealed additional putative exonic region of current genes and identified other regions of significant sequence conservation that could represent other forms of functional sequence. The human population study revealed that 80% of the DNA samples could be described by three common haplotypes. Such knowledge of chromosomal block structure and inheritance patterns enables the choice of a smaller set of variants with which one can perform disease-gene association studies (see Chapter 4). The sequencing-by-hybrization approach clearly necessitates that all the sequence being tested is present on the microarray. While this may be a limitation for discovering novel sequence in some genomes, even for the multiple, yet to be sequenced, human genomes, the methodology is well suited and has been extensively employed for the genotyping of known SNPs in disease conditions and deletions in arrayed sequence. Recent whole genome association studies employing large-scale SNP assays performed on microarrays have identified genes as either biomarkers or therapeutic targets in Parkinson's disease (~200,000 loci) (Maraganore et al., 2005) and copy number variations and deletions in mental retardation (100,000 loci) (Friedman et al., 2006).

2009 UPDATE

Individual human genome sequencing

DNA sequencing has witnessed an incredible increase in throughput, cost reduction and facility of application over the last two years. The so-called "Next Generation" sequencing platforms as provided by Roche-454 GS (**G**enome **S**equencer), Illumina-Solexa GA (**G**enome **A**nalyzer) and ABI SOLiD (**S**upported **O**ligonucleotide **Li**gation and **D**etection) have been used in a variety of sequencing approaches. Most recently, 454 and Illumina GA have been employed to detect comprehensive sets of polymorphisms in individual human genomes (Bentley et al., 2008; Wang et al., 2008; Wheeler et al., 2008). These data reveal a trend in the catalogue of variants found in now four individuals of distinct ethnicity whose genomes have been sequenced, whereby SNPs occur in the millions, insertion/deletion polymorphisms (<1 kb) occur in the hundreds of thousands and structural variants (>1 kb) occur in the thousands.

The sequencing platforms that produced these individual human genomes are continuing to improve in sequence accuracy, read length and the complexity of DNA libraries to enable their use for *de-novo* genome assembly, which to date has only be accomplished by mature ABI Sanger chemistry and a dedicated genome assembly tool (Levy et al., 2007). The sequence variation discovered using the new platforms has been achieved by mapping of individual reads or paired end reads to the existing human genome reference sequence (http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/index.shtml) using a variety of tools like Newbler for Roche 454, MAQ (Li et al., 2008) and SOAP (Li et al., 2008a) of Illumina-Solexa, and Corona Lite and SHRiMP for ABI SOLiD. Typically these software tools provide the placement of reads on the reference genome and knowledge of any variant bases contained within any read. Some statistical inference can then be performed to assess the presence of SNPs or indels, as has been developed for the 454 platform (Quinlan et al., 2008). Larger scale human polymorphism, like structural variation, can be detected by comparing the mapping distance between paired reads from known insert size libraries onto the universal human reference genome. This pair-end mapping approach has been performed for fosmid inserts (~40 kb) whose ends were sequenced with ABI Sanger chemistry to detect structural variants in humans (Kidd et al., 2008) and has shown to work with library sizes as small as 3 kb (Korbel et al., 2007) that were sequenced with Roche-454. An ongoing challenge remains the detection of copy number variants in the human genome using sequencing approaches

alone, since they have been more extensively studied with comparative hybridization approaches to date (Redon et al., 2006).

It is clear that some variation detection can be achieved via sequence read mapping using "Next Generation" platform data sets. However, the relationships among variants (that is, haplotype data) will only be revealed by some preceding genome assembly or haplotype inference. Therefore, the need to perform *de-novo* genome assembly with short reads generated by these platforms still remains an important if difficult challenge and some groups have produced tools that appear to work with non-mammalian genomes (Jeck et al., 2007; Warren et al., 2007; Butler et al., 2008; Hernandez et al., 2008; Zerbino and Birney, 2008). The size and repeat sequence diversity of the human genome render genome assembly a complex problem; however, the continued improvement of read length on "Next Generation" platforms promises to reduce this complexity.

Targeted sequencing

Despite the incredible sequencing throughput capabilities of the current commercial platforms, genetic association studies employing the whole genome shotgun sequencing strategy alone are still quite costly and time consuming to accomplish at a large scale. Consequently, the use of targeted sequencing approaches that employ locus-specific capture methods coupled with sequencing is still favored. The approaches have been supplemented beyond PCR by the development of microarray-based methodologies to target regions (currently exons), resulting in the specific isolation of up to 200,000 exons *en masse* (Albert et al., 2007; Hodges et al., 2007; Okou et al., 2007; Porreca et al., 2007). These methods have been able to show, in some cases, the identification of 98% of targeted exons with 55-85% of captured sequence reads derived from the targeted regions. These massively multiplexed regional capture methods, coupled with a "Next Generation" sequencing platform, promise to provide variant detection at high accuracy and throughput.

Future projects

The vastly improved sequence data throughput provided by the Roche-454 GS, Illumina-Solexa GA and ABI-SOLiD platforms have enabled large scale, international sequencing projects in human populations (e.g., 1,000 Genomes and The Cancer Genome Atlas Projects). These projects attempt relatively low coverage (2-6x) of the human genome to detect variants and/or mutations in either chronic diseases or distinct tumor types. These project designs are likely to reveal the complex nature of polymorphism in disease populations, and such study designs have been fueled by radical technology improvement in sequencing. Further improvements in sequencing technologies are likely to push study design toward the detection and analysis of polymorphism in multiple individual human genomes, including both common and rare variants.

2009 UPDATE REFERENCES

Albert, T.J., Molla, M.N. et al. (2007). Direct selection of human genomic loci by microarray hybridization. *Nat Methods* 4(11), 903–905.

Bentley, D.R., Balasubramanian, S. et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456(7218), 53–59.

Butler, J., Maccallum, I. et al. (2008). ALLPATHS: *De novo* assembly of whole-genome shotgun microreads. *Genome Res* 18, 810–820.

Hernandez, D., Francois, P. et al. (2008). *De novo* bacterial genome sequencing: Millions of very short reads assembled on a desktop computer. *Genome Res* 18, 802–809.

Hodges, E., Xuan, Z. et al. (2007). Genome-wide *in situ* exon capture for selective resequencing. *Nat Genet* 39(12), 1522–1527.

Jeck, W.R., Reinhardt, J.A. et al. (2007). Extending assembly of short DNA sequences to handle error. *Bioinformatics* 23(21), 2942–2944.

Kidd, J.M., Cooper, G.M. et al. (2008). Mapping and sequencing of structural variation from eight human genomes. *Nature* 453(7191), 56–64.

Korbel, J.O., Urban, A.E. et al. (2007). Paired-end mapping reveals extensive structural variation in the human genome. *Science* 318(5849), 420–426.

Levy, S., Sutton, G. et al. (2007). The diploid genome sequence of an individual human. *PLoS Biol* 5(10), e254.

Li, R., Li, Y. et al. (2008a). SOAP: Short oligonucleotide alignment program. *Bioinformatics* 24(5), 713–714.

Li, H., Ruan, J. et al. (2008b). Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* 18(11), 1851–1858.

Okou, D.T., Steinberg, K.M. et al. (2007). Microarray-based genomic selection for high-throughput resequencing. *Nat Methods* 4(11), 907–909.

Porreca, G.J., Zhang, K. et al. (2007). Multiplex amplification of large sets of human exons. *Nat Methods* 4(11), 931–936.

Quinlan, A.R., Stewart, D.A. et al. (2008). Pyrobayes: An improved base caller for SNP discovery in pyrosequences. *Nat Methods* 5(2), 179–181.

Redon, R., Ishikawa, S. et al. (2006). Global variation in copy number in the human genome. *Nature* 444(7118), 444–454.

Wang, J., Wang, W. et al. (2008). The diploid genome sequence of an Asian individual. *Nature* 456(7218), 60–65.

Warren, R.L., Sutton, G.G. et al. (2007). Assembling millions of short DNA sequences using SSAKE. *Bioinformatics* 23(4), 500–501.

Wheeler, D.A., Srinivasan, M. et al. (2008). The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452(7189), 872–876.

Zerbino, D. and Birney, E. (2008). Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res* 18, 821–829.

ACKNOWLEDGEMENTS

The authors would like to thank Drs Robert Strausberg, Jiaqi Huang and Pauline Ng for their suggestions and critical review of the manuscript.

REFERENCES

Ball, E.V., Stenson, P.D., Abeysinghe, S.S., Krawczak, M., Cooper, D.N. and Chuzhanova, N.A. (2005). Microdeletions and microinsertions causing human genetic disease: Common mechanisms of mutagenesis and the role of local DNA sequence complexity. *Hum Mutat* 26, 205–213.

Bashkin, J., Marsh, M., Barker, D. and Johnston, R. (1996a). DNA sequencing by capillary electrophoresis with a hydroxyethylcellulose sieving buffer. *Appl Theor Electrophor* 6, 23–28.

Bashkin, J.S., Bartosiewicz, M., Roach, D., Leong, J., Barker, D. and Johnston, R. (1996b). Implementation of a capillary array electrophoresis instrument. *J Capillary Electrophor* 3, 61–68.

Behr, S., Matzig, M., Levin, A., Eickhoff, H. and Heller, C. (1999). A fully automated multicapillary electrophoresis device for DNA analysis. *Electrophoresis* 20, 1492–1507.

Bhangale, T.R., Rieder, M.J., Livingston, R.J. and Nickerson, D.A. (2005). Comprehensive identification and characterization of diallelic insertion-deletion polymorphisms in 330 human candidate genes. *Hum Mol Genet* 14, 59–69.

Bhangale, T.R., Stephens, M. and Nickerson, D.A. (2006). Automating resequencing-based detection of insertion-deletion polymorphisms. *Nat Genet* 38, 1457–1462.

Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X.C., Stern, D., Winkler, J., Lockhart, D.J., Morris, M.S. and Fodor, S.P. (1996). Accessing genetic information with high-density DNA arrays. *Science* 274, 610–614.

Consortium, The International HapMap (2005). A haplotype map of the human genome. *Nature* 437, 1299–1320.

Cox, D.G., Kraft, P., Hankinson, S.E. and Hunter, D.J. (2005). Haplotype analysis of common variants in the BRCA1 gene and risk of sporadic breast cancer. *Breast Cancer Res* 7, R171–R175.

Crawford, D.C., Carlson, C.S., Rieder, M.J., Carrington, D.P., Yi, Q., Smith, J.D., Eberle, M.A., Kruglyak, L. and Nickerson, D.A. (2004). Haplotype diversity across 100 candidate genes for inflammation, lipid metabolism, and blood pressure regulation in two populations. *Am J Hum Genet* 74, 610–622.

Cronin, M.T., Fucini, R.V., Kim, S.M., Masino, R.S., Wespi, R.M. and Miyada, C.G. (1996). Cystic fibrosis mutation detection by hybridization to light-generated DNA probe arrays. *Hum Mutat* 7, 244–255.

Davies, J., Alton, E. and Griesenbach, U. (2005). Cystic fibrosis modifier genes. *J R Soc Med* 98 Suppl 45, 47–54.

Dean, F.B., Hosono, S., Fang, L., Wu, X., Faruqi, A.F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J. et al. (2002). Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci USA* 99, 5261–5266.

Fan, J.B., Chee, M.S. and Gunderson, K.L. (2006). Highly parallel genomic assays. *Nat Rev Genet* 7, 632–644.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M. et al. (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* 269, 496–512.

Frazer, K.A., Sheehan, J.B., Stokowski, R.P., Chen, X., Hosseini, R., Cheng, J.F., Fodor, S.P., Cox, D.R. and Patil, N. (2001). Evolutionarily conserved sequences on human chromosome 21. *Genome Res* 11, 1651–1659.

Friedman, J.M., Baross, A., Delaney, A.D., Ally, A., Arbour, L., Asano, J., Bailey, D.K., Barber, S., Birch, P., Brown-John, M. et al. (2006). Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation. *Am J Hum Genet* 79, 500–513.

Gibbs, R.A., Weinstock, G.M., Metzker, M.L., Muzny, D.M., Sodergren, E.J., Scherer, S., Scott, G., Steffen, D., Worley, K.C., Burch, P.E. et al. (2004). Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428, 493–521.

Goldberg, S.M., Johnson, J., Busam, D., Feldblyum, T., Ferriera, S., Friedman, R., Halpern, A., Khouri, H., Kravitz, S.A., Lauro, F.M. et al. (2006). A Sanger/pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes. *Proc Natl Acad Sci USA* 103, 11240–11245.

Green, R.E., Krause, J., Ptak, S.E., Briggs, A.W., Ronan, M.T., Simons, J.F., Du, L., Egholm, M., Rothberg, J.M., Paunovic, M. et al. (2006). Analysis of one million base pairs of Neanderthal DNA. *Nature* 444, 330–336.

Hacia, J.G., Brody, L.C., Chee, M.S., Fodor, S.P. and Collins, F.S. (1996). Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. *Nat Genet* 14, 441–447.

Havlak, P., Chen, R., Durbin, K.J., Egan, A., Ren, Y., Song, X.Z., Weinstock, G.M. and Gibbs, R.A. (2004). The Atlas genome assembly system. *Genome Res* 14, 721–732.

Ju, J., Ruan, C., Fuller, C.W., Glazer, A.N. and Mathies, R.A. (1995). Fluorescence energy transfer dye-labeled primers for DNA sequencing and analysis. *Proc Natl Acad Sci USA* 92, 4347–4351.

Kapranov, P., Cawley, S.E., Drenkow, J., Bekiranov, S., Strausberg, R.L., Fodor, S.P. and Gingeras, T.R. (2002). Large-scale transcriptional activity in chromosomes 21 and 22. *Science* 296, 916–919.

Kirkness, E.F., Bafna, V., Halpern, A.L., Levy, S., Remington, K., Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. et al. (2003). The dog genome: Survey sequencing and comparative analysis. *Science* 301, 1898–1903.

Klein, R.J., Zeiss, C., Chew, E.Y., Tsai, J.Y., Sackler, R.S., Haynes, C., Henning, A.K., SanGiovanni, J.P., Mane, S.M., Mayne, S.T. et al. (2005). Complement factor H polymorphism in age-related macular degeneration. *Science* 308, 385–389.

Kozal, M.J. et al. (1996). Extensive polymorphisms observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays. *Nat Med* 2, 753–759.

Kukita, Y., Miyatake, K., Stokowski, R., Hinds, D., Higasa, K., Wake, N., Hirakawa, T., Kato, H., Matsuda, T., Pant, K. et al. (2005). Genome-wide definitive haplotypes determined using a collection of complete hydatidiform moles. *Genome Res* 15, 1511–1518.

Kwok, P.Y. and Chen, X. (2003). Detection of single nucleotide polymorphisms. *Curr Issues Mol Biol* 5, 43–60.

Laitinen, T., Polvi, A., Rydman, P., Vendelin, J., Pulkkinen, V., Salmikangas, P., Makela, S., Rehn, M., Pirskanen, A., Rautanen, A. et al. (2004). Characterization of a common susceptibility locus for asthma-related traits. *Science* 304, 300–304.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.

Lee, L.G., Spurgeon, S.L., Heiner, C.R., Benson, S.C., Rosenblum, B.B., Menchen, S.M., Graham, R.J., Constantinescu, A., Upadhya, K.G. and Cassel, J.M. (1997). New energy transfer dyes for DNA sequencing. *Nucleic Acids Res* 25, 2816–2822.

Levy, S., Sutton, G., Ng, P.C., Feuk, L., Halpern, A.L., Walenz, B.P., Axelrod, N., Huang, J., Kirkness, E.F., Denisov, G. et al. (2007). The Diploid Genome Sequence of an Individual Human. *PLoS Biol* 5, e254.

Manaster, C., Zheng, W., Teuber, M., Wachter, S., Doring, F., Schreiber, S. and Hampe, J. (2005). InSNP: A tool for automated detection and visualization of SNPs and InDels. *Hum Mutat* 26, 11–19.

Maraganore, D.M., de Andrade, M., Lesnick, T.G., Strain, K.J., Farrer, M.J., Rocca, W.A., Pant, P.V., Frazer, K.A., Cox, D.R. and Ballinger, D.G. (2005). High-resolution whole-genome association study of Parkinson disease. *Am J Hum Genet* 77, 685–693.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z. et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380.

McCombie, W.R., Heiner, C., Kelley, J.M., Fitzgerald, M.G. and Gocayne, J.D. (1992). Rapid and reliable fluorescent cycle sequencing of double-stranded templates. *DNA Seq* 2, 289–296.

Mullikin, J.C. and Ning, Z. (2003). The phusion assembler. *Genome Res* 13, 81–90.

Mullis, K.B., Faloona, F.A., Scharf, S.J., Saiki, R.K., Horn, G.T. and Erlich, H.A. (1986). Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. *Cold Spring Harbor Symp Quant Biol* 51, 263–273.

Myers, E.W., Sutton, G.G., Delcher, A.L., Dew, I.M., Fasulo, D.P., Flanigan, M.J., Kravitz, S.A., Mobarry, C.M., Reinert, K.H., Remington, K.A. et al. (2000). A whole-genome assembly of Drosophila. *Science* 287, 2196–2204.

Nyren, P., Pettersson, B. and Uhlen, M. (1993). Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. *Anal Biochem* 208, 171–175.

Osoegawa, K., Mammoser, A.G., Wu, C., Frengen, E., Zeng, C., Catanese, J.J. and de Jong, P.J. (2001). A bacterial artificial chromosome library for sequencing the complete human genome. *Genome Res* 11, 483–496.

Patil, N., Berno, A.J., Hinds, D.A., Barrett, W.A., Doshi, J.M., Hacker, C.R., Kautzer, C.R., Lee, D.H., Marjoribanks, C., McDonough, D.P. et al. (2001). Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* 294, 1719–1723.

Prober, J.M., Trainor, G.L., Dam, R.J., Hobbs, F.W., Robertson, C.W., Zagursky, R.J., Cocuzza, A.J., Jensen, M.A. and Baumeister, K. (1987). A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* 238, 336–341.

Rand, V., Huang, J., Stockwell, T., Ferriera, S., Buzko, O., Levy, S., Busam, D., Li, K., Edwards, J.B., Eberhart, C. et al. (2005). Sequence survey of receptor tyrosine kinases reveals mutations in glioblastomas. *Proc Natl Acad Sci USA* 102, 14344–14349.

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W. et al. (2006). Global variation in copy number in the human genome. *Nature* 444, 444–454.

Rowntree, R.K. and Harris, A. (2003). The phenotypic consequences of CFTR mutations. *Ann Hum Genet* 67, 471–485.

Rozen, S. and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 132, 365–386.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual, 2nd ed*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74, 5463–5467.

Sanger, F., Coulson, A.R., Hong, G.F., Hill, D.F. and Petersen, G.B. (1982). Nucleotide sequence of bacteriophage lambda DNA. *J Mol Biol* 162, 729–773.

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S., Massa, H., Walker, M., Chi, M. et al. (2004). Large-scale copy number polymorphism in the human genome. *Science* 305, 525–528.

Shendure, J., Mitra, R.D., Varma, C. and Church, G.M. (2004). Advanced sequencing technologies: Methods and goals. *Nat Rev Genet* 5, 335–344.

Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D. and Church, G.M. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 309, 1728–1732.

Sladek, R., Rocheleau, G., Rung, J., Dina, C., Shen, L., Serre, D., Boutin, P., Vincent, D., Belisle, A., Hadjadj, S. et al. (2007). A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445, 881–885.

Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B. and Hood, L.E. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature* 321, 674–679.

Stephens, M., Sloan, J.S., Robertson, P.D., Scheet, P. and Nickerson, D.A. (2006). Automating sequence-based detection and genotyping of SNPs from diploid samples. *Nat Genet* 38, 375–381.

Strausberg, R.L., Simpson, A.J. and Wooster, R. (2003). Sequence-based cancer genomics: Progress, lessons and opportunities. *Nat Rev Genet* 4, 409–418.

Thomas, R.K., Nickerson, E., Simons, J.F., Janne, P.A., Tengs, T., Yuza, Y., Garraway, L.A., LaFramboise, T., Lee, J.C., Shah, K. et al. (2006). Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. *Nat Med* 12, 852–855.

Venter, J.C., Smith, H.O. and Hood, L. (1996). A new strategy for genome sequencing. *Nature* 381, 364–366.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A. et al. (2001). The sequence of the human genome. *Science* 291, 1304–1351.

CHAPTER

Genome-Wide Association Studies and Genotyping Technologies

Kevin V. Shianna

INTRODUCTION

Since the completion of the International HapMap Project (International HapMap Consortium, 2005; The International HapMap Consortium, 2003), a wealth of genetic information within the human genome has been uncovered. Over 12 million single nucleotide polymorphisms (SNPs) have been identified (Wheeler et al., 2007). This dense map of common genetic variation has revolutioni zed the field of human genetics. Genome-wide linkage scans can now be performed using efficient and cost-effective high-throughput SNP genotyping technologies, while at the same time achieving a fine map at a density not previously possible when performing earlier-generation genome scans based on microsatellite polymorphisms. The SNP map, along with recent technological gains, has also made it possible to comprehensively study complex diseases and traits by allowing for the genotyping of hundreds of thousands of SNPs for many samples. Also, with the identification of a specific type of SNP, called tagging SNPs, almost complete genomic coverage of common variation can be attained.

This chapter will provide an introduction to genome-wide association studies, followed by an overview of two commonly used genome-wide genotyping platforms, the Illumina Infinium Assay and the Affymetrix GeneChip. The methods and concepts discussed here complement those described elsewhere in this book.

PRINCIPLES OF GENOME-WIDE ASSOCIATION STUDIES

Before performing a genome-wide association study, it is essential to begin with a fully developed plan. The key issues to consider are the kind of study to perform, the required sample size, the importance of a well-defined phenotype, the need to correct for population stratification and the technologies available for genome-wide genotyping. This section will provide an introduction to the basic experimental design for a genome-wide association study.

Type of Study

First, the study should be focused on a key question based on a specific phenotype or disease. As an example, a study could be established to identify genetic variants associated with epilepsy. The key would be to collect a large cohort of individuals that have a definitive and well-defined diagnosis of epilepsy (cases) and compare these individuals to a group that does not have epilepsy (controls). In the past it was always necessary to collect matched controls, but with an abundance of control samples containing data from the various genome-wide array-based studies, it is now possible to use control genotyping data from the available databases. One example is the database made available by Illumina (iControlDB) where individual users have deposited genotyping data for control samples from previously processed samples. It has been demonstrated that the use of general control

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

TABLE 4.1 Examples of successful genome-wide association studies

Disease	Chip type	Gene/locus	Sample size (cases/controls)	Reference
HIV	Illumina HumanHap 550	HLA-C, HCP5/HLA-B*5701	486	(Fellay et al., 2007)
Type 2 Diabetes	Illumina HumanHap 300	TCF7L2, SLC30A8, HHEX, FTO, PPARG and KCNJ11	1161/1174	(Scott et al., 2007)
	Affymetrix Human Mapping 500	CDKN2A, CDKN2B, IGF2BP2, CDKAL1	1464/1467	(Saxena et al., 2007)
	Illumina HumanHap 300	TCF7L2, AHI1-LOC441171 region	500/497	(Salonen et al., 2007)
	Illumina HumanHap 300	TCF7L2, IDE–KIF11–HHEX and EXT2–ALX4	1363	(Sladek et al., 2007)
Type 1 Diabetes	Illumina HumanHap 550	KIAA0350	563/1146	(Hakonarson et al., 2007)
	Affymetrix Human Mapping 500	Chromosome regions 18q22, 12q24, 12q13, 16p13 and 18p11	4000/5000	(Todd et al., 2007)
	Affymetrix Human Mapping 100	IFIH1	2029/1775	(Smyth et al., 2006)
Restless leg syndrome	Affymetrix Human Mapping 500	MEIS1, BTBD9	401/1644	(Winkelmann et al., 2007)
Colorectal cancer	Illumina HumanHap 550	8q24.21 locus	930/960	(Tomlinson et al., 2007)
	Affymetrix Human Mapping 100	8q24 locus	1257/1336	(Zanke et al., 2007)
Prostate cancer	Illumina HumanHap 300	TCF2	1501/11290	(Gudmundsson et al., 2007)
	Illumina HumanHap 300	8q24	1453/3064	(Gudmundsson et al., 2007)
	Illumina HumanHap 300	8q24	1172/1157	(Yeager et al., 2007)
Atrial fibrillation	Illumina HumanHap 300	Chromosome region 4q25 near PITX2	550/4476	(Gudbjartsson et al., 2007)
Celiac disease	Illumina HumanHap 300	IL2, IL21	778/1422	(van Heel et al., 2007)
Crohn's disease	Illumina HumanHap 300	IL23R	2877/1345	(Tremelling et al., 2007)
	Illumina HumanHap 300	IL23R	567/571	(Duerr et al., 2006)
Macular degeneration	Affymetrix Human Mapping 100	CFH	96/50	(Klein et al., 2005)
Asthma	Illumina HumanHap 300	ORMDL3	994/1243	(Moffatt et al., 2007)

samples across studies is feasible (Wellcome Trust Case Control Consortium, 2007). This greatly decreases the cost to run a project and potentially allows for the inclusion of more cases (if available). A study could also be arranged to look at a specific continuous phenotype or quantitative trait within a group of individuals. Using the previous example, one could study how individuals with epilepsy responded to a specific drug or series of drugs.

Sample Size

The sample sizes required for a genome-wide association study are dependent on several factors. The first is the effect size of an unknown genetic variable. Of course, this is impossible to know in advance; however, power calculations can be performed to determine how many samples would be required to identify a variant with a specific effect size. These calculations can serve as a guide when deciding how many samples to include in the study. Purcell and colleagues (2003) have created a web-based genetic power calculator (http://pngu.mgh.harvard.edu/~purcell/gpc/) that can be used to determine power of a study. Another factor that is important when determining sample size is the type of study: case/control or a continuous phenotype. The current view for case/control studies is that large sample sizes are needed (1,000 of each), but this is not always true. A few studies have identified associated variants with only 150–200 samples (Table 4.1). Again, it is impossible to know the effect size of an unknown variant so it is best to include a large number of samples when designing a study. For a continuous phenotype, studies have had success while using fewer samples (Table 4.1).

Importance of Phenotype

The most important aspect when designing a genome-wide association study is the quality and discrete nature of the phenotype. It is essential that the studied phenotype is evaluated carefully for accuracy and is well defined across all samples. The inclusion of incorrectly phenotyped samples in a study could lead to a decrease in the statistical power to identify associated variants. This could consist of control samples that are in fact cases, but have yet to be identified with the specific phenotype/diagnosis or a misdiagnosis

of an individual within the cases. For some complex diseases, the phenotypic characteristics overlap with other closely related diseases. This can make it difficult when choosing which samples should be included in the study and, because of genetic heterogeneity, will decrease the power of the study.

Population Stratification

When analyzing the data from genome-wide association studies, it is essential to evaluate and account for population stratification between cases and controls. Cryptic population structure can generate false positive and/or negative associations. This can happen if allele frequencies are different in any unknown subgroups. One solution to correct for this difference in ancestry is described by Price and colleagues (2006). Left uncorrected, population stratification can lead directly to false associations.

Genotyping Technologies, Tagging SNPs and Genomic Coverage

Another necessity for genome-wide association studies is to have a method to genotype hundreds of thousands of genetic variants around the genome. Fortunately, there are now standard products that allow for screening of very large numbers of SNPs. Depending on the chosen platform, the SNPs on the arrays have been selected based either on even spacing or on the basis of certain SNPs being able to represent multiple SNPs. The SNPs in the second approach are termed "tagging SNPs" and greatly increase the genomic coverage that can be obtained compared to SNPs that are evenly spaced across the genome (Barrett and Cardon, 2006; Pe'er et al., 2006). Reflecting now the well-established patterns of linkage disequilibrium (Hinds et al., 2005; International HapMap Consortium, 2005), these tagging SNPs work by acting as proxies for other nearby associated SNPs, making it possible to genotype only the tagging SNP and yet capture the content of other associated SNPs in the region. There are currently commercially available arrays (SNP chips) that contain close to 1 million tagging SNPs, representing most of the common variation in the human genome. Products where the SNPs have not been specifically chosen based on the close association with other SNPs will have much less genomic coverage, because the SNPs have not been optimally chosen for the ability to tag. For example, Illumina's HumanHap300 array, containing around 300,000 tagging SNPs, has essentially the same statistical genomic coverage as Affymetrix's Mapping 500K Array Set, which contains 500,000 evenly spaced SNPs (Barrett and Cardon, 2006; Pe'er et al., 2006). Further background information on genome-wide association studies and tagging SNPs has been elegantly presented elsewhere (Hirschhorn and Daly, 2005).

PLATFORM OVERVIEW

Illumina Infinium Assay

The Illumina genotyping technology is based on 3 μ silicon beads that have hundreds of thousands of DNA oligomers attached to

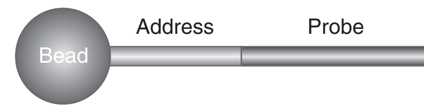


Figure 4.1 An example of a bead on the Illumina Infinium BeadChip. A schematic of a single bead, containing a 75-nucleotide oligomer consisting of an address and a specific probe sequence, as described in the text.

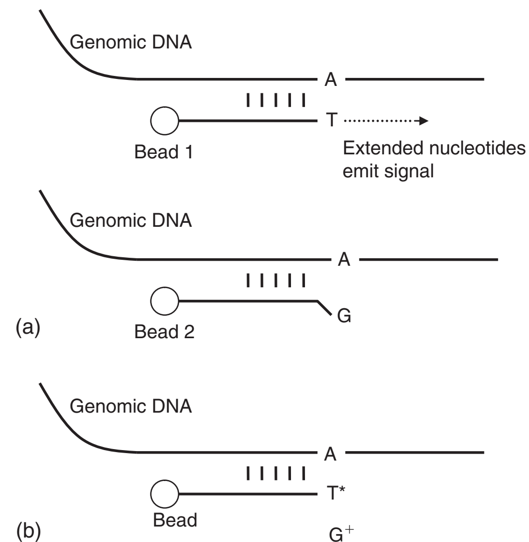


Figure 4.2 Infinium I and II assay. (a) A graphical representation of the Infinium I assay. This assay uses an ASPE and requires two bead types to detect a single SNP. (b) A depiction of an SBE reaction as used for the Infinium II assay. Only one bead type is required to detect a single SNP. The diagram shows a single dideoxy base (T) being added to the 50-base oligomer on the bead.

the surface (Shen et al., 2005). For Illumina's Infinium whole-genome genotyping BeadChip, the attached oligomers consist of a stretch of 75 nucleotides (Gunderson et al., 2005; Steemers et al., 2006) (Figure 4.1). The first 25 nucleotides (the address) serve as a barcode to identify the location of specific beads on the array, and the next 50 bases (the probe) are complementary to a genomic region containing a specific SNP (or copy number variant). On average, there is an 18-fold redundancy of each bead type on the BeadChip. The placement of the beads on the BeadChip is a random process, thus requiring a decoding step. The decoding involves multiple hybridizations to the 25 oligomer address. This not only identifies the location of each bead, but performs a functional quality control for each bead on the BeadChip.

There are two Infinium chemistries, type I and II (Figure 4.2). The Infinium I assay uses an allele-specific primer extension (ASPE) reaction in which fluorescently labeled nucleotides are inserted at the 3′ end of the primer. With this assay it takes two bead types to determine the genotype at a single SNP. The Infinium II assay involves a single-base extension (SBE) reaction that occurs on a single bead type for each SNP. Most of Illumina's off-the-shelf BeadChips use the Infinium II assay.

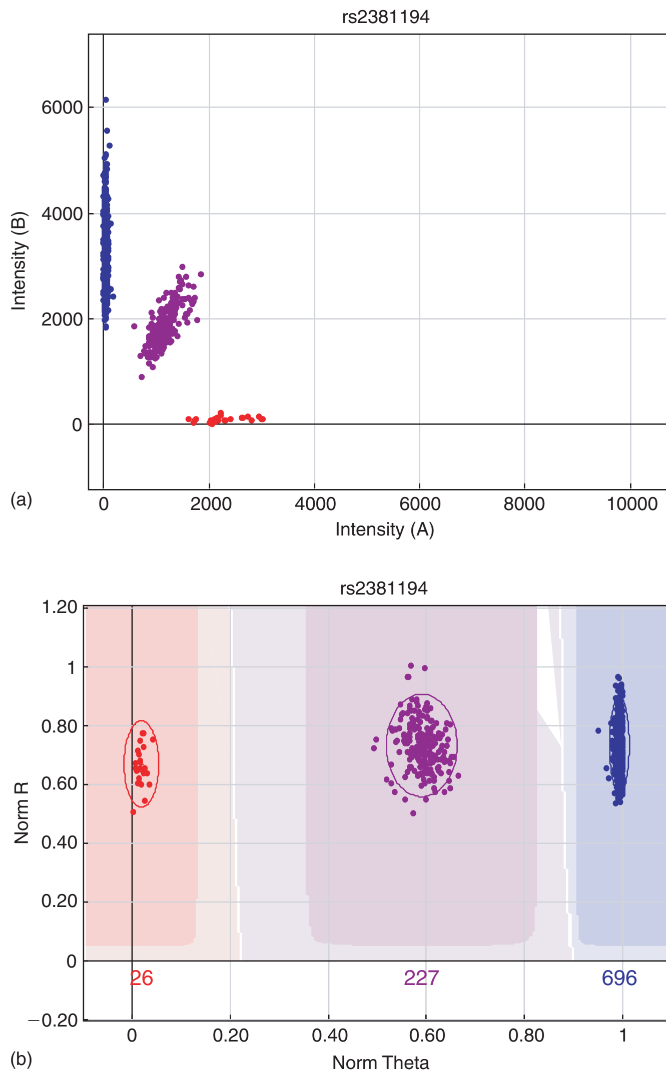


Figure 4.3 Raw and normalized data for SNP genotyping, using the Illumina platform. (a) Raw data from genotype analysis of one SNP (rs2381194) from 949 individuals. Each dot corresponds to a single individual. Blue, homozygote for one allele ($n = 696$); Purple, heterozygous ($n = 227$); Red, homozygote for other allele ($n = 26$). (b) Same SNP and individuals as in (a), but after normalization using Illumina's BeadStudio software.

The steps for the Infinium II assay are as follows: (1) 750 ng of intact genomic DNA is whole-genome amplified (WGA); (2) the amplified product is fragmented; (3) the fragmented DNA is hybridized to a BeadChip; (4) an SBE reaction is performed directly on the chip with the insertion of a labeled (2,4-dinitrophenol or biotin) dideoxy base; (5) the inserted base is stained with a fluorescently labeled antibody and (6) fluorescence on the beads on the BeadChip is detected by scanning with the Illumina BeadStation. An average of the raw intensity per bead type is then imported into the Illumina BeadStudio software, where the data can be efficiently managed and individual SNPs can be automatically or manually called. The quality of the raw and normalized genotype calls can be visually confirmed within this software (Figure 4.3).

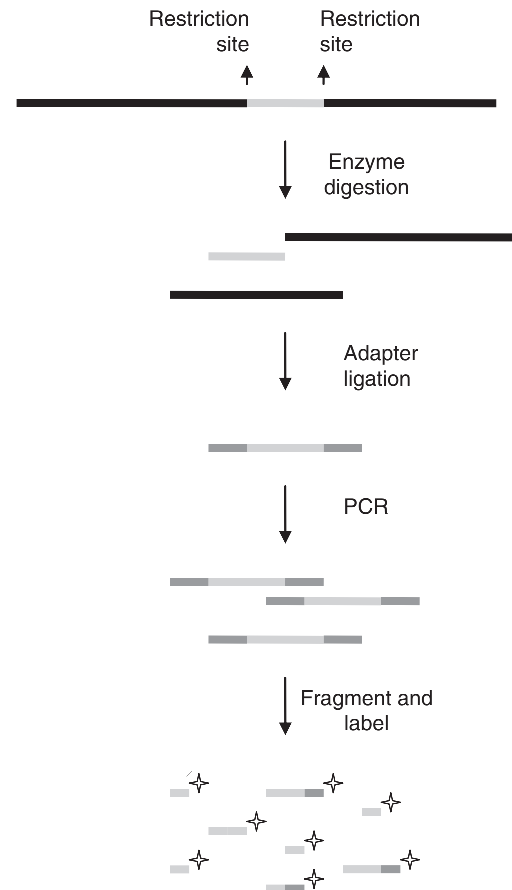


Figure 4.4 Overview of DNA processing steps for the AffymetrixGeneChip. The diagram demonstrates the steps implemented to reduce the genomic complexity allowing for efficient allele-specific hybridization.

Affymetrix GeneChip

The Affymetrix genotyping GeneChip is manufactured using the same principles as for the expression GeneChip, using a technology called photolithography. This has been described in great detail (Dalma-Weiszhausz et al., 2006; Matsuzaki et al., 2004), including elsewhere in this book. However, the genotyping GeneChip contains additional probes for each interrogated SNP, to allow for differentiation of true signal over noise (Dalma-Weiszhausz et al., 2006).

The steps required to process a DNA sample in preparation for hybridization to the GeneChip lead to a reduction in genome complexity allowing for allele-specific hybridization (Figure 4.4). The following is the procedure for processing a sample: (1) perform separate genomic DNA (250 ng) digestions using the *NspI* and *StyI* restriction enzymes; (2) ligate specific adapters containing a single universal primer binding site to the digested DNA; (3) pool the DNA fragments from the separate digestions and amplify using a single universal PCR primer specific to the primer site within the adaptor; (4) fragment the amplified products; (5) end label the DNA; (6) hybridize to a GeneChip and (7) scan the chip. The raw data can then be converted into genotype calls by using the appropriate algorithms (from Affymetrix).

Platform Comparison

Numerous laboratories worldwide have successfully used both the Illumina and Affymetrix platforms to identify genetic variants involved with complex disease (Table 4.1). This section will describe potential benefits of each platform, paying particular attention to (1) differences in the basic design of the chips; (2) statistical tagging power (genomic coverage); (3) ability to detect copy number variation (CNV); (4) processing issues and (5) overall quality of the final raw genotyping data.

The length of the probes to detect an SNP during hybridization constitutes one of the major differences between the platforms. The probe length on the Affymetrix GeneChip is 25 nucleotides, while on the Illumina BeadChip the probes contain 50 nucleotides. This difference in length for hybridization provides for higher specificity for probes on the Illumina BeadChip. The other main difference is the content of the probe for each SNP. For Affymetrix, each SNP is represented by a series of probes consisting of mismatches and perfect homology, allowing for differentiation between signal and noise. The Illumina BeadChips do not require specific mismatch probes for every SNP because of the increased length of the probes for hybridization, but contain control beads that detect the overall level of background noise. The probe length constitutes the major difference between the two platforms and is most likely directly linked to the differences in quality between the platforms.

Most of the SNPs on the Illumina Infinium BeadChips were selected as tagging SNPs, resulting in excellent genomic coverage for most populations (based on HapMap populations) (International HapMap Consortium, 2005). On the other hand, the Affymetrix GeneChips are based on SNPs that are evenly spaced throughout the genome. The positive aspect of even spacing is better physical coverage for detecting CNVs (Redon et al., 2006). One disadvantage of this approach is less statistical power to capture the known common variation when compared to arrays that contain tagging SNPs. A multimarker or haplotype-tagging approach may gain increased power/coverage, but the downstream statistical analyses become more difficult. Multiple groups (Barrett and Cardon, 2006; Magi et al., 2007; Pe'er et al., 2006) have performed statistical comparisons of the Illumina HumanHap300 and the HumanHap550 BeadChips with the Affymetrix Mapping 500 K Set. From these analyses it is evident that the Illumina array with only around 300,000 tagging SNPs has essentially the same statistical genomic coverage as the Affymetrix Mapping 500 K Set, while the Illumina HumanHap550 array is statistically superior for genome-wide association studies.

Overall, both platforms provide good quality results (Table 4.1). However, a comparison of the SNP conversion rate (percentage of SNPs that work) between the two technologies reveals a significant difference. In our hands, for example, when using the Illumina pre-defined cluster, an average sample call rate of 99.6% is achieved, with an average reproducibility of over 99.99%. With the Affymetrix platform, the average sample call rate is variable and depends greatly on the chosen stringency within the calling algorithm, but a typical call rate would be in the range of 95–98% (personal communications). As discussed above, the increased SNP conversion rate with the Illumina BeadChip is most likely related to the size of the probe (50 nucleotides) which increases the specificity during the hybridization step.

Recently, there has been a push to detect CNVs throughout the human genome by using genome-wide genotyping chips (Komura et al., 2006; Redon et al., 2006) (see Chapter 5). Using these SNP-based arrays is an efficient method to detect large deletions or duplications. The Affymetrix GeneChip with its even spacing of SNPs has great genomic coverage to detect large CNVs. Most of the BeadChips from Illumina have specific SNPs or non-polymorphic probes added in regions where there are known CNVs. In general, both arrays are capable of detecting large CNVs. However, with the current SNP densities, the detection of small CNVs is difficult with the use of SNP-based arrays.

Raw Data Quality Control

Genetic association studies using genome-wide genotyping chips produce very large amounts of data. A typical project with 2,000 samples, each containing over 500,000 SNPs per sample, will result in over one billion data points. With this much data, the probability for false positives becomes much greater. A few reasons that can attribute to this are samples being prepared using different purification methods and the processing of the chips in different labs. These slight variations can lead to minor differences in intensity, thus making it more difficult to accurately call the genotypes. To prevent false positives from clouding the results from the statistical analyses, it is of utmost importance to begin with a clean or curated data set. This is accomplished by establishing a high threshold of sample and SNP quality. The number of SNPs that are obtained for each sample (the "call rate") and the raw fluorescent intensity are the main variables that can be analyzed to determine whether a sample should be included or excluded. If any sample's call rate is much lower (>1–2%) than the average call rate for the project, the sample should be excluded. A 1% difference can represent as many as 5,000 fewer SNPs being called in a sample when compared with the average number of SNP calls in a population of samples.

There are many variables that can be used to assess SNP quality. Metrics such as call frequency (number of samples that are being called for an SNP), cluster separation (degree of split between the heterozygote grouping and homozygotes, e.g. Figure 4.3), extreme deviation from Hardy–Weinberg proportions and overall raw signal intensity should be evaluated for all SNPs. A very aggressive approach to maintaining a clean data set is to adhere to a rigorous level of acceptable SNP call frequency. We have established a high threshold for SNP inclusion and have coined the phrase the "1% rule"; we delete any SNP that has greater than 1% of the samples not being called or called ambiguously. This rule was established from personal experiences when attempting to extract as much genotype data as possible. After performing statistical analyses, we noticed a large number of false positives due entirely to poor genotyping calls. For the 1% rule to work effectively, each new experiment needs to be reclustered using its own raw data. With the 1% rule, the use of predetermined cluster files will not work, because slight variations

from experiment to experiment can result in slight shifts of intensity and cluster separation (i.e., separation of heterozygote and homozygote clusters), resulting in incomplete calling of truly callable samples for a large number of SNPs.

When applying the 1% rule, it is a balance between including a greater number of samples or a greater number of SNPs. Attempting to keep good but lower quality samples can result in the deletion of many SNPs that fail to meet the 1% rule standard. By deleting a few extra samples across the entire data set, it is possible to keep many of the SNPs that would have violated the 1% rule. In practice, in our experience, a typical project will have between 1% and 3% samples deleted (not all because of 1% rule) across all SNPs and between 1% and 2% of SNPs deleted because of violating the 1% rule. We have successfully used this approach for multiple studies (Fellay et al., 2007).

After completing the first pass of statistical tests, the SNPs with the lowest 100–200 *p*-values should be manually checked for genotyping data quality. Visualization of the raw/normalized genotype call can increase one's confidence in any potentially associated SNP with a low *p*-value. Inspecting the data is especially important if a data set has not been stringently curated under the 1% rule. For a given SNP, if only 95% of samples are called, over 50 samples from a sample set of 1,000 samples will not have a genotype call. Across a very large data set (over 500,000 SNPs), this loosened threshold can result in many false positives solely because of the increased opportunity for an imbalance of the case/control ratio in the uncalled samples.

Sample Collection, Processing and Throughput

It is now feasible for institutional core facilities or small- to medium-sized labs to run hundreds of samples per week using genome-wide genotyping chips. For each platform described here, there are off-the-shelf kits allowing for easy processing of many chips/samples at once with minimal variation across experiments. Also, with the implementation of automation developed specifically for each assay, sample throughput and tracking have all been simplified.

Genome-wide genotyping chips require less than 1 μg of DNA, a very low amount considering the amount of data generated from this sample. In general, high-quality genomic DNA should be used. DNA of lesser quality or WGA DNA can be used, but will result in a lower sample call rate. The Affymetrix platform is more amenable to the use of WGA DNA, because of the lack of an initial WGA step as in the Illumina Infinium assay. However, WGA DNA will work for the Infinium platform, although lower call rates will be achieved (personal communication). Formalin-fixed, paraffin-embedded (FFPE) tissue samples represent a rich and abundant source of well-annotated material, but DNA isolated from FFPE samples using currently available methods is not amenable to genome-wide genotyping assays. The main reason for this is that the interaction of the formalin with the DNA prevents efficient enzymatic amplification of the DNA. Creating methods to successfully perform whole-genome genotyping on FFPE samples will be of great importance. This will benefit current projects by allowing for the potential increase in sample sizes. Also, precious samples obtained only via FFPE tissues could then be included in genome-wide association studies.

2009 UPDATE

Since the completion of this chapter, there has been an exponential increase in the number of published genome-wide association studies. Examples of some of the major findings include the identification of associated variants with prostate cancer (Gudmundsson et al., 2008; Eeles et al., 2008; Thomas et al., 2008), type 2 diabetes (Yasuda et al., 2008; Unoki et al., 2008; Bouatia-Naji et al., 2009; Lyssenko et al., 2009; Prokopenko et al., 2009), lung cancer (Thorgeirsson et al., 2008; Hung et al., 2008; Amos et al., 2008) and obesity (Thorleifsson et al., 2009; Willer et al., 2009; Loos et al., 2008). The raw data for some of these projects as well as others may be accessible at the dbGaP database (http://www.ncbi.nlm.nih.gov/gap).

As for the genotyping platforms, the rapid advances that have occurred over the past few years with the genome-wide genotyping chips have now stabilized. This is positive for researchers who are now pursuing a genome-wide association study or are screening for copy number variants (CNVs). The current options for the most complete genomic coverage (SNP and CNV) are the Illumina Human 1M-Duo BeadChip and the Affymetrix Human Genome-Wide SNP Array 6.0. Both of these platforms contain tagging SNPs to provide greater coverage of the common variation across the genome. Specific probes have also been added to detect known CNVs and to cover regions of the genome that until now were sparsely covered by probes.

Looking forward, the creation of a new genome-wide genotyping chip with increased SNP content will not add much value or power to most studies screening for common variants associated with a specific phenotype. However, there will be a great need for platforms that allow for the screening of tens of thousands SNPs specific to one's phenotype. This type of custom chip will be a valuable component of the initial next-generation sequencing projects in which small subsets of samples (25-50) are sequenced. Any newly identified single nucleotide variants/CNVs can then be screened across a larger set of relevant samples using the custom designed SNP chip. This type of two-step approach will be useful until it becomes feasible to simply and cost-effectively sequence the entire genomes of thousands of individuals.

2009 UPDATE REFERENCES

Amos, C.I. et al. (2008). Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet* 40, 616–622.

Bouatia-Naji, N. et al. (2009). A variant near MTNR1B is associated with increased fasting plasma glucose levels and type 2 diabetes risk. *Nat Genet* 41, 89–94.

Eeles, R.A. et al. (2008). Multiple newly identified loci associated with prostate cancer susceptibility. *Nat Genet* 40, 316–321.

Gudmundsson, J. et al. (2008). Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. *Nat Genet* 40, 281–283.

Hung, R.J. et al. (2008). A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* 452, 633–637.

Loos, R.J. et al. (2008). Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat Genet* 40, 768–775.

Lyssenko, V. et al. (2009). Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion. *Nat Genet* 41, 82–88.

Prokopenko, I. et al. (2009). Variants in MTNR1B influence fasting glucose levels. *Nat Genet* 41, 77–81.

Thomas, G. et al. (2008). Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet* 40, 310–315.

Thorgeirsson, T.E. et al. (2008). A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* 452, 638–642.

Thorleifsson, G. et al. (2009). Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet* 41, 18–24.

Unoki, H. et al. (2008). SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat Genet* 40, 1098–1102.

Willer, C.J. et al. (2009). Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet* 41, 25–34.

Yasuda, K. et al. (2008). Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nat Genet* 40, 1092–1097.

REFERENCES

Barrett, J.C. and Cardon, L.R. (2006). Evaluating coverage of genome-wide association studies. *Nat Genet* 38, 659–662.

Dalma-Weiszhausz, D.D., Warrington, J., Tanimoto, E.Y. and Miyada, C.G. (2006). The affymetrix GeneChip platform: an overview. *Methods Enzymol* 410, 3–28.

Duerr, R.H., Taylor, K.D., Brant, S.R., Rioux, J.D., Silverberg, M.S., Daly, M.J., Steinhart, A.H., Abraham, C., Regueiro, M., Griffiths, A., Dassopoulos, T., Bitton, A., Yang, H., Targan, S., Datta, L.W., Kistner, E.O., Schumm, L.P., Lee, A.T., Gregersen, P.K., Barmada, M.M., Rotter, J.I., Nicolae, D.L. and Cho, J.H. (2006). A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* 314, 1461–1463.

Fellay, J., Shianna, K.V., Ge, D., Colombo, S., Ledergerber, B., Weale, M., Zhang, K., Gumbs, C., Castagna, A., Cossarizza, A. et al. (2007). A Whole-Genome Association Study of Major Determinants for Host Control of HIV-1. *Science* 317, 944–947.

Gudbjartsson, D.F., Arnar, D.O., Helgadottir, A., Gretarsdottir, S., Holm, H., Sigurdsson, A., Jonasdottir, A., Baker, A., Thorleifsson, G., Kristjansson, K. et al. (2007). Variants conferring risk of atrial fibrillation on chromosome 4q25. *Nature* 448, 353–357.

Gudmundsson, J., Sulem, P., Manolescu, A., Amundadottir, L.T., Gudbjartsson, D., Helgason, A., Rafnar, T., Bergthorsson, J.T., Agnarsson, B.A., Baker, A. et al. (2007). Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet* 39, 631–637.

Gudmundsson, J., Sulem, P., Steinthorsdottir, V., Bergthorsson, J.T., Thorleifsson, G., Manolescu, A., Rafnar, T., Gudbjartsson, D., Agnarsson, B.A., Baker, A. et al. (2007). Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet* 39, 977–983.

Gunderson, K.L., Steemers, F.J., Lee, G., Mendoza, L.G. and Chee, M.S. (2005). A genome-wide scalable SNP genotyping assay using microarray technology. *Nat Genet* 37, 549–554.

Hakonarson, H., Grant, S.F., Bradfield, J.P., Marchand, L., Kim, C.E., Glessner, J.T., Grabs, R., Casalunovo, T., Taback, S.P., Frackelton, E.C. et al. (2007). A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature* 448, 591–594.

Hinds, D.A., Stuve, L.L., Nilsen, G.B., Halperin, E., Eskin, E., Ballinger, D.G., Frazer, K.A. and Cox, D.R. (2005). Whole-genome patterns of common DNA variation in three human populations. *Science* 307, 1072–1079.

Hirschhorn, J.N. and Daly, M.J. (2005). Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 6, 95–108.

International HapMap Consortium (2003). The International HapMap Project. *Nature* 426, 789–796.

International HapMap Consortium (2005). A haplotype map of the human genome. *Nature* 437, 1299–1320.

Klein, R.J., Zeiss, C., Chew, E.Y., Tsai, J.Y., Sackler, R.S., Haynes, C., Henning, A.K., SanGiovanni, J.P., Mane, S.M., Mayne, S.T. et al. (2005). Complement factor H polymorphism in age-related macular degeneration. *Science* 308, 385–389.

Komura, D., Shen, F., Ishikawa, S., Fitch, K.R., Chen, W., Zhang, J., Liu, G., Ihara, S., Nakamura, H., Hurles, M.E. et al. (2006). Genome-wide detection of human copy number variations using high-density DNA oligonucleotide arrays. *Genome Res* 16, 1575–1584.

Magi, R., Pfeufer, A., Nelis, M., Montpetit, A., Metspalu, A. and Remm, M. (2007). Evaluating the performance of commercial whole-genome marker sets for capturing common genetic variation. *BMC Genomics* 8, 159.

Matsuzaki, H., Loi, H., Dong, S., Tsai, Y.Y., Fang, J., Law, J., Di, X. et al. (2004). Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array. *Genome Res* 14, 414–425.

Moffatt, M.F., Kabesch, M., Liang, L., Dixon, A.L., Strachan, D., Heath, S., Depner, M., von Berg, A., Bufe, A., Rietschel, E. et al. (2007).

Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 448, 470–473.

Pe'er, I., de Bakker, P.I., Maller, J., Yelensky, R., Altshuler, D. and Daly, M.J. (2006). Evaluating and improving power in whole-genome association studies using fixed marker sets. *Nat Genet* 38, 663–667.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38, 904–909.

Purcell, S., Cherny, S.S. and Sham, P.C. (2003). Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 19, 149–150.

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H. et al. (2006). Global variation in copy number in the human genome. *Nature* 444, 444–454.

Salonen, J.T., Uimari, P., Aalto, J.M., Pirskanen, M., Kaikkonen, J., Todorova, B., Hypponen, J., Korhonen, V.P., Asikainen, J., Devine, C. et al. (2007). Type 2 Diabetes Whole-Genome Association Study in Four Populations: The DiaGen Consortium. *Am J Hum Genet* 81, 338–345.

Saxena, R., Voight, B.F., Lyssenko, V., Burtt, N.P., de Bakker, P.I., Chen, H., Roix, J.J., Kathiresan, S., Hirschhorn, J.N., Daly, M.J. et al. (2007). Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316, 1331–1336.

Scott, L.J., Mohlke, K.L., Bonnycastle, L.L., Willer, C.J., Li, Y., Duren, W.L., Erdos, M.R., Stringham, H.M., Chines, P.S., Jackson, A.U. et al. (2007). A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316, 1341–1345.

Shen, R., Fan, J.B., Campbell, D., Chang, W., Chen, J., Doucet, D., Yeakley, J. et al. (2005). High-throughput SNP genotyping on universal bead arrays. *Mutat Res* 573, 70–82.

Sladek, R., Rocheleau, G., Rung, J., Dina, C., Shen, L., Serre, D., Boutin, P., Vincent, D., Belisle, A., Hadjadj, S. et al. (2007). A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445, 881–885.

Smyth, D.J., Cooper, J.D., Bailey, R., Field, S., Burren, O., Smink, L.J., Guja, C., Ionescu-Tirgoviste, C., Widmer, B., Dunger, D.B. et al. (2006). A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. *Nat Genet* 38, 617–619.

Steemers, F.J., Chang, W., Lee, G., Barker, D.L., Shen, R. and Gunderson, K.L. (2006). Whole-genome genotyping with the single-base extension assay. *Nat Methods* 3, 31–33.

Todd, J.A., Walker, N.M., Cooper, J.D., Smyth, D.J., Downes, K., Plagnol, V., Bailey, R., Nejentsev, S., Field, S.F., Payne, F. et al. (2007). Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet* 39, 857–864.

Tomlinson, I., Webb, E., Carvajal-Carmona, L., Broderick, P., Kemp, Z., Spain, S., Penegar, S., Chandler, I., Gorman, M., Wood, W. et al. (2007). A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet* 39, 984–988.

Tremelling, M., Cummings, F., Fisher, S.A., Mansfield, J., Gwilliam, R., Keniry, A., Nimmo, E.R., Drummond, H., Onnie, C.M., Prescott, N.J. et al. (2007). IL23R variation determines susceptibility but not disease phenotype in inflammatory bowel disease. *Gastroenterology* 132, 1657–1664.

van Heel, D.A., Franke, L., Hunt, K.A., Gwilliam, R., Zhernakova, A., Inouye, M., Wapenaar, M.C., Barnardo, M.C., Bethel, G., Holmes, G.K. et al. (2007). A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet* 39, 827–829.

Wellcome Trust Case Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661–678.

Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M. et al. (2007). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 35, D5–D12.

Winkelmann, J., Schormair, B., Lichtner, P., Ripke, S., Xiong, L., Jalilzadeh, S., Fulda, S., Putz, B., Eckstein, G., Hauk, S. et al. (2007). Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. *Nat Genet* 39, 1000–1006.

Yeager, M., Orr, N., Hayes, R.B., Jacobs, K.B., Kraft, P., Wacholder, S., Minichiello, M.J., Fearnhead, P., Yu, K., Chatterjee, N. et al. (2007). Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet* 39, 645–649.

Zanke, B.W., Greenwood, C.M., Rangrej, J., Kustra, R., Tenesa, A., Farrington, S.M., Prendergast, J., Olschwang, S., Chiang, T., Crowdy, E. et al. (2007). Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet* 39, 989–994.

CHAPTER

Copy Number Variation and Human Health

Charles Lee, Courtney Hyland, Arthur S. Lee, Shona Hislop and Chunhwa Ihm

INTRODUCTION

Early cytogenetic studies recognized that microscopically visible aberrations such as duplications or deletions of entire chromosomes (aneuploidy) were associated with specific congenital developmental disorders. For example, an extra copy of chromosome 21 (i.e., trisomy 21) has long been established as being correlative with the mongoloid phenotype that was first recognized in 1866 by John Langdon Down (Down 1866; Lejeune et al., 1959). However, recent development of high-resolution assays, capable of detecting small segmental genetic alterations in a genome-wide fashion, have led to the detection of widespread copy number variation (CNV) among the genomes of healthy individuals (Iafrate et al., 2004; Sebat et al., 2004). This finding has now sparked intense efforts to identify and characterize the extent of this type of genetic variation in human populations and to understand its impact on human health.

BASIC PRINCIPLES OF CNVs

CNVs have been operationally defined as genomic gains and losses of 1 kb or larger (Freeman et al., 2006; Feuk et al., 2006). This definition helps to differentiate it from other forms of polymorphism and/or repeated DNAs in the human genome (see Chapter 2), which include indels, microsatellites, minisatellites, simple repeats (e.g., dinucleotide repeats, trinucleotide repeats, etc.), telomeric and centromeric repetitive DNAs, and most interspersed repetitive elements (although LINES and other long interspersed repeats have repeat elements greater than 1 kb in size). In addition, since CNVs are considered subchromosomal imbalances, they are differentiated from whole chromosomal aneuploidies, such as trisomy 21 or monosomy X.

When CNV data are available for many individuals within or across different human populations, CNVs can be categorized into biallelic or multiallelic states (Figure 5.1). Biallelic CNVs have only two alleles and thus produce three different genotypes. For example, a CNV that exists solely in the form of a 1-copy allele or a 2-copy allele (Figure 5.1a, right) can produce diploid copy numbers of 2 (1 allele+1 allele), 3 (1 allele+2 allele), or 4 (2 allele+2 allele) (Figure 5.1a, left). CNVs with greater than two alleles are considered multiallelic and result in more than three different genotypes (Figure 5.1b and Figure 5.2).

Heritable CNVs are thought to arise from germline genomic rearrangements (or in some cases, possibly very early somatic events). The genomic rearrangements (mutational events) that are thought to cause CNVs can be broadly categorized as arising via one of two mechanisms: (i) Non-allelic homologous recombination (NAHR) – where homologous recombination

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

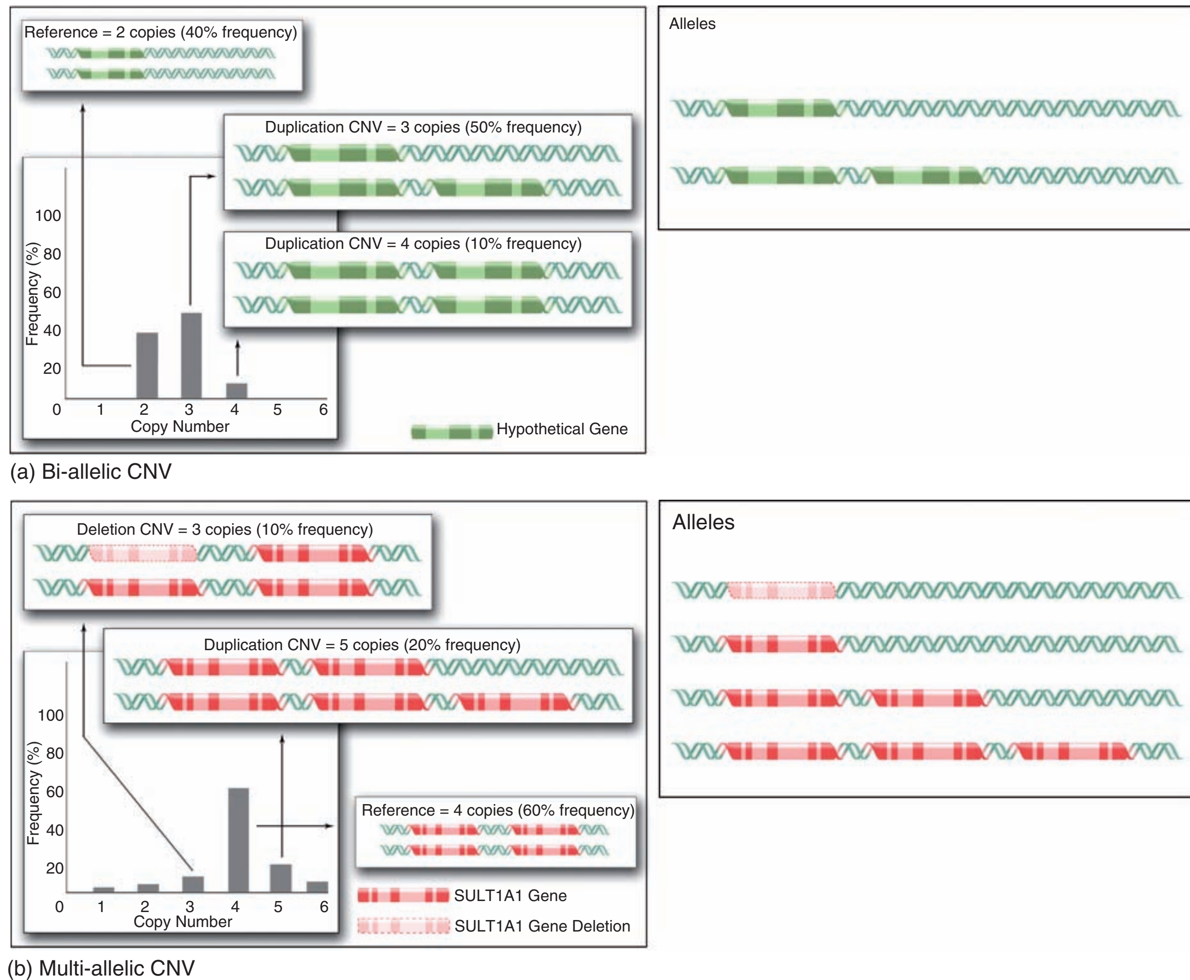


Figure 5.1 Examples of bi-allelic and multi-allelic CNVs. a) An example of a bi-allelic CNV that has a 1-copy allele and a 2-copy allele. The reference individual has two 1-copy alleles but 50% of individuals in this population have a total of three copies of this gene per cell. All bi-allelic CNVs have three genotypes per diploid cell, and in this case, copy numbers of 2, 3, and 4 per diploid cell. b) An example of a multi-allelic CNV that has a 0, 1, 2 and 3 copy alleles, resulting in six genotypes in this population. Only the allelic combinations for the three most common genotypes are shown.

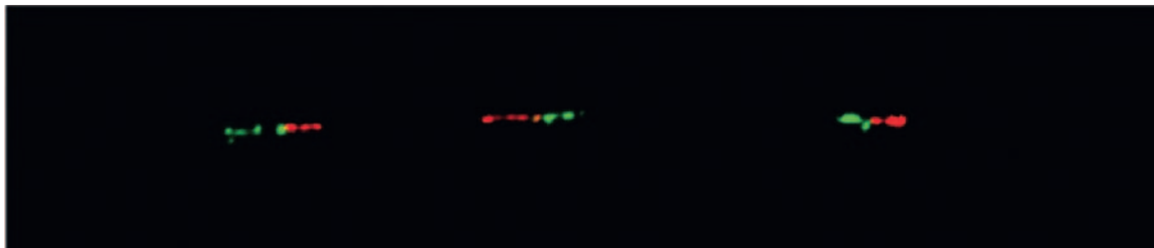


Figure 5.2 The amylase gene is considered a multiallelic CNV with diploid copy number estimates ranging from 2 to 15 among humans (Perry et al., 2007). Here, a two-color fiber-FISH experiment of the amylase gene (each copy of the gene is represented by a red+green combination) shows three copies on a DNA fiber from a single chromosome. It is interesting to note here that the second amylase gene copy is inverted relative to the others.

occurs between highly identical sequences in the genome, such as segmental duplications or related interspersed repetitive elements (Figure 5.3) and (ii) non-homologous end joining (NHEJ) – a repair mechanism whereby double-strand breaks that occur in the genome are ligated together with the assistance of specific protein complexes that form at the sites of the double-strand breaks (Figure 5.4). The rate for NHEJ is likely influenced significantly by environmental factors and localized DNA

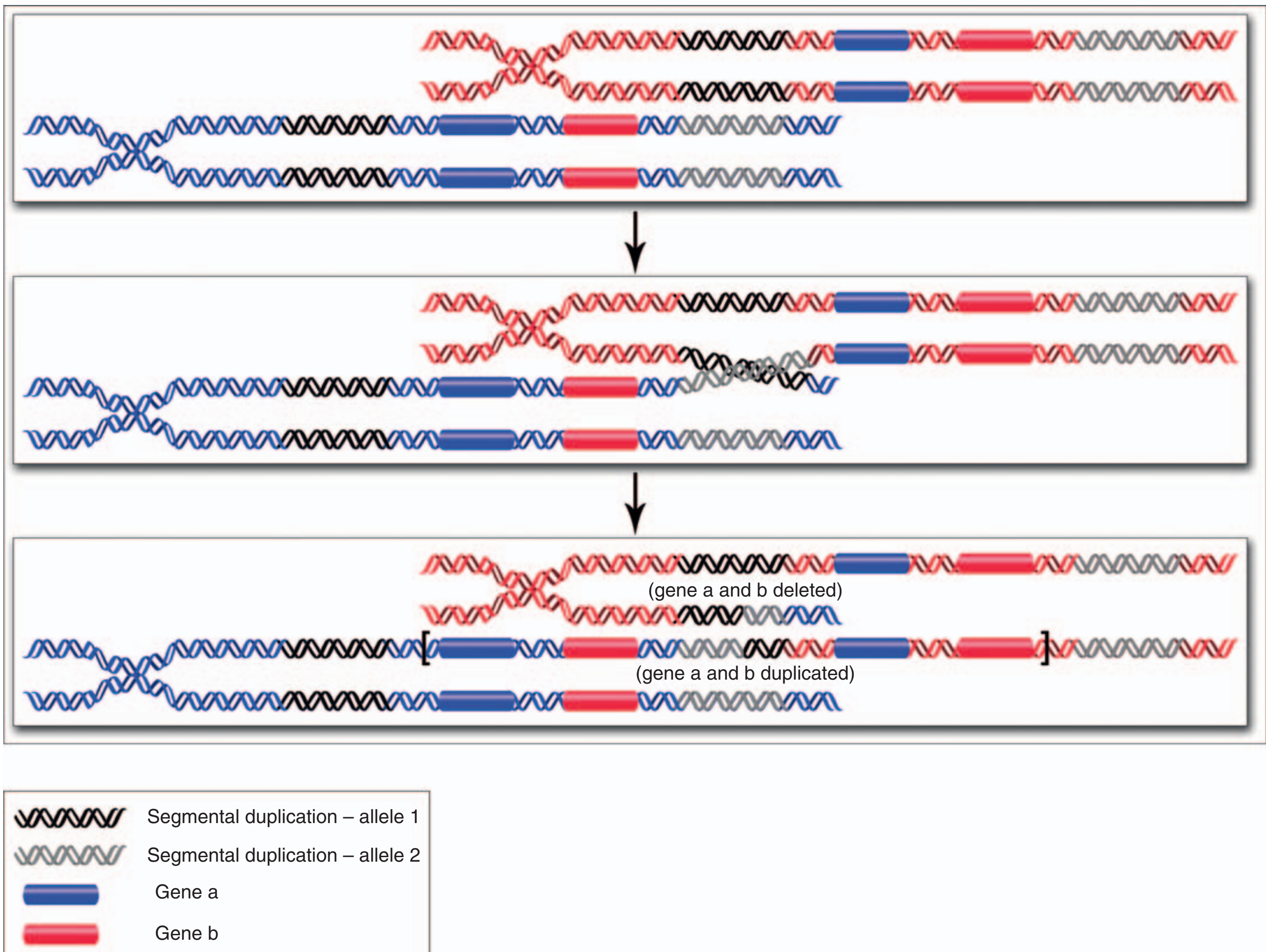


Figure 5.3 Non-allelic homologous recombination (NAHR) is a mechanism for generating CNVs where recombination between non-allelic repeats with >90% sequence homology (indicated by black- and gray-colored DNA segments). Intervening DNA sequences are deleted and duplicated on different chromatids.

conformations, but in general has been estimated to occur at a rate of less than 10^{-7} per generation, similar to the 10^{-8} per generation estimated mutation rate of single nucleotide polymorphisms (SNPs) (Conrad and Hurles, 2007). NAHR events are believed to occur more frequently, with estimates of up to 10^{-4} per locus per gamete per generation (Shaffer and Lupski 2000). Since NAHR events lead to duplication and deletion of DNA sequences that lie between highly identical sequences in the genome, this mutational event tends to be associated with larger CNVs (Redon et al., 2006).

CNVs are scattered throughout the human genome with at least 6% of a chromosome's total DNA content being potentially copy number variable (Redon et al., 2006). Taken together, current estimates are that more than 500 Mb (or 18.8%) of the reference human genome is copy number variable (Scherer et al., 2007). In fact, one recent study suggested that structural genomic variation (i.e., non-SNP variation that is predominantly in the form of CNVs) account for as much as 22% of all genetic variable events found in a given individual and total 74% of all variant bases (Levy et al., 2007). Many of the earlier CNV discovery projects relied on array comparative genomic hybridization (aCGH) technologies that utilized large-insert clones (on the order of 120kb–150kb each) as individuals probles and therefore yielded CNV data with ill-defined boundaries. This has likely led to a size over-estimation for many currently identified CNVs.

Documented CNVs are being cataloged and collated in several databases. For example, CNV data from healthy individuals can be found in the Database of Genomic Variants (http://projects.tcag.ca/variation) and the human paralogy database (http://humanparalogy.gs.washington.edu/structuralvariation), while CNVs from patients affected with a neurodevelopmental disorder are being collated in databases such as DECIPHER

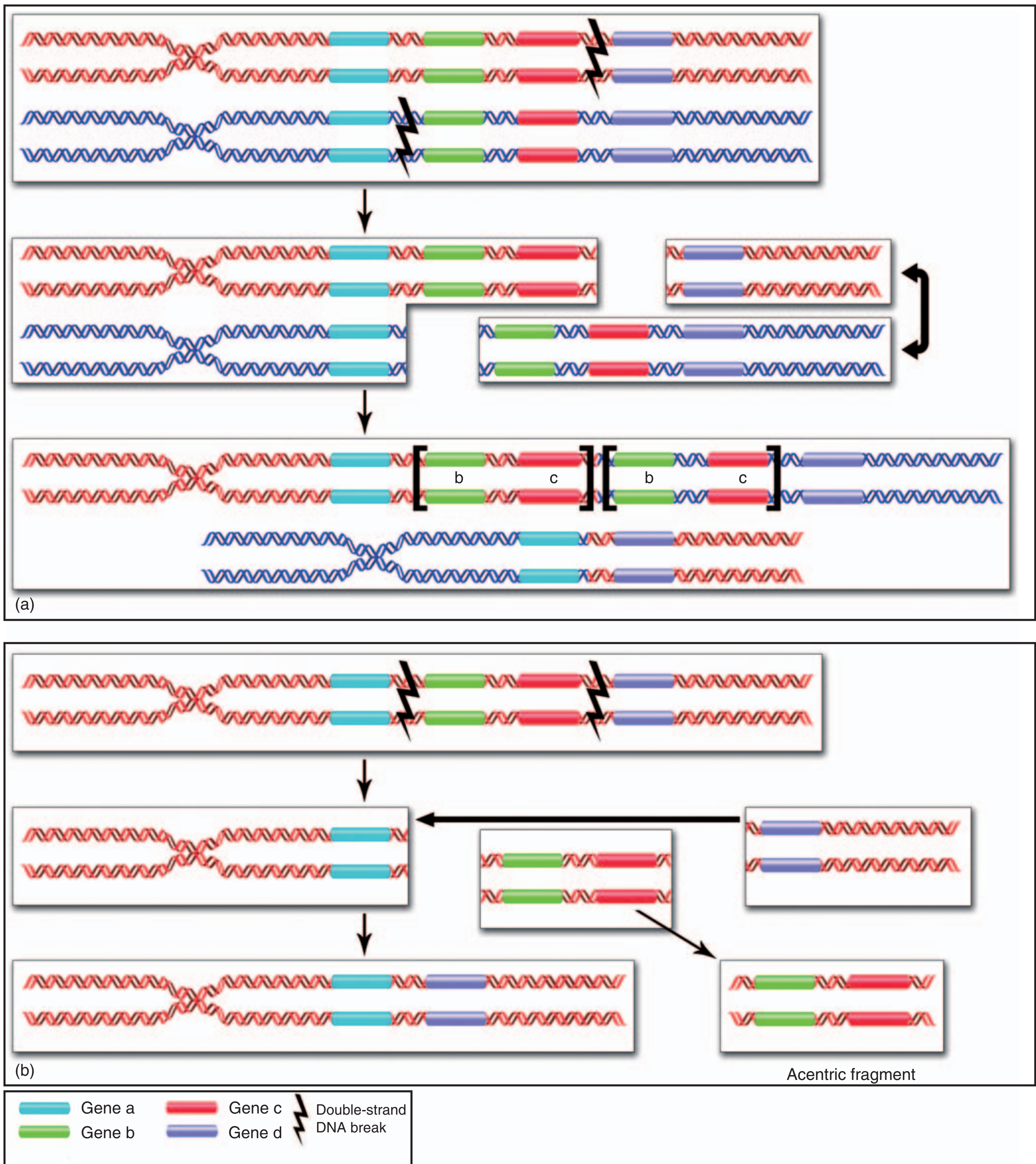


Figure 5.4 Non-homologous end joining (NHEJ) can result in a genomic rearrangement that immediately or eventually results in the gain or loss of genetic material after multiple double-strand DNA breaks occur and are repaired. (a) A balanced rearrangement has occurred and there is no immediate net gain or loss of genomic material within the cell, but it will result in gains or losses of DNA in subsequent generations depending on which chromosome is inherited. (b) In an unbalanced rearrangement, genomic material can be lost immediately (e.g., in this case, the acentric fragment containing one copy of genes b and c will be lost in subsequent cell divisions).

(http://www.sanger.ac.uk/PostGenomic/decipher/), the Chromosome Abnormality Database (http://www.ukcad.org.uk/cocoon/ukcad/) and the European Cytogenetics Association Register of Unbalanced Chromosome Aberrations (http://www.ecaruca.net).

Most CNVs found in healthy individuals are biased away from genes and reside within intergenic regions (i.e., DNA sequences between genes) (Conrad et al., 2006; Nguyen et al., 2006). Nevertheless, the remaining CNVs have been shown to overlap some 3,000 RefSeq genes and 300 genes implicated in genetic disease (as listed in the *Online Mendelian Inheritance in Man* database, http://www.ncbi.nlm.nih.gov/sites/entrez?db=OMIM). Ontology analyses of the genes thought to be copy number variable demonstrate a substantial number that can be classified as "environmental sensor/interaction" genes (Tuzun et al., 2005). These are genes that are involved in sensory perception, neurophysiological processes, drug detoxification, immunity and inflammation, as well as cell surface integrity and cell surface antigens. Clearly, such genes are not critical for early development but rather are involved in our perception and interaction with external stimuli, helping us to adapt to our ever-changing environment.

The functional impact of certain CNVs can be relatively straightforward. For example, reduced copy number of a given gene can often be correlated with reduced expression levels, while additional copies of a gene could lead to increased expression levels of the CNV gene (McCarroll et al., 2006). CNVs that involve parts of a gene could result in fusion gene products or aberrant proteins with addition or loss of specific protein domains. Some CNVs in intergenic regions could overlap regulatory elements that affect the expression of genes as far away as 4 Mb (Stranger et al., 2007), and correlations of CNVs with transcriptional levels do not necessarily have to be positive. For example, deletion of a repressor element may cause upregulation of an associated gene.

DETECTING CNVs IN A GENOME-WIDE MANNER

Array-Based Comparative Genomic Hybridization

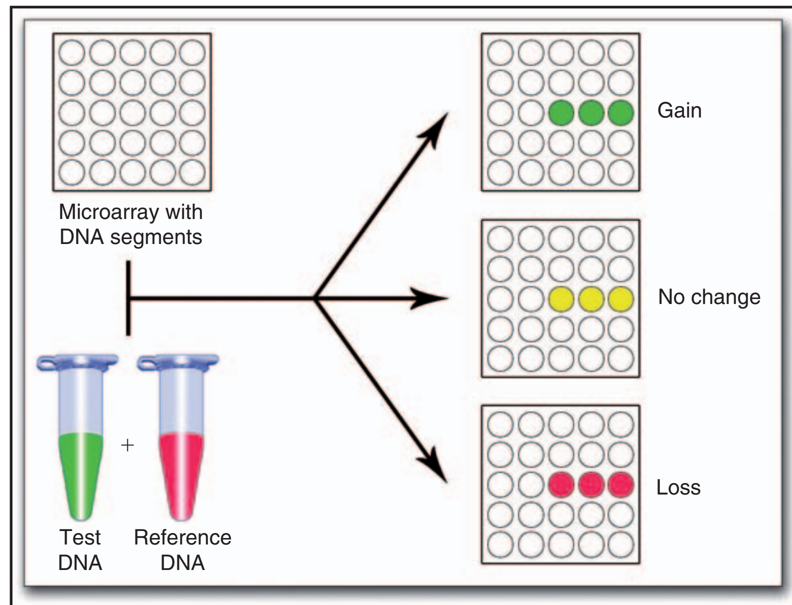
There are different genome-wide methods for detecting CNVs. By far, the most widely used method has been array-based comparative genomic hybridization (aCGH). This technology was first introduced as "matrix-CGH" (Solinas-Toldo et al., 1997) and later referred to as "array CGH" (Pinkel et al., 1998). In aCGH, the "test" genome being interrogated is labeled with one type of fluorescent molecule (e.g., Cy5), and a "normal" or "reference" control genome is labeled with another type of fluorescent molecule (e.g., Cy3) (Figure 5.5). The labeled DNAs are combined, denatured and hybridized to an array of DNA fragments or oligonucleotides on a microscope slide, with each DNA fragment or oligonucleotide representing a unique part of the human genome. The labeled DNAs are then allowed to hybridize to their complementary DNA sequences on the array, in a stoichiometric fashion, such that – by measuring the fluorescence ratio of the two fluorescent dyes at each spot on the array – one can infer the relative copy number of that particular DNA sequence in the genome being tested, with respect to the reference genome (Figure 5.5).

There have been many advances in aCGH technology over the past 5 years. Originally, aCGH platforms contained several hundred or even up to a thousand, large-insert DNA clones (e.g., BAC clones that have an average insert size of about 120–150 kb), which recapitulated the human genome with a clone per ~3 Mb. More recently, the trend has been to manufacture arrays that use smaller DNA sequences as hybridization targets but with increasing number of targets on an array. On such arrays, targets can be oligonucleotides of 45–75 bases in length that have been designed to have similar annealing temperatures (isothermic), based primarily on the length of the oligonucleotide and its GC base pair content. Two companies that produce such arrays are NimbleGen Systems, Inc. (www.nimblegen.com) and Agilent, Inc. (www.agilent.com/chem/goCGH). NimbleGen uses a programmable mirror array to synthesize 385,000 oligonucleotides (more recently 2.1 million targets) directly on a glass surface using photolithography. Agilent, on the other hand, uses ink-jet technology to synthesize 244,000 oligonucleotides (more recently 1.1 million targets), on a spot-by-spot basis.

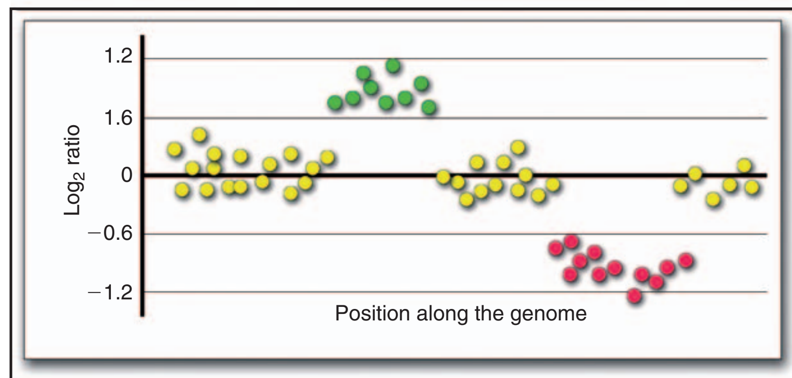
When using oligonucleotide-based arrays, it is important to note that their primary disadvantage is that each oligonucleotide probe tends to have lower signal-to-noise ratios than a large-insert genomic clone target. This results in more experimental "noise" for oligonucleotide-based aCGH assays. Indeed, the typical standard deviation of $\log_2$ ratios for an oligonucleotide-based array is approximately 0.25–0.3, five times the standard deviation of $\log_2$ ratios obtained for BAC-based arrays. However, the inclusion of hundreds of thousands to millions of targets on a given oligonucleotide-based platform provides an assay with an increased resolution. The "effective resolution" of one of these platforms depends largely on the minimum number of consecutive probes needed to confidently call a CNV, which in turn is a function of how well the target sequences were chosen to accurately and consistently respond to copy number changes. Hence, a particular platform that has 500,000 targets and requires only three consecutive probes to make a confident CNV call actually has a higher effective resolution than an array platform with one million targets, but requiring 10 consecutive probes to make a CNV call, assuming that both platforms distribute targets evenly throughout the genome.

Genotyping Arrays

High-throughput array technologies for identifying SNPs can also be used to identify CNVs (see Chapter 4). In general, these arrays contain short targets (20–30 base oligonucleotides) that make them ideal for detecting single base alterations, but less ideal for identifying CNVs (especially when compared to long oligonucleotide-based arrays). For genotyping arrays, only a single labeled DNA source ("test sample") is hybridized, and the signal intensities obtained at each hybridized target appear to have a linear relationship with respect to copy numbers of that particular



(a)



(b)

Figure 5.5 (a) A schematic of an aCGH assay where a test genome (labeled with Cy5, denoted in green) is cohybridized with a reference genome (labeled with Cy3, denoted in red). The DNA probes are mixed and allowed to hybridize to its complementary sequences on the array in a stoichiometric fashion. Fluorescence intensities of the spots on the microarray (each containing a specific DNA sequence) are measured and DNA sequences occurring in greater copy number in the test than in the reference will result in more green fluorescence for those spots on the microarray. A lower copy number of the same DNA sequences will result in more red fluorescence. (b) Typically the $\log_2$ of the fluorescence ratios for each DNA segment on the array is then plotted from one end of a chromosome to the other. A gain is indicated in green and a loss in red. DNA segments having no significant change in DNA copy number in the test sample (with respect to the reference DNA) is indicated in yellow.

DNA sequence in the test genome. For example, if a given DNA sequence has four copies in test genome "X" and only two copies in test genome "Y," the signal intensity obtained for that DNA sequence when test genome "X" is hybridized would be twice that of when test genome "Y" is hybridized. In general, genotyping platforms can detect larger CNVs (especially when there is a higher level of copy number change) or smaller CNVs (that are detectable by numerous targets on the array platform). Affymetrix (www.affymetrix.com) SNP arrays contain targets that are ~25 bases long and the predecessor to the Affymetrix 500K array was recently used to identify CNVs in 270 HapMap individuals (Redon et al., 2006). Illumina Inc. (www.illumina.com) has designed a genotyping platform that uses 50 base oligonucleotides attached to indexed beads on a glass slide. Labeled test DNAs are hybridized to this slide, followed by primer extension and then immunofluorescence detection. Peiffer et al. (2006) recently showed that the ability of Illumina genotyping platforms have also been used to detect CNVs in both constitutional and tumor samples.

Clearly, there would be significant benefits to having platforms that are capable of accurately determining both SNP and CNV genotypes, including reduced reagent costs and expenditure of minimal amounts of DNA. However, the fluorescence intensity

data obtained from these genotyping platforms typically have more noise when trying to obtain copy number information than do long oligonucleotide-based arrays. Hence, both Affymetrix and Illumina are now designing "next generation" arrays that incorporate thousands of non-polymorphic probes (i.e., probes that do not target known SNPs) that fall within known CNV regions. It is thought that a CNV can be confidently detected if enough targets are strategically chosen for a given CNV region and included in the array. In other words, one probe may not be able to reliably detect a single copy loss in a CNV region but cumulative data from one hundred targets in the same CNV region may result in a consistent and confident CNV call.

Whole-Genome Sequence Comparisons

CNVs can also be detected via whole-genome sequence comparison analyses. The main advantage of this method for identifying CNVs is that the acquisition of fine-scale genomic architecture of CNVs (i.e., accurate CNV sizes and breakpoint information). A major disadvantage of this approach is the limited number of individuals for which whole-genome sequence data are currently available. The first human whole-genome sequence was made available from the Human Genome Project. This reference human genome is actually a compilation of DNA sequences from over ten different individuals with ~67% of the DNA sequences originating from the RPCI-11 DNA library, derived from a single individual male. Tuzun et al. (2005) aligned end-sequence data from thousands of fosmid clones from the G248 DNA library (derived from a single North American female) and compared these with the human reference genome sequence. Taking advantage of the tight size restriction of fosmid clones, they were able to identify genomic gains/losses in the reference genome, when pairs of end-sequences aligned with intervening spacing significantly greater or less than the expected 40 kb. In this manner, 241 CNVs were identified in one of these two healthy and presumably normal individuals. More recently, sequencing technologies have advanced to a point where complete human genome sequences can be obtained more efficiently and cost-effectively than previously possible. For example, using Solexa or 454 DNA sequencing approaches, an individual's complete genome sequence can now be obtained within months and at costs of under $250,000. Undoubt-edly, these advances will be used to comprehensively identify and define CNVs at a DNA sequence resolution, provided that a minimal fold-coverage of the person's genome is achieved. Recent studies that have obtained complete genome sequences for different individuals have reported thousands of CNVs in a given individual, encompassing hundreds of millions of bases of DNA (Korbel et al., 2007; Levy et al., 2007).

A major advantage of utilizing DNA sequence comparison strategies is the ability to identify balanced chromosomal rearrangements that cannot be detected by aCGH-based methods. For example, Tuzun et al. (2005) found evidence for 56 inversion breakpoints in their comparative analysis of the two genomes of two individuals, and Korbel et al. (2007) found 132 inversion breakpoints when comparing the genomes of two different individuals.

ASSOCIATION OF CNVs WITH DISEASE AND DISEASE SUSCEPTIBILITY

Genomic imbalances, including CNVs, can contribute to human diseases in at least two ways. First, certain genomic imbalances appear to directly cause neurodevelopmental diseases, that can occur at birth or even later on in life. These genomic imbalances (referred by some as "pathogenic" CNVs) are usually de novo in nature and recent estimates have associated specific genomic imbalances with as many as 50 such genetic syndromes (http://www.sanger.ac.uk/PostGenomics/decipher/). Most of the remaining known genomic imbalances – sometimes referred to as "benign" CNVs because of their identification in healthy individuals – actually have more subtle consequences on human health.

Fcgr3 Variation in Glomerulonephritis

Glomerulonephritis is a major contributor to human kidney failure. *Fcgr3* is a gene that encodes for a receptor found on the surfaces of macrophages and that has low-affinity binding properties to immunoglobulin G. The copy number of *Fcgr3* can vary in humans and among rat strains from 0 to 4 per diploid cell. Individuals with fewer copies of this gene, due to deletions of paralogous *Fcgr3* genes (which appear to have a negative regulatory effect on the "full length" *Fcgr3* gene/gene products), demonstrate increased macrophage activity and an autoimmune response (Aitman et al., 2006; Fanciulli et al., 2007).

DEFB4 CNV in Crohn Disease

Human β-defensins are a family of genes predominantly secreted from leukocytes and epithelial tissues. β-defensins are small proteins (15–20 residues) that function in antimicrobial defense by penetrating a microbe's cell membrane and cause microbial death in a manner similar to that of antibiotics. In the presence of interleukin 1-alpha (IL-1α), which is secreted by macrophages and other immunologically relevant cell types at the site of tissue inflammation, the expression levels of the β-defensin gene, *DEFB4*, increases (O'Neil et al., 1999), to protect the tissue from further microbial invasion. Therefore, individuals with a lower copy number of this β-defensin have decreased immunity against microbes and increased susceptibility to Crohn disease (Inflammatory bowel disease 1) (Fellermann et al., 2006; Naser et al., 2004).

UGT2B17 Variation in Prostate Cancer

Park and colleagues (2006) found that CNV of the *UGT2B17* gene (Murata et al., 2003; Redon et al., 2006; Wilson et al., 2004) confers differential susceptibility to prostate cancer (Figure 5.6). Testosterone is normally processed into dihydrotestosterone

(DHT) and other androgens within human cells. In the presence of functioning UGT2B17 gene product, excess DHT molecules are converted into water-soluble glucuronic acid which can be subsequently eliminated from the cell. When *UGT2B17* is deleted, this process becomes less efficient and results in greater amounts of DHT within the cells. Increased intracellular DHT concentrations lead to increased interactions with androgen receptors that cause elevated cell proliferation. When this occurs in the prostate cells, it can result in prostatic tumorigenesis. Interestingly, while *UGT2B17* null deletions significantly increases prostate cancer susceptibility as a whole, stratified analysis revealed that this correlation to prostate cancer susceptibility is only significant in the Caucasian populations studied, but not among African-Americans.

The CNV of Complement Component C4 in Systemic Lupus Erythematosus (SLE)

Although a link between the complement component C4 (and its isotypes, C4A and C4B) and SLE has been previously reported (Hauptmann et al., 1974a,b), a recent study suggests that the gene's variable copy number actually serves as a significant risk factor for the disease (Yang et al., 2007). The complement system is comprised of over 20 proteins/protein fragments that normally circulate in the blood but, when activated, cause a biochemical cascade that clears pathogens from the human body, often by forming new transmembrane channels in the pathogen and causing osmotic lysis of the target cell. The median copy number of complement component C4 is 4 but can range from 0 to >5 copies among humans (Yang et al., 2007). Low copy numbers of this gene are correlated with increased risk for SLE (Fanciulli et al., 2007; Yang et al., 2007).

CCL3L1 CNVs in HIV/AIDS

A number of studies on HIV/AIDS have focused on the gene copy number and protein levels of CCL3L1 and CCR5 (Gonzalez et al., 2001, 2005) in diseased and healthy individuals. CCR5 is a major co-receptor for the HIV-1 virus as well as the CCL3L1 chemokine. It was recently found that the combination of low *CCL3L1* gene copy number, which is population-specific, and high protein levels of CCR5 showed the greatest increase in susceptibility to AIDS (Gonzalez et al., 2005). It is speculated that increased CCR5 protein levels lead to greater

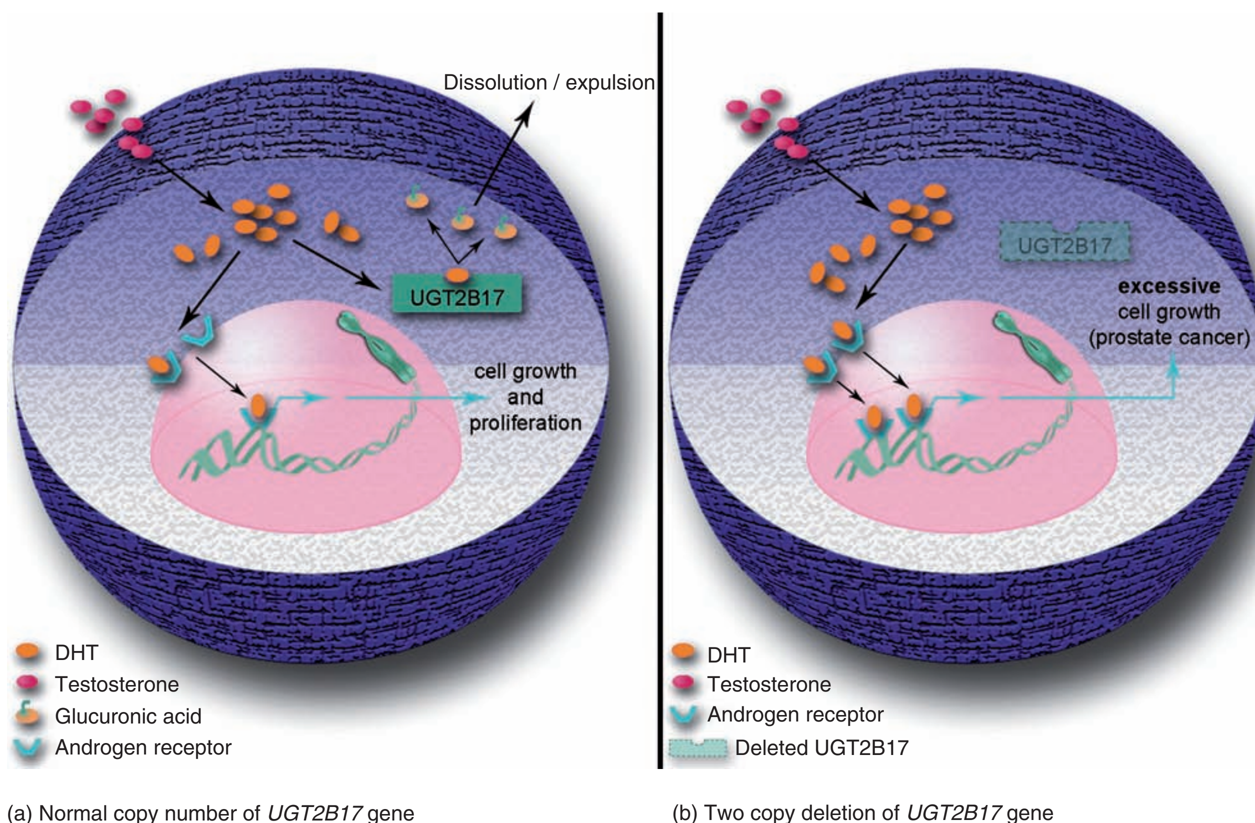


Figure 5.6 Potential mechanism of the *UGT2B17* CNV in prostate cancer pathogenesis. (a) Testosterone is normally processed into DHT and other androgens. When there are two copies of *UGT2B17*, the gene product converts excess molecules of DHT into the water-soluble glucuronic acid and is subsequently eliminated from the cell. (b) When there is a homozygous deletion of *UGT2B17*, the DHT-sequestration pathway no longer exists, increasing endogenous interaction with androgen receptors (AR) and leading to elevated cell proliferation, which in some cases can result in the development of prostate cancer (Park et al., 2006).

number of CD4+CCR5 receptor complexes on the surfaces of T cells and decreased copy number of *CCL3L1* leads to decreased expression (and protein levels) of this chemokine, leaving more CD4+CCR5 receptor complexes available for binding with the HIV-1 virus.

IMPLICATIONS OF CNVs

Disease Association Studies

SNPs have become powerful markers for identifying important disease loci in genetic association studies. However, recent analysis of the complete DNA sequence from a single human individual has shown that CNVs (and other structural genomic variants including small balanced chromosomal rearrangements) account for some 22% of all genetic variation events in the individual and 74% of the total DNA sequence variation, when compared to the human reference genome (Levy et al., 2007). Although it is premature to predict the relative contribution of CNVs to the etiology of common, complex diseases (compared to SNPs), it is clear that CNVs represent a substantial component of human genetic variation that should not be ignored in future disease association studies. Indeed, the presence of CNVs in the human genome has actually reduced the power of certain SNPs. For example, SNPs that lie within CNV regions are difficult to genotype, result in Hardy–Weinberg equilibrium distortions, and are often erroneously calculated as having reduced linkage disequilibrium to nearby causative genomic regions. Moreover, CNVs are far more complex in nature than SNPs and demand appreciation of several factors in order to use this form of genetic variation appropriately in disease association studies. First, the *absolute* copy number (rather than *relative* copy number obtained from aCGH-based experiments) needs to be established for each CNV. Second, the exact genomic location of duplicated CNVs should also be considered. For example, the presence of a third copy of a gene may have dramatically different phenotypic effects depending on where the third copy occurs in the genome. Finally, the precise boundaries (at the DNA sequence level) of each CNV should also be known, as should the specific allelic state of a CNV (e.g., two copies of a gene could be distributed as one copy per chromosome or both copies on a single chromosome). Without this level of information, the power of CNVs in genetic association studies diminishes. Indeed, of the several thousand CNVs that have been identified to date, only a small percent have actually been annotated and genotyped to this precision.

Some efforts have been made to determine if CNVs are in linkage disequilibrium with nearby SNPs. If so, this would allow for specific CNV alleles to be assayed indirectly with a subset of well-characterized SNPs (i.e., "tagging" SNPs) (see Chapter 4). Initial observations suggested that a subset of CNVs did appear to be ancestral in nature and therefore taggable by specific SNPs (McCarroll et al., 2006). Larger CNVs, especially those that are in segmental, duplication-rich areas of the genome, appear to be less taggable by known SNPs (Locke et al., 2006). Part of the reason for this may be that the mutation rate for such CNVs is higher than the mutation rates for nearby SNPs (recall that NAHR mutations rates are estimated to be substantially higher than that of SNPs). Overall, this suggests that, for association studies, a substantial number of CNVs may need to be genotyped directly.

Pharmacogenetics

Since many CNV genes are involved in metabolism and drug detoxification, it has been speculated that CNVs may also make significant contributions to future pharmacogenomic studies (Ouahchi et al., 2006). For example, the *CYP2D6* genes code for enzymes that are involved in the metabolism of more than 30 medications that include antiarrhythmics, antihypertensives, beta-blockers, etc. *CYP2D6* CNVs have been identified and shown to result in gene products with differential metabolism efficiency (Rotger et al., 2007). Similarly, CNVs of genes involved in metabolism may explain some cases of interindividual variation in drug toxicity. McCarroll et al. (2005) and Conrad et al. (2006) found more than 120 CNV genes that were homozygously deleted. Individuals harboring such homozygous deletions presumably have low toxicity tolerance to medications that depend heavily on the homozygously deleted CNV gene product for proper metabolism. Rapid and accurate identification of these individuals should be made a priority in pharmacogenomic research (see also Chapter 15).

Clinical Cytogenetic Diagnostics

Array CGH-based techniques are now being more widely used in the clinical cytogenetic diagnostic arena to identify smaller genomic imbalances that may be associated with neurodevelopmental disorders. Indeed, it has been estimated that aCGH-based assays are now detecting apparently pathogenic genomic imbalances in as much as 20% of cases that have had normal results from chromosome-banded karyotyping tests (Figure 5.7). However, the recent recognition of widespread CNVs among the genomes of healthy individuals has made interpretation of aCGH-based assays more difficult [for a recent review on this, see Lee et al. (2007)]. Specifically, all CNVs detected in an aCGH clinical test need to be assessed for their potential to contribute to the clinical presentation.

One of the primary means for assessing pathogenicity of a CNV is by determining whether or not the CNV is *de novo*. Indeed, 2–3% of children are born with a major birth defect that is sporadic in nature, and the genetic alteration causing these birth defects is similarly assumed to be sporadic in nature. This strategy for determining the pathogenicity of a genomic imbalance in "constitutional" genetic testing is not new and has been applied by cytogeneticists for decades (albeit at the chromosomal banding analysis level). Therefore, for genomic imbalances detected by aCGH assays, it must be determined if the same genomic imbalance is observed in either of the biological

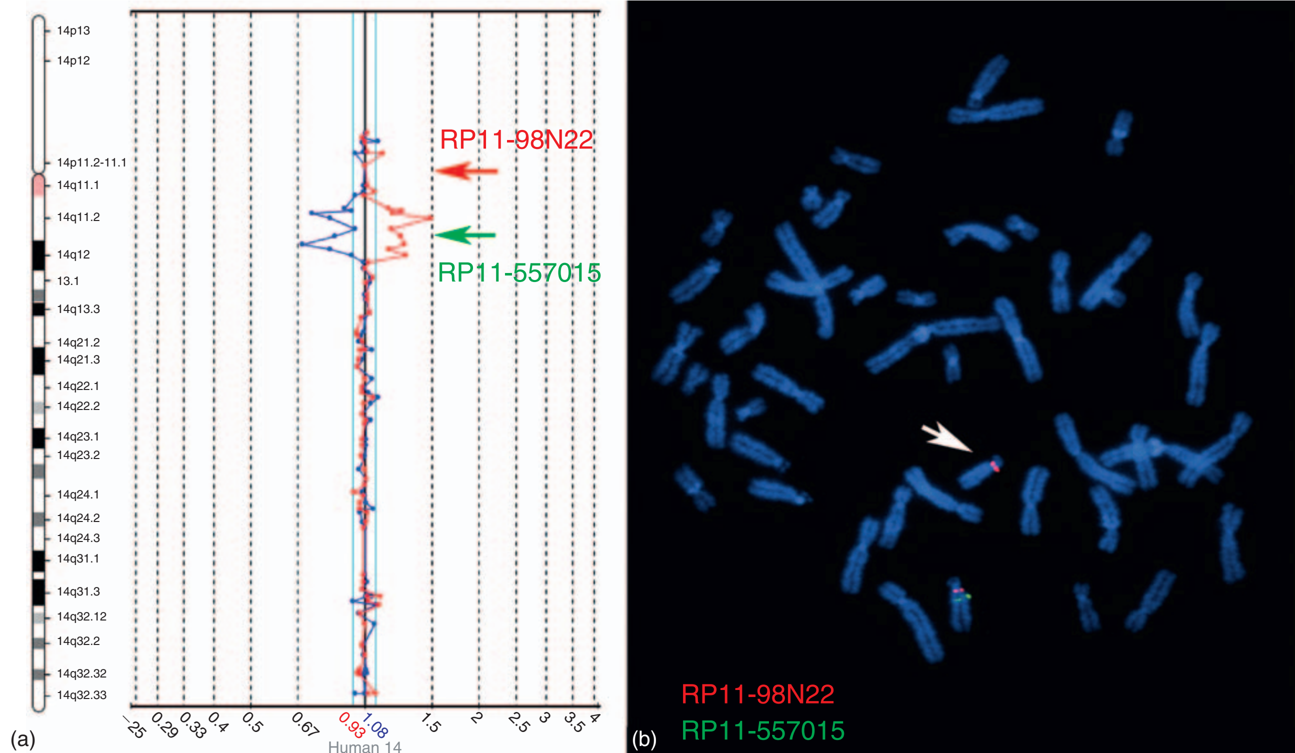


Figure 5.7 Example of a *de novo* genomic imbalance that is likely causative of the clinical phenotype of a patient seen in the clinical genetic clinic. (a) A chromosome 14 profile from a 1 Mb resolution aCGH assay that has been run in dye swap. Deviation of the red line to the right of the expected 1:1 ratio and the blue line to the left of the expected 1:1 ratio is indicative of a genomic loss in the patient DNA. Here, nine BAC clones (each clone represented by a dot on the array profile) are within the deleted region, suggesting a 9–10 Mb sized deletion. The position of BAC clones used in FISH confirmation studies are indicated in red and green. (b) A metaphase spread from the patient after a two-color FISH confirmation assay using a clone within the deletion interval (RP11-557O15) and a control clone outside of the deletion interval (RP11-98N22). The absence of the green signal on one of the chromosome 14s (indicated with a white arrowhead) confirms the aCGH findings. Parental FISH studies were normal (not shown), indicating that this chromosomal aberration was a *de novo* event.

parents, and close attention must be paid to the absolute copy number of the CNV (i.e., not just the relative copy number that is obtained from aCGH experiments), breakpoint information when available and genomic distribution of the CNV DNA segment. Similarly, determining whether a given CNV is present in affected (or unaffected) relatives can also be highly informative.

Another factor to consider is whether the CNV has been found in other unrelated but affected individuals with similar clinical phenotypes (using CNV databases such as DECIPHER) or among healthy individuals (using CNV databases such as the Database of Genomic Variants). This information is only as useful as the accuracy of the CNV entries in these databases (i.e., minimal false positives and precise definition of the architecture of each CNV entry). Moreover, the gene content within a given CNV will also dictate the clinical interpretation of that CNV. Finally, the overall size of the CNV can be a minor indicator of its pathogenicity, with the caveat that "benign" CNVs, as large as 10 Mb in size in gene-poor regions of the genome, have been documented (Hansson et al., 2007).

Accurate clinical interpretation of CNVs will improve as our knowledge of the structure and function of the human genome increases. Clearly, CNVs cannot be considered as isolated genomic aberrations and eventually should be interpreted in the context of all genomic variants, including other CNVs as well as SNPs. For example, the same deletion CNV in a healthy individual may uncover a recessive SNP mutation on the other allele in an unrelated individual – leading to a drastically different phenotype. Since we are at a stage in clinical cytogenetic diagnostics where the technology and the data being obtained are beyond our understanding of human genome function, caution continues to be advised in the application of genome-wide aCGH assays in prenatal diagnostics. Instead, targeted aCGH platforms that primarily utilize probes with well-defined associations with genetic disorders have been used for prenatal diagnoses.

2009 UPDATE

Generating CNVs during DNA replication

Genomic rearrangements that occur in germ cells can potentially lead to *de novo* CNVs in the individual's offspring. The two main mechanisms for generating CNVs have been proposed (see Figures 5.3 and 5.4 in main chapter): (1) NAHR – where homologous recombination occurs between highly identical and multiple sequences in the genome leading to duplication or deletion of DNA sequences that lie between the homologous sequences; and (2) NHEJ – where erroneous repair of double strand breaks lead to gains and losses of DNA segments near the breakpoints. NAHR events are thought to mainly occur during meiosis (although NAHR-like event could also occur mitotically – please see below) and NHEJ events occur in either G1 or G2 of the cell cycle. Most recently, it has been proposed that CNVs could also be generated from errors occurring during DNA replication (i.e. during S-phase). Occasionally, during DNA replication, the DNA replication fork stalls. At these times, the DNA replication machinery may disengage and reattach to template DNA sequences nearby ("template switching"), resulting in duplication or deletion of sequences near the site of the template switching (Lee et al., 2007). If the template switching is facilitated by the presence of cruciform(s) or other non-B DNA structures, one may expect to generate more simple CNV structures at those DNA regions. Otherwise, non-recurrent and complex CNV formations are expected. Evidence to support this mechanism for CNV formation has been provided for the Pelizaeus-Merzbacher disease region (Lee et al., 2007) as well as for MECP2 duplications (Bauters et al., 2008).

CNVs within an individual and between identical twin pairs

Genomic rearrangements that occur somatically (i.e. not in germ cells or germ cell progenitors) can lead to cell-to-cell variability with respect to genomic content per cell. Consistent with this notion, copy number variants have been detected between different tissue types within a given individual (e.g. Sharp et al., 2005; Piotrowski et al., 2008). Furthermore, the fact that CNVs have now also been identified between monozygotic twins (Bruder et al., 2008) suggests that some of these events may be occurring early during embryogenesis. The finding that monozygotic twins may differ with respect to CNVs (and possibly other structural genomic variants) is likely to lead to renewed interest in analyzing the genomic profiles of phenotypically-discordant monozygotic twins to rapidly identify candidate genes involved in the etiology of that specific phenotype.

The role of CNVs in schizophrenia and autism

Evidence is now accumulating that CNVs contribute to the etiology of autism spectrum disorder (ASD) as well as schizophrenia. Sebat et al. (2007) suggested that the burden of *de novo* CNVs was higher in sporadic rates of autism (*de novo* CNVs in 10% of patients from "simplex" families – containing a single child with autism) compared to patients with an affected first-degree relative (*de novo* CNVs in 3% of patients from "multiplex" families – containing multiple affected siblings) or controls (*de novo* CNVs in 1% of the controls). This suggested that *de novo* germline CNVs may be significant risk factors for ASD. One of the most frequently observed *de novo* CNVs identified (aside from the maternal duplication of 15q11-q13) was a 600-kb microdeletion/microduplication at chromosome region 16p11.2 (Sebat et al., 2007, Weiss et al., 2008, Kumar et al., 2008, Marshall et al., 2008). Interestingly, several neuronal synaptic complex genes have been identified within these *de novo* CNV regions including *SHANK3*, *NLGN4*, *NRXN1* (Sebat et al., 2007, Marshall et al., 2008, Weiss et al., 2008, Kumar et al., 2008), as well as genes involved in neuronal activity: *PCDH10*, *DIA1*, and *NHE9* (Morrow et al., 2008).

De novo CNVs also seem to play a role in schizophrenia. One recent study identified three deletions (CNVs ranging from 400 kb to 1.6 Mb size at chromosome regions 1q21.1, 15q11.2, and 15q13.3) that was associated with schizophrenia and psychoses (Stefansson et al., 2008). Rare, *de novo* CNVs, that are larger than 100 kb in length, also appear to contribute to schizophrenia (The International Schizophrenia Consortium, 2008), especially for sporadic (Xu et al., 2008) and early onset cases (Walsh et al., 2008).

New associations with CNVs

Psoriasis. Defensins are small, secreted antimicrobial peptides that are encoded by the *DEFB* genes. Two clusters of *DEFB* genes are found on chromosome 20 and one cluster is found on chromosome 8. On chromosome 8, at 8p23.1, a ~250 kb repeat unit exists containing six *DEFB* genes and among Europeans, the copy number of these defensin genes has been shown to vary from 1 copy to 12 copies per diploid cell (Hollox et al., 2003).

Previously, it was documented that individuals with less than 4 copies of these genes may have a significantly higher risk of developing Crohn's disease, a severe chronic inflammatory bowel disease characterized by intestinal ulceration (Fellerman et al., 2006). It is thought that in these individuals, a reduction in defensin expression leads to the breakdown of the defensin barrier lining the intestinal walls, thereby allowing the adherence and infiltration of the microbial flora into the underlying mucosa. More recently, it has been shown that individuals with increased copy number of these defensin genes, while presumably having less susceptibility to Crohn's disease, have an increased risk for developing psoriasis, a common inflammatory skin disorder. Among the Europeans studied, individuals with > 5 copies of these genes were found to have

a 1.7-fold increased relative risk of developing psoriasis (Hollox et al., 2008). In these individuals, it is thought that increased expression of defensin elicits an elevated inflammatory response upon any minor injury, by activating epidermal growth factor receptor as well as *STAT* (Signal Transducer and Activator of Transcription) signaling pathways.

In addition to the defensin genes, deletion of the late cornified envelope *LCE3B* and *LCE3C* genes also appear to be a susceptibility factor for psoriasis (de Cid et al. 2009).

Crohn's disease. Crohn's disease has now also shown to be associated with a 20 kb common deletion upstream to the autophagy-inducing *IRGM* (Immunity-related GTPase family, M) gene (McCarroll et al., 2008). The 20 kb deletion is in perfect linkage disequilibrium with the strongest Crohn's disease-associated SNP, rs13361189, and leads to different expression patterns in different cell types. Taken together, the data suggest that the deletion haplotype may be a more likely causal variant for the disease, rather than the rs13361189 SNP.

Osteoporosis. Deletion of the *UGT2B17* gene was previously shown to be associated with increased risk of prostate cancer (Park et al., 2006). While deletions of this gene may have a detrimental effect in prostate tissue, they appear to have a protective effect in bones (Yang et al., 2008). *UGT2B17* encodes for a uridine diphospho (UDP)-glucuronosyltransferase enzyme that is involved in the homeostasis and metabolism of steroid hormones. Deletions of this gene are thought to ultimately affect androgen levels such that bone formation is favored over bone resorption. Conversely, increased copy number of *UGT2B17* results in increased expression of the corresponding enzyme, ultimately leading to the reduced bone density that is characteristic of osteoporosis.

Idiopathic generalized epilepsy. A 1.5 Mb microdeletion at chromosome region 15q13.3 was previously identified in individuals with mental retardation and other neuropsychiatric features (Sharp et al. 2008). The critical genomic region of this microdeletion syndrome contains at least seven genes including *CHRNA7*, encoding for the α7 subunit of the nicotinergic acetylcholine receptor. Some individuals diagnosed with this genomic syndrome also exhibit epilepsy. Extending the screen for 15q13.3 microdeletions to 1223 individuals with idiopathic generalized epilepsy, 12 individuals were found to carry this deletion (Helbig et al. 2009). However, out of 3,699 control individuals, none had this same microdeletion. This supports the notion that deletion of *CHRNA7* is a significant risk factor for idiopathic generalized epilepsy.

Starch diet. On the short arm of chromosome 1 lies a cluster of genes that are important for starch digestion. *Amy1* genes are expressed in the saliva and *amy2* genes are expressed in the pancreas. The copy number of *amy1* genes range from 2 – 15 per diploid cell and in individuals from traditionally high starch diets, *amy1* copy number was generally higher, compared to individuals from traditionally low starch diets. Chimpanzees, having low starch in their diets, were found to consistently have 2 copies of *amy1* in each of their cells. Such data support the notion that *amy1* is under positive selection: human populations with high starch diets tend to maintain higher copy numbers of *amy1* because it is advantageous, whereas the copy number of this gene is undergoing neutral evolution or genetic drift in populations having low starch diets (Perry et al. 2007).

Development of a new and improved human reference genome

When the human reference genome was initially described, the existence of widespread structural genomic variation (including CNVs) was not known. We now know that the male individual from whom 66.3% of the reference human genome was derived (the RPCI-11 BAC library), like all individuals, actually had a substantial amount of structural genomic variation in his genome. When comparing this individual's genome with that of another (e.g. Craig Venter), 132 regions of homozygous deletions were detected in one or the other individual (Khaja et al. 2006). Extending such sequence comparisons to eight other individuals revealed another 525 novel DNA segments, that appear to have been homozygously deleted in the individual from which most of the human reference genome was built (Kidd et al. 2008). These "new" human DNA sequences, and other structural genomic variants, now need to be annotated onto a new and improved human reference sequence (http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/index.shtml). Many common structural variants will likely be annotated first, followed by rare structural variants that are discovered via large-scale genomic studies (e.g. such as the 1000 genome project – http://www.1000genomes.org).

2009 UPDATE REFERENCES

Bauters, M., Van Esch, H., Friez, M.J., Boespflug-Tanguy, O. et al. (2008). Nonrecurrent MECP2 duplications medicated by genomic architecture-driven DNA breaks and break-induced replication repair. *Genome Res* 18, 847–858.

Bruder, C.E., Piotrowski, A., Gijsberss, A.A., Anderson, R., Erickson, S. et al. (2008). Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *Am J Hum Genet* 82, 763–771.

De Cis, R., Riveira-Munoz, E., Zeeuwen, P.L., Robarge, J., Liao, W., Dannhauser, E.N. et al. (2009). Deletion of the late cornified envelope *LCE3B and LCE3C* genes as a susceptibility factor for psoriasis. *Nat Genet* 41, 211–215.

Helbig, I., Mefford, H.C., Sharp, A.J., Guipponi, M., Fichera, M., Franke, A. et al. (2009). 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. *Nat Genet* 41, 160–162.

Hollox, E.J., Armour, J.A., Barber, J.C. (2003). Extensive normal copy number variation of a beta-defensin anitmicrobial-gene cluster. *Am J Hum Genet* 73, 591–600.

Holloz, E.J., Huffmeier, U., Zeeuwen, P.L., Palla, R., Lascorz, J., Rodijk-Olthuis, D., et al. (2008). Psoriasis is associated with increased β-defensin genomic copy number. *Nat Genet* 40, 23–25.

Kidd, J.M., Cooper, G.M., Donahue, W.F., Hayden, H.S., Sampas, N., Graves, T., et al. (2008). Mapping and sequencing of structural variation from eight human genomes. *Nature* 453, 56–64.

Khaja, R., Zhang, J., MacDonald, J.R., He, Y., Joseph-George, A.M., Wei, J. et al. (2006). Genome assembly comparison identifies structural variants in the human genome. *Nat Genet* 38, 1413–1418.

Kumar, R.A., KaraMohamed, S., Sudi, J., Conrad, D.F., Brune, C., BadnerJ.A. et al. (2008). Recurrent 16p11.2 microdeletions in autism. *Hum Mol Genet* 17, 628–638.

Lee, J.A., Carvalho, C.M. and Lupski, J.R. (2007b). A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* 131, 1235–1247.

Marshall, C.R. et al. (2008). Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* 82, 477–488.

McCarroll, S.A., Huett, A., Kuballa, P., Chilewski, S.D., Landry, A., Goyette, P. et al. (2008). Deletion polymorphism upstream of *IRGM* associated with altered *IRGM* expression and Crohn's disease. *Nat Genet* 40, 1107–1112.

Morrow, E.M., Yoo, S-Y., Flavell, S.W., Kim, T-K., Lin Y, Hill, R.S. et al. (2008). Identifying autism loci and genes by tracing recent shared ancestry. *Science* 321, 218–223.

Park, J., Chen, L., Ratnashinge, L., Sellers, T.A., Tanner, J.P., Lee, J.H. et al. (2006). Deletion polymorphism of UDP-glucuronosyltransferase 2B17 and risk of prostate cancer in African American and Caucasian men. *Cancer Epidemiol Biomarkers Prev* 15, 1473–1478.

Perry, G.H., Dominy, N.J., Claw, K.G., Lee, A.S., Giegler, H., Redon, R. et al. (2007). Diet and evolution of human amylase gene copy number variation. *Nat Genet* 39, 1256–1260.

Piotrowski, A., Bruder, C.E., Andersson, R., de Stahl, T.D., Menzel, U. et al. (2008). Somatic mosaicism for copy number variation in differentiated human tissues. *Hum Mutat* 29, 1118–1124.

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T. et al. (2007). Strong association of de novo copy number mutations with autism. *Science* 316, 445–449.

Sharp, A.J., Locke, D.P., McGrath, S.D., Cheng, Z., Bailey, J.A. et al. (2005). Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet* 77, 78–88.

Stefansson, H. et al. (2008). Large recurrent microdeletions associated with schizophrenia. *Nature* 455, 232–236.

The International Schizophrenia Consortium. (2008). Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 455, 237–241.

Xu, B., Roos, J.L., Levy, S., van Rensburg E.J., Gogos, J.A., Karayiorgou, M. (2008). Strong association of *de novo* copy number mutations with sporadic schizophrenia. *Nat Genet* 40, 880–885.

Walsh, T., McClellan J.M., McCarthy, S.E., Addington, A.M., Pierce, S.B., Cooper, G.M. et al. (2008). Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 320, 539–543.

Weiss, L.A., Shen, Y., Korn, J.M., Arking, D.E., Miller, D.T., Fossdal, R. et al. (2008). Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* 358, 667–675.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to George Perry for providing the image for Figure 5.2 and Marc Listewnik for his assistance in generating Figure 5.7.

REFERENCES

Aitman, T.J., Dong, R., Vyse, T.J., Norsworthy, P.J., Johnson, M.D., Smith, J. et al. (2006). Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. *Nature* 439, 851–855.

Conrad, D.F., Andrews, T.D., Carter, N.P., Hurles, M.E. and Pritchard, J.K. (2006). A high-resolution survey of deletion polymorphism in the human genome. *Nat Genet* 38, 75–81.

Conrad, D.F. and Hurles, M.E. (2007). The population genetics of structural variation. *Nat Genet* 39, 530–536.

Down, J.L.H. (1866). Observations on an ethnic classification of idiots. *London Hosp Clin Lect Rep* 3, 259.

Fanciulli, M., Norsworthy, P.J., Petretto, E., Dong, R., Harper, L., Kamesh, L., Heward, J.M. et al. (2007). FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. *Nat Genet* 39, 721–723.

Fellermann, K., Stange, D.E., Schaeffeler, E., Schmalzl, H., Wehkamp, J., Bevins, C.L., Reinisch, W. et al. (2006). A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. *Am J Hum Genet* 79, 439–448.

Feuk, L., Marshall, C.R., Wintle, R.F. and Scherer, S.W. (2006). Structural variants: Changing the landscape of chromosomes and design of disease studies. *Hum Mol Genet* 15, R57–R66.

Freeman, J.L., Perry, G.H., Feuk, L., Redon, R., McCarroll, S.A., Altshuler, D.M. et al. (2006). Copy number variation: new insights in genome diversity. *Genome Res* 16, 949–961.

Gonzalez, E., Dhanda, R., Bamshad, M., Mummidi, S., Geevarghese, R., Catano, G., Anderson, S.A. et al. (2001). Global survey of genetic variation in CCR5, RANTES, and MIP-1alpha: Impact on the epidemiology of the HIV-1 pandemic. *Proc Natl Acad Sci USA*, 5199–5204.

Gonzalez, E., Kulkarni, H., Bolivar, H., Mangano, A., Sanchez, R., Catano, G. et al. (2005). The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* 307, 1434–1440.

Hansson, K., Szuhai, K., Knijnenburg, J., van Haeringen, A. and de Pater, J. (2007). Interstitial deletion of 6q without phenotypic effect. *Am J Med Genet A* 143, 1354–1357.

Hauptmann, G., Grosshans, E. and Heid, E. (1974a). Lupus erythematosus syndrome and complete deficiency of the fourth component of complement. *Boll Ist Sieroter Milan* 53(Suppl), 228.

Hauptmann, G., Grosshans, E., Heid, E., Mayer, S. and Basset, A. (1974b). Acute lupus erythematosus with total absence of the C4 fraction of complement. *Nouv Presse Med* 3, 881–882.

Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y. et al. (2004). Detection of large-scale variation in the human genome. *Nat Genet* 36, 949–951.

Korbel, J.O., Urban, A.E., Affourtit, J.P., Godwin, B., Grubert, F., Simons, J.F. et al. (2007). Paired-end mapping reveals extensive structural variation in the human genome. *Science* 318, 420–426.

Lee, C., Iafrate, A.J. and Brothman, A.R. (2007). Copy number variations and clinical cytogenetic diagnosis of constitutional disorders. *Nat Genet* 39, S48–S54.

Lejeune, J., Gautier, M. and Turpin, R. (1959). Etude des chromosomes somatiques de neuf enfants mongoliens. *C R Hebd Seances Acad Sci* 248, 1721–1722.

Levy, S., Sutton, G., Ng, P.C., Feuk, L., Halpern, A.L., Walenz, B.P. et al. (2007). The diploid genome sequence of an individual human. *PLoS Biol* 5, e254.

Locke, D.P., Sharp, A.J., McCarroll, S.A., McGrath, S.D., Newman, T.L., Cheng, Z. et al. (2006). Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. *Am J Hum Genet* 79, 275–290.

McCarroll, S.A., Hadnott, T.N., Perry, G.H., Sabeti, P.C., Zody, M.C., Barrett, J.C. et al. (2006). Common deletion polymorphisms in the human genome. *Nat Genet* 38, 86–92.

Murata, M., Warren, E.H. and Riddell, S.R. (2003). A human minor histocompatibility antigen resulting from differential expression due to a gene deletion. *J Exp Med* 197, 1279–1289.

Naser, S.A., Ghobrial, G., Romero, C. and Valentine, J.F. (2004). Culture of Mycobacterium avium subspecies paratuberculosis from the blood of patients with Crohn's disease. *Lancet* 364, 1039–1044.

Nguyen, D.Q., Webber, C. and Ponting, C. (2006). Bias of selection on human copy-number variants. *PLoS Genet* 2, e20.

O'Neil, D.A., Porter, E.M., Elewaut, D., Anderson, G.M., Eckmann, L., Ganz, T. and Kagnoff, M.F. (1999). Expression and regulation of the human beta-defensins hBD-1 and hBD-2 in intestinal epithelium. *J Immunol* 163, 6718–6724.

Ouahchi, K., Lindeman, N. and Lee, C. (2006). Copy number variants and pharmacogenomics. *Pharmacogenomics* 7, 25–29.

Park, J., Chen, L., Ratnashinge, L., Sellers, T.A., Tanner, J.P., Lee, J.H. et al. (2006). Deletion polymorphism of UDP-glucuronosyltransferase 2B17 and risk of prostate cancer in African American and Caucasian men. *Cancer Epidemiol Biomarkers Prev* 15, 1473–1478.

Perry, G.H., Dominy, N.J., Claw, K.G., Lee, A.S., Fiegler, H., Redon, R., Werner, J., Villanea, F.A., Mountain, J.L., Misra, R., Carter, N.P., Lee, C. and Stone, A.C. (2007). Diet and the evolution of human amylase gene copy number variation. *Nat Genet* 39, 1256–1260.

Peiffer, D.A., Le, J.M., Steemers, F.J., Chang, W., Jenniges, T. and Garcia, F.K. (2006). High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res* 16, 1136–1148.

Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D. et al. (1998). High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* 20, 207–211.

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D. et al. (2006). Global variation in copy number in the human genome. *Nature* 444, 444–454.

Rotger, M., Saumoy, M., Zhang, K., Flepp, M., Sahli, R., Decosterd, L. et al. (2007). The Swiss HIV Cohort Study. Partial deletion of CYP2B6 owing to unequal crossover with CYP2B7. *Pharmacogenet Genomics* 17, 885–890.

Scherer, S.W., Lee, C., Birney, E., Altshuler, D.M., Eichler, E.E., Carter, N.P. et al. (2007). Challenges and standards in integrating surveys of structural variation. *Nat Genet* 39, S7–S15.

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S. et al. (2004). Large-scale copy number polymorphism in the human genome. *Science* 305, 525–528.

Shaffer, L.G. and Lupski, J.R. (2000). Molecular mechanisms for constitutional chromosomal rearrangements in humans. *Annu Rev Genet* 34, 297–329.

Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nickolenko, J., Benner, A., Dohner, H. et al. (1997). Matrix-based comparative genomic hybridization: Biochips to screen for genomic imbalances. *Genes Chromosomes Cancer* 20, 399–407.

Stranger, B.E., Forrest, M.S., Dunning, M., Ingle, C.E., Beazley, C., Thorne, N. et al. (2007). Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 315, 848–853.

Tuzun, E., Sharp, A.J., Bailey, J.A., Kaul, R., Morrison, V.A., Pertz, L.M. et al. (2005). Fine-scale structural variation of the human genome. *Nat Genet* 37, 727–732.

Wilson, W., III, Pardo-Manuel de Villena, F., Lyn-Cook, B.D., Chatterjee, P.K., Bell, T.A., Detwiler, D.A. et al. (2004). Characterization of a common deletion polymorphism of the UGT2B17 gene linked to UGT2B15. *Genomics* 84, 707–714.

Yang, Y., Chung, E.K., Wu, Y.L., Savelli, S.L., Nagaraja, H.N., Zhou, B. et al. (2007). Gene copy-number variation and associated polymorphisms of complement component c4 in human systemic lupus erythematosus (SLE): Low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. *Am J Hum Genet* 80, 1037–1054.

RECOMMENDED RESOURCES

Affymetrix, Inc. (http://www.affymetrix.com)

Agilent, Inc. (www.agilent.com/chem/goCGH)

Chromosome Abnormality Database (http://www.ukcad.org.uk/cocoon/ukcad/)

Database of Genomic Variants (http://projects.tcag.ca/variation)

DECIPHER (http://www.sanger.ac.uk/PostGenomic/decipher/)

Ensembl Genome Browser (http://www.ensembl.org/index.html)

European Cytogenetics Association Register of Unbalanced Chromosome Aberrations (http://www.ecaruca.net)

Human Paralogy Database (http://humanparalogy.gs.washington.edu/structuralvariation)

Human Segmental Duplication Database (http://projects.tcag.ca/humandup/)

Illumina Inc. (www.illumina.com)

NimbleGen Systems, Inc. (www.nimblegen.com)

UCSC Genome Browser (http://www.genome.ucsc.edu/)

CHAPTER

DNA Methylation Analysis: Providing New Insight into Human Disease

Susan Cottrell, Theo deVos, Juergen Distler, Carolina Haefliger, Ralf Lesche, Achim Plum and Matthias Schuster

INTRODUCTION

The human genome contains four bases – guanine, adenine, thymine, and cytosine; but the cytosines can be either methylated or unmethylated at the fifth carbon position in the pyrimidine ring (Figure 6.1a). In general, cytosines can only be methylated when they are in the context of a CpG dinucleotide, or in other words, a cytosine immediately followed by a guanine. These CpG dinucleotides are under-represented in the genome, occurring at only about 20% of the frequency expected assuming random distribution. In some regions, called CpG islands, the frequency of CpG dinucleotides is only 65% of the expected frequency (Takai and Jones, 2002). These stretches of 500–1000 base pairs typically overlap the promoter or exon 1 region of genes, and there are estimated to be 30,000 CpG islands in the human genome (Lander et al., 2001). CpGs in islands are predominantly unmethylated, while CpGs outside of islands are typically methylated.

The methylation status of a CpG island is normally correlated with the chromatin structure and expression levels of nearby genes (Stirzaker et al., 2004). CpG islands associated with actively transcribed genes are typically unmethylated. When a CpG island is methylated, methyl-CpG-binding domain (MBD) proteins recognize the methylated CpG and recruit the necessary factors for chromatin condensation and gene inactivation. This DNA methylation state is maintained during cell division by a family of enzymes called DNA methyltransferases (DNMTs). DNMT1 is thought to be responsible for maintaining methylation patterns through cell divisions, and DNMT3b is the primary enzyme involved in *de novo* methylation (Lopatina et al., 2002). Mutations in one of the DNMT genes, *DNMT3B*, are responsible for the rare genetic disorder ICF syndrome, which is characterized by severe immunodeficiency and DNA hypomethylation (Hansen et al., 1999). Mice lacking any of the DNMTs die either *in utero* or soon after birth (Okano et al., 1999).

Although most CpG islands in the human genome are typically unmethylated in healthy cells, methylation of some islands plays a role in several normal physiological processes. In females, CpG islands on one of the X chromosomes become hypermethylated in the process of chromosomal inactivation (Hellman and Chess, 2007; Mohandas et al., 1981). On autosomes, methylation is also important during development for inactivation of maternally- or paternally-derived alleles of certain genes in a process called genomic imprinting. For instance, a region 2 kb upstream

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

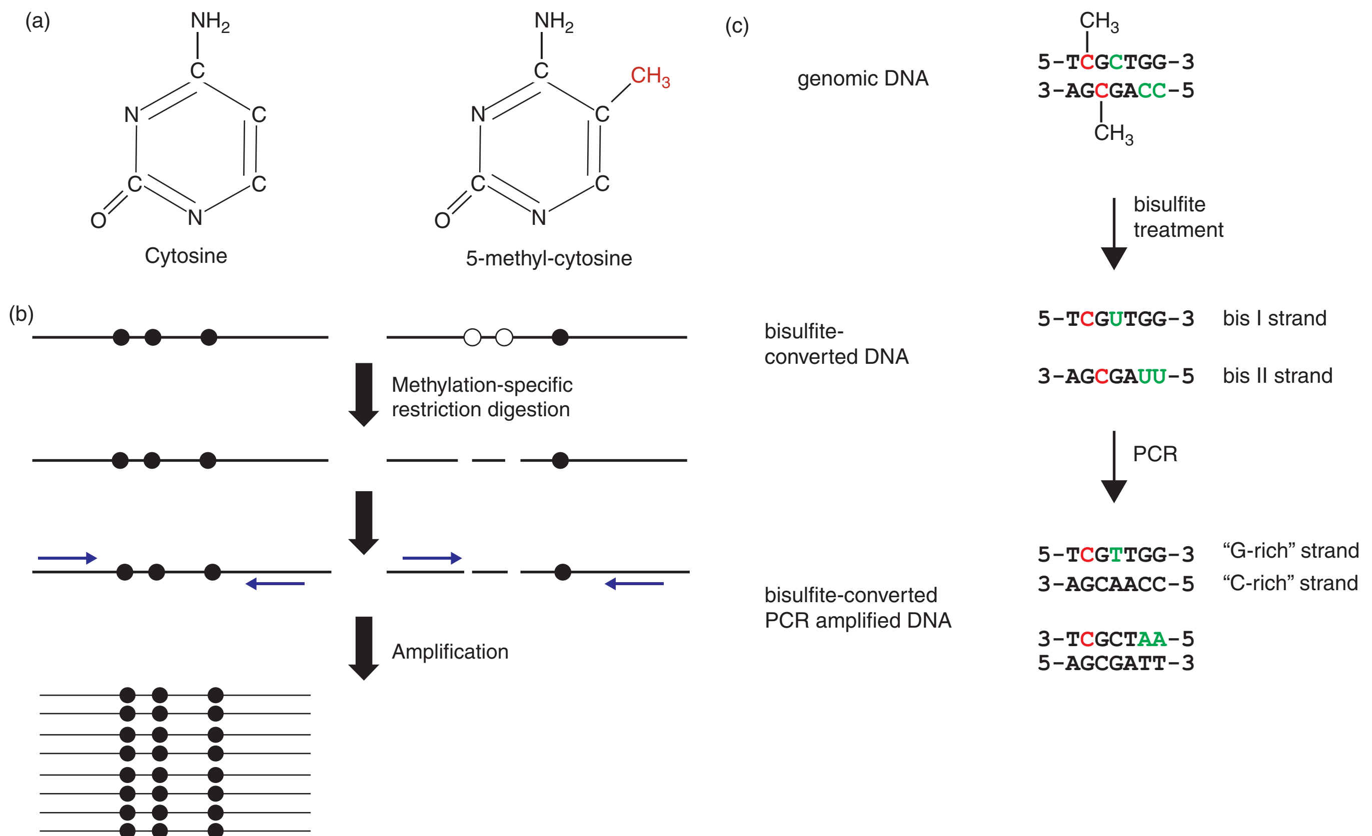


Figure 6.1 Detection of methylated cytosines. (a) Cytosine and 5-methyl-cytosine. (b) Cleavage by some restriction enzymes is blocked by methylation of cytosines in the recognition site. After digestion, various methods can be used to detect the intact methylated DNA, including locus-specific PCR amplification. (c) Genomic DNA is shown at the top with methylated (red) and unmethylated (green) cytosines. The bisulfite treatment converts unmethylated cytosines into uracils. The resulting strands are no longer complementary, and each strand can be amplified by conversion-specific PCR primers.

of the H19 gene for a non-coding RNA is paternally methylated, leading to repressed IGF2 expression (Bartolomei et al., 1993).

During the process of carcinogenesis, two important changes in the methylation pattern of the genome occur. First, sporadic CpGs (outside of CpG islands) become hypomethylated (Ehrlich et al., 1982). This global hypomethylation has been linked to genomic instability, possibly due to mobilization of repetitive elements (Chen et al., 1998; Feinberg, 2004). Meanwhile, some CpG islands, particularly near tumor-suppressor genes, can become hypermethylated during carcinogenesis (Herman and Baylin, 2003; Jones and Baylin, 2002). The hypermethylation is functionally similar to a deletion or mutation and can serve as one method of gene inactivation in Knudson's classic two-hit hypothesis (Jones and Laird, 1999). Different combinations of aberrantly methylated genes are found in nearly all types of cancer (Costello et al., 2000).

TECHNOLOGY TO ASSESS DNA METHYLATION

Cytosine methylation has minimal impact on both thermodynamics and kinetics of common DNA polymerases and can, therefore, not be directly detected by any amplification or hybridization-based methods. Instead, there are three basic principles that can be employed to distinguish differentially methylated DNA sequences: (i) methylation-specific restriction enzymes (Figure 6.1b), (ii) methylation-specific antibodies and proteins, and (iii) methylation-specific bisulfite modification (Figure 6.1c). All existing methods for discovering and evaluating DNA methylation markers are based on one or more of these principles.

Methylation-sensitive restriction enzymes, which typically digest unmethylated but not methylated DNA, have been in use for many years (Bird and Southern, 1978). The recognition sites contain one or more CpG sites. Sometimes cleavage is blocked completely by methylation of the CpG site, but more often the rate of cleavage is affected. Therefore, these methods must be optimized for complete and specific digestion. More recently, restriction enzymes have been introduced that specifically digest methylated DNA (Krueger et al., 1995). Methods based on restriction enzymes are limited to CpGs that are present within a restriction site.

The second major category of DNA methylation methodology relies on naturally occurring proteins whose activity requires methylation detection. Methylation-specific antibodies

have been generated that specifically bind to DNA containing methylated cytosines (Gebhard et al., 2006). In addition, methyl-binding proteins, which play an integral part in bridging DNA methylation information with cellular function, have been successfully used to separate differentially methylated DNA fractions (Cross et al., 1994).

Methylation analysis methods based on bisulfite conversion of sample DNA rely on the unique property of inorganic bisulfite to transform unmethylated cytosine to uracil (Figure 6.1) without affecting methylated cytosines (Frommer et al., 1992). Thus, the bisulfite reaction converts the epigenetic methylation information into sequence information, which can be analyzed by common molecular biology methods, most notably PCR. By converting most of the cytosines to uracils, bisulfite conversion of DNA has the effect of reducing the complexity of the genome, thereby making primer design and multiplexing more challenging. In addition, the chemical reaction can cause DNA damage if sub-optimal conditions are used (Grunau et al., 2001). Despite these caveats, bisulfite-based methods are extremely powerful for sensitive and quantitative analysis of DNA methylation at one or many CpG sites.

Methods for Methylation Marker Discovery

The human genome contains tens of thousands of CpG islands associated with individual genes and their activity. The number of individual methylation positions that could be aberrantly methylated in diseased states is several orders of magnitude higher. Several genome-wide discovery methods to identify differentially methylated regions associated with the specific diagnostic question have been developed. Many of these methods are based on digestion of sample DNA with a methylation-sensitive restriction enzyme. The individual discovery techniques differ in their methods for detection of these restriction cuts, but they include techniques such as linker ligation combined with subtractive hybridization (MCA; Toyota et al., 1999), arbitrarily primed PCR (AP-PCR; Liang et al., 2002), array hybridization (Differential methylation hybridization, DMH; Huang et al., 1999), and two-dimensional gel electrophoresis (RLGS; Costello et al., 2002; Hatada et al., 1991).

These methylation marker discovery methods use methylation-sensitive restriction enzymes to differentiate two classes of samples, such as a DNA from samples of responders to a drug and DNA from non-responders. For a thorough analysis, the entire marker discovery process often involves multiple comparisons. Furthermore, the throughput is low enough that only a few samples can reasonably be processed. Using pooled DNA from multiple patients reduces the likelihood of identifying polymorphic differences associated with individual patients, and it allows for the enrichment of methylation differences associated with disease, which is essential for identification of markers with high prevalence for screening applications.

One approach that is becoming more commonly used is array-based discovery of aberrant methylation (Fukasawa et al., 2006; Hayashi et al., 2007; Taylor et al., 2007). In DMH, originally used for the identification of epigenetic alterations in breast and ovarian cancer (Huang et al., 1999; Yan et al., 2001), genomic DNA is digested by a methylation-insensitive restriction enzyme and then linkers are ligated to the fragments. These products are then digested with a methylation-sensitive restriction enzyme and finally amplified by PCR via the linker sequence. Amplificates are only generated from methylated templates that are not digested by the methylation-sensitive restriction enzymes. The PCR products can be labeled with fluorescent dyes and hybridized to a CpG-island microarray. A recent paper describes the combination of a modified and optimized pre-chip DMH workflow and a custom-made oligonucleotide array representing approximately 50,000 regions of the human genome (Lewin et al., 2007). The advantages of DMH are its reproducibility, speed, and throughput (allowing the option of analyzing individual samples instead of pools). The readout of the microarray not only identifies the differentially methylated sequence, but also provides semi-quantitative methylation values.

Recently, the Human Epigenome Project, a large-scale effort to identify, catalog, and interpret genome-wide DNA methylation patterns of all human genes in all major tissues, has been initiated by The Wellcome Trust Sanger Institute (Cambridge, UK), the Centre National de Génotypage (Paris, France), and Epigenomics AG (Berlin, Germany) (Table 6.1). Based on public versions of the sequence of the human genome established by the Human Genome Project, the Human Epigenome Project will systematically uncover the epigenetic information layer that is still hidden to a large extent. A second Human Epigenome Project was recently proposed by an international group of 40 scientists (Esteller, 2006; Jones and Martienssen, 2005) (Table 6.1).

Methods for Analyzing Markers in Tissue Samples

Methylation analysis in tissues is often not limited by the quantity of sample material available, but the quality of the material, particularly formalin-fixed and paraffin-embedded samples, can be an issue (Schuster, 2004). The isolated DNA is usually damaged to some extent depending on the fixation procedure. In addition, the proportion of the tissue type with aberrant methylation (i.e., the percent tumor) can vary from sample to sample. Laser capture microdissection (LCM), which is used to isolate the tissue type of interest, helps overcome this issue.

There are very different levels at which DNA methylation can be assessed (Figure 6.2) (Laird, 2003), ranging from genome-wide to a particular locus or CpG. The "methylation content" of a cell population corresponds to the overall proportion of methylated cytosines within the entire genome. In contrast, the "methylation level" designates the percentage of methylated DNA strands at one genomic locus, while the "methylation pattern" describes the individual methylation status of a specific set of CpG positions on one DNA strand within a region of interest. Tissue testing is best performed using marker information contained in the "methylation level".

In principle, the methylation level of a DNA sample can be measured in a straightforward manner after bisulfite treatment because the methylation ratio at the CpG site of interest is

Table 6.1 Multicenter efforts for large-scale epigenetic analysis

	Goals	Approach	Institutions	References
Human Epigenome Project	• "To identify, catalogue and interpret genome-wide DNA methylation patterns of all human genes in all major tissues" • Identify "Methylation Variable Positions"	Bisulfite DNA sequencing	Wellcome Trust Sanger Institute; Epigenomics, AG; The Centre National de Genotypage	Eckhardt et al., 2006 http://www.epigenome.org/
International Alliance for the Human Epigenome and Disease (AHEAD)	• A comprehensive catalog of potential chromatin alterations • Describe "reference epigenomes"	To be determined; likely to include a variety of methods to assess DNA methylation and histone modifications	To be determined	Jones and Martienssen 2005
Encyclopedia of DNA Elements (ENCODE)	• Identify all functional elements in the human genome	Chip-on-chip assays for DNA methylation and chromatin modifications; comparative genomics; DNase I hypersensitivity site mapping	Welcome Trust Sanger Institute; Stanford University; University of Virginia; Affymetrix; Municipal Institute of Medical Research; Ludwig Institute for Cancer Research; Yale University; University of Washington; others	ENCODE project consortium, 2007 http://www.genome.gov/10005107

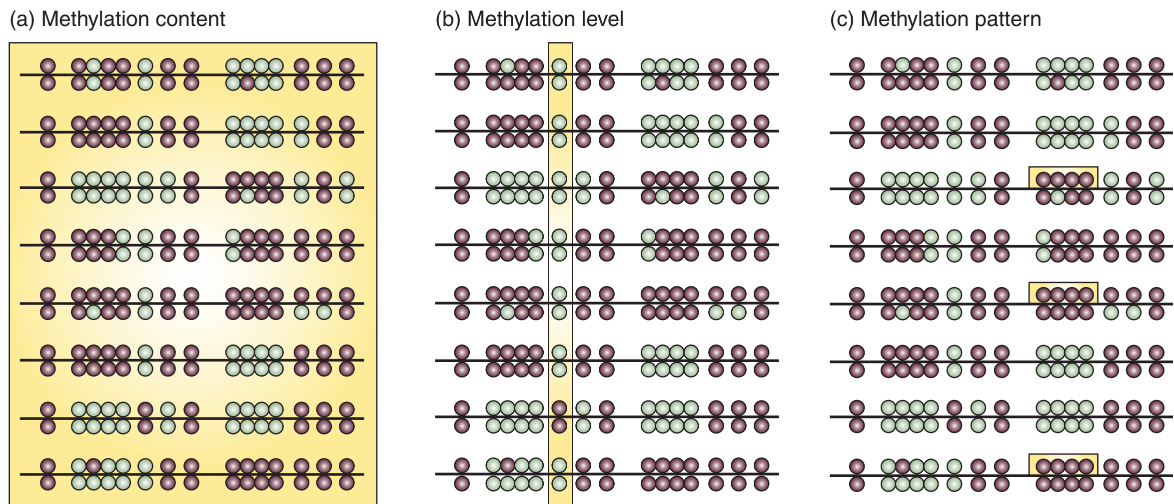


Figure 6.2 Methods for assessment of DNA methylation. The "methylation content" of a cell population corresponds to the overall proportion of methylated cytosines within the entire genome (a), the "methylation level" designates the percentage of methylated DNA strands at a one genomic locus (b), and the "methylation pattern" describes the individual methylation status of a specific set of CpG positions on one DNA strand within a region of interest (c).Reproduced with permission of Macmillan Magazines Ltd from *Nature Reviews Cancer* 3(4), April 2003, 253–266. Copyright (2003) Macmillan Magazines Ltd.

translated into an allele ratio. However, in most cases not enough DNA will be available for direct analysis. The bisulfite-treated DNA can be amplified without disturbing this allele ratio using conventional PCR with the primers designed to avoid CpG sites. After amplification, the product can be analyzed using any of a number of analysis methods typically used for the quantification of allele ratios. In MS-SNuPE (Methylation-Specific Single Nucleotide Primer Extension), extension primers

designed to hybridize just upstream of a CpG site are extended by one nucleotide in a methylation-specific manner. Detection and quantification based on HPLC (El-Maarri et al., 2002) and incorporation of radioactivity have been described (Gonzalgo et al., 1997). Another example is pyrosequencing, in which the sequential incorporation of additional bases into the extension primer is translated into a light flash of the same intensity (Colella et al., 2003). Alternatively, hybridization of methylation-specific probes to amplified bisulfite-treated DNA has been combined with different readout systems, including microarrays (Adorjan et al., 2003) and real-time PCR (Quantitative MethyLight; Eads et al., 2000). All of the methods mentioned measure relative levels of methylation, but can be used to quantify DNA methylation levels when they are appropriately calibrated on mixtures of methylated and unmethylated DNA templates.

The methods described in the previous paragraph use an unbiased amplification of bisulfite-treated DNA for measuring methylation levels. Methylation in tissues can also be estimated using various forms of methylation-specific amplification. The most commonly used method in methylation research is methylation-specific PCR (MSP) or the real-time version, MethyLight (Eads et al., 2000). These methods use primers designed for a certain methylation pattern of bisulfite-converted DNA. MethyLight technology is a combination of methylation-specific primers and a methylation-specific real-time PCR probe (Figure 6.3). Since these methods only amplify and detect the methylated portion of the DNA sample, they can only provide qualitative methylation pattern information. In combination with an estimation of total DNA amount, which can be measured for instance by a control real-time PCR assay, methylation-specific assays can be used to estimate relative methylation pattern levels (Ogino et al., 2006a).

For tissue-based analyses, a quantitative methylation assay is likely to be important, because in many cases the required information will depend not on the presence or absence of methylation but on the level of methylation. Responders to a certain drug, for instance, might have fewer inactivated copies of a gene than non-responders. In some instances, the methylation difference could be quite subtle, and it is not clear which of the methods will provide the level of precision required.

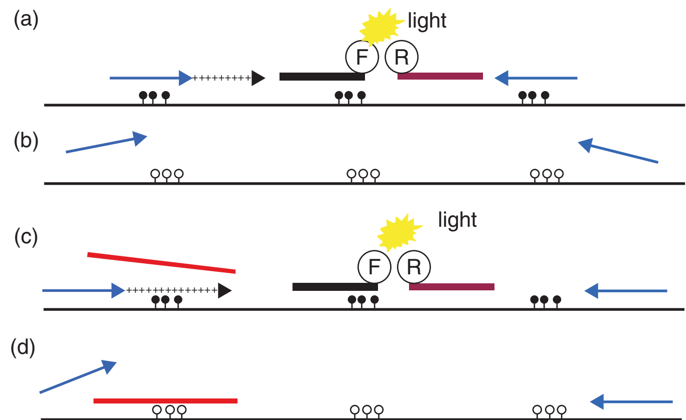


Figure 6.3 Sensitive real-time PCR assays for methylation detection MethyLight assays (a, b) have two methylation-specific primers that bind only to the methylated (a) but not the unmethylated (b) versions of their binding sites. If both primer binding sites are methylated, PCR products are formed and will be detected by methylation-specific real-time PCR probes (a). In contrast, HM assays (c, d) use primers that amplify all methylation states. The methylation specificity of the priming is provided by the blocker oligonucleotide, which prevents the primer from binding to unmethylated DNA (d). Therefore, only methylated DNA is amplified and will be detected by a methylation-specific fluorescent probe (c). Filled circles indicate methylated CpGs, and open circles indicate unmethylated CpGs.

Methods for Analyzing Markers in Disease Screening Samples

The samples of choice for early detection or screening are body fluids that can be obtained by non-invasive procedures. Serum and plasma are optimal targets for methylation marker analyses, but investigators have also analyzed methylation in urine (Hoque et al., 2005), sputum, bronchial lavage, ductal lavage, ejaculate, and other "remote samples" (reviewed in Laird, 2003). In healthy individuals, only small amounts of circulating DNA are observed in plasma and serum, whereas high amounts have been described in patients with various conditions. The main sources of circulating DNA are thought to be apoptotic or necrotic cell death (Holdenrieder et al., 2001). DNA is released from degrading cells after nuclease cleavage of the chromatin between the nucleosomes. Therefore, the tumor- or disease-derived DNA fragments can be as short as the smallest nucleosomal fragment, approximately 180 base pairs (Cottrell and Laird, 2003; Jahr et al., 2001).

In contrast to tissue-based applications, methylation markers for screening must fulfill additional biological requirements. Even if the majority of free DNA in blood is derived from cancer cells, a significant amount of DNA from other cell types, particularly leukocytes, will be present. Therefore, the marker must be methylated in a high percentage of tumor cells and relatively unmethylated in leukocytes and other tissues contributing background DNA. Qualitative markers, which are methylated only in tumor-derived DNA, are well suited for screening applications.

The clinical sensitivity (proportion of diseased people testing positive) and specificity (proportion of healthy individuals testing negative) of the marker is not only dependent on its biological properties, but also on the analytical performance of the assay. The challenge for the final diagnostic assay is a high analytical sensitivity (detecting small amounts of methylated tumor DNA) at a maximum specificity (no detection of unmethylated DNA). In practice, methylation assays suitable for screening must be capable of detecting a few copies of the methylated marker DNA in a background of 1–10 thousand-fold excess unmethylated DNA. In addition, the DNA extraction method must be optimized for high recovery yield, particularly of short DNA fragments found in plasma or serum. Bisulfite treatment is known to cause damage to the DNA, but milder treatment

conditions can be used without risk of incomplete conversion of the DNA (Berlin et al., 2005).

MethyLight may be an assay that is suitable for detection of disease in remote samples. In contrast to MSP, MethyLight minimize false-positive results caused by possible mis-priming of the methylation-specific primer, as the methylation status of the amplificates are analyzed on a second level by the methylation-specific fluorescent probe. MethyLight has been successfully used for various applications, but primers and probes must be carefully designed in order to ensure sensitive and specific detection.

HeavyMethyl (HM) technology is another way to achieve methylation-specific priming, and, like MSP, can be combined with a probe for real-time detection (Cottrell et al., 2004). HM assays combine methylation-non-specific primers and a non-extendable 3′-modified blocker oligonucleotide. The blocker binds to unmethylated DNA, preventing the primers from accessing the priming site (Figure 6.3). The blocker does not bind to the methylated version, allowing the primers to access the priming sites. A methylation-specific fluorescent probe is used to ensure specificity and allow for methylation quantification. HM assays typically have a limit of detection of 25–50 pg of methylated DNA and can detect this methylated DNA in a background of 50 ng of unmethylated DNA. Although HM design requires more effort compared to a MethyLight assay, the HM assay has advantages. Unmethylated template that is amplified unintentionally will be targeted by the blocker in the next cycle. Therefore, the analytical specificity of HM assays is very high.

CLINICAL IMPACT OF DNA METHYLATION ANALYSIS

Cancer has been viewed as an accumulation of chromosomal aberrations and therefore been called a "genetic disease". However, it has become clear over the last two decades that epigenetic changes play a crucial role in carcinogenesis. Research on DNA methylation in cancer has been expanding rapidly, and for this reason, in the following sections, we will rely on examples from oncology to demonstrate how methylation profiles can be applied for basic disease research, early detection, disease stratification, treatment response prediction, and drug target identification.

While attention has focused on methylation in carcinogenesis, a similar groundswell of research is emerging on methylation in other diseases, especially autoimmune and cardiovascular conditions (see Table 6.2). For example, aberrantly hypomethylated DNA in circulation in systemic lupus erythematosus might induce an immune response due to its similarity to unmethylated microbial DNAs (Januchowski et al., 2004). Further contributions of defective methylation to lupus pathogenesis could be due to transcription of endogenous retroviruses (Okada et al., 2002) or the increase in the expression of certain genes related to autoreactivity (Richardson, 2002). In cardiovascular disease, DNA methylation is associated with atherosclerosis pathogenesis in two ways: First, the decrease in essential factors (e.g., folate and vitamin D) for the synthesis of *S*-adenosyl methionine (the main methyl group donor in the methylation reactions) leads to global DNA hypomethylation and coexisting hyperhomocystinemia (Zaina et al., 2005). Second, Lund et al. (2004) provided evidence that methylation profiles are early markers of the disease in a mouse model. Taken together, these data reinforce the notion that methylation profiles can be highly valuable to study the pathogenesis of a variety of conditions (Table 6.2).

Impact of Methylation Research on Basic Disease Understanding

Researchers have spent many years studying the role of permanent alterations in DNA sequence, such as mutations, deletions, and insertions, in carcinogenesis. Cancer has long been thought of as a disease that arises after accumulation of mutational events in growth control genes. In recent years, RNA expression profiling has been used to examine the functional consequences of these sequence alterations, as outlined in other chapters in this book. There is now overwhelming evidence that epigenetic modification can also control RNA and protein levels. As the methylation analysis toolbox expands, we are gaining more insight into how alterations in these patterns of methylation influence disease.

As an example, methylation analysis has provided a new molecular understanding of colon cancer. Colon cancer develops in a well-described pathway from normal epithelium, to aberrant crypts, to adenomas, then to adenocarcinoma, each with specific histological features and molecular alterations. The first alterations to be studied were sequences changes: for instance, *APC* mutations have been implicated in the formation of adenomas (Powell et al., 1992) and *TP53* mutations have been implicated in the transition from adenoma to adenocarcinoma (Baker et al., 1990).

More recently, the role of aberrant methylation events of genes in these same pathways has been elucidated. For example, a hereditary form of colon cancer is commonly caused by mutations in the *MLH1* gene, but this same gene can induce sporadic forms of the disease when it is aberrantly methylated (Cunningham et al., 1998). Some colon cancers are characterized by microsatellite instability (MSI), and the majority of these tumors have methylated *MLH1* alleles. Furthermore, research on methylation in colon cancer has established that tumors can not only be divided based on microsatellite stability status, but also on their CpG island methylator phenotype (CIMP) status (Toyota et al., 1999). Colon tumors, particularly those with MSI, have a bimodal distribution based on the methylation status of multiple genes. Those tumors with multiple methylated genes (CIMP+) are more likely to have *BRAF* mutations and wild-type *KRAS* (Ogino et al., 2006a, b). Clearly, a complete understanding of the pathogenesis of colon cancer requires knowledge of both genetic and epigenetic events.

Methylation Markers for Early Detection and Diagnosis

Early detection is an effective success factor in the fight against any type of cancer. Aberrant DNA methylation patterns show

TABLE 6.2 Examples of DNA methylation in human disease

Disease/Disorder	Phenotype	Mechanism	Genes involved	Reference
Cancer	Tumorigenesis; Increased cell division; metastasis	Hypermethylation	Tumor-suppressor genes	Jones and Baylin, 2002; Herman and Baylin, 2003
		Hypomethylation	Global	Ehrlich et al., 1982
		Loss of imprinting	Insulin-like grow factor 2 (IGF2), GTP-binding RAS-like 3 (DIRAS3), mesoderm-specific transcript homolog (MEST)	Feinberg, 2007
		Heritable germline hypermethylation	Mut L Homolog 1 (MLH1)	Hitchins et al., 2007
Beckwith–Wiedemann syndrome	Embryonal tumors; omphalocele; macroglossia (large tongue); gigantism	Disruption of an imprinting locus	11p15 (IGF2, H19 ncRNA, cyclin-dependent kinase inhibitor 1C (CDKN1C), pleckstrin homology-like domain, family A, member 2 (PHLDA2), solute carrier family 22, member 1 (SLC22A1), potassium voltage-gated channel (KCNQ1))	Weksberg et al., 2003
Prader–Willi syndrome	Developmental delay; obesity; genital hypoplasia; distinct facial features	Disruption of an imprinting locus	15q11-13 (small nuclear ribonucleoprotein polypeptide N (SNRPN), ncRNAs)	Nicholls and Knepper, 2001
Angelman syndrome	Developmental delay; ataxia; microcephaly; distinct facial features	Disruption of an imprinting locus	15q11-13 (Ubiquitin E3 ligase (UBE3A))	Nicholls and Knepper, 2001
Pseudohypoparathyroidism	Hypoparathyroidism without sensitivity to parathyroid hormone	Disruption of an imprinting locus	20q13 (GNAS complex locus (GNAS1))	Bastepe et al., 2001
Immunodeficiency/centromeric instability/facial anomalies (ICF syndrome)	Immunodeficiency; chromosomal breakage; facial defects	Mutations	DNA Methyltransferase 3b (DNMT3b)	Hansen et al., 1999
Rett syndrome	Motor abnormalities; ataxia; hand-wringing; poor verbal skills	Mutations	Methyl-CpG-binding protein 2 (MeCP2)	Amir et al., 1999
Systemic Lupus Erythematosus	Autoimmune disorder; inflammation; tissue damage	Hypomethylation	global, perforin (PRF1), CD70, integrin, alpha L (ITGAL)	Januchowski et al., 2004
Atherosclerosis	Thickening and hardening of arterial walls	Hypermethylation	Estrogen receptor, Monocarboxylate Transporter 3 (SLC16A3)	Kim et al., 2007; Zhu et al., 2005
Schizophrenia	Hallucinations; delusions	Hypomethylation	Global	Shimabukuro et al., 2006
		Hypermethylation	Reelin (RELN)	Grayson et al., 2005
Endometriosis	Growth of endometrial cells outside of the uterus	Hypermethylation	Homeobox A10 (HOXA10)	Wu et al., 2005
Fragile X syndrome	Mental retardation; autism; large ears; macroorchidism	Hypermethylation of an expanded trinucleotide repeat	Fragile X mental retardation 1 (FMR1)	Oberle et al., 1991

promise as biomarkers for early detection of cancer for several reasons: Tumor-specific methylation changes occur early in tumorigenesis, appear to be stable, provide an amplifiable signal, and can be assayed with high analytical sensitivity. DNA methylation patterns characteristic of tumor cells have been found for many types of cancer in several types of body fluids (Laird, 2003).

DNA methylation changes occur frequently in colon cancer and can be present early in malignant transformation (Grady, 2005). Feasibility studies have shown that tumor-derived methylated DNA markers can be detected in serum, plasma, and stool (reviewed in Laird, 2003). These studies analyzed a limited number of patient samples, and larger clinical studies will be necessary to assess the quality of these markers for a screening approach.

Prostate-specific antigen (PSA) based screening has resulted in detection of many cases of prostate cancer, but, due to the low specificity of this assay, its usefulness for prostate cancer screening is no longer certain. DNA methylation markers may improve screening and diagnosis of prostate cancer considerably in the future. *GSTP1*, encoding glutathione *S*-transferase P1, is the best-characterized methylation marker for prostate cancer. Methylation of *GSTP1* in prostate cancer has been consistently reported with a frequency of about 70–90% (reviewed in Henrique and Jeronimo, 2004). Interestingly, *GSTP1* methylation is also reported in some high-grade prostatic intraepithelial neoplasia (PIN) lesions, a precursor to prostate cancer, indicating that *GSTP1* hypermethylation is an early event in prostate cancer development (Henrique et al., 2006). Several groups have demonstrated the feasibility of using *GSTP1* methylation and other methylation markers to detect prostate cancer in bodily fluids. For example, Hoque et al. (2005) analyzed a panel of nine genes in urine sediments from 52 cancer patients and 91 controls using MethyLight assays. A panel containing *GSTP1* and three other genes (*CDKN2A*, *ARF*, and *MGMT*) detected the prostate cancer cases with 87% sensitivity and 100% specificity.

Recently, Lofton-Day et al. (2006) conducted a study to identify and validate DNA methylation-based markers for colorectal cancer using plasma samples. At a set specificity of 95%, the lead methylation marker (Septin 9) showed in several studies sensitivity values of 51, 65, and 50%, respectively. In total, approximately 2000 samples were tested, including over 600 plasma samples from colorectal cancer patients, 600 with related non-malignant diseases, and over 600 normal controls from an age-matched, colonoscopy-verified healthy population. Importantly, early-stage cancers were identified with the same sensitivity as later stage cancers (Lofton-Day et al., 2006). The test was able to detect colorectal cancers regardless of their location, addressing a critical medical need and shortfall of the existing fecal occult blood tests. Once a methylation-based screening test is developed, its use can potentially be extended to monitoring for disease recurrence after treatment, similar to the role of PSA for prostate cancer.

Methylation Markers for Disease Prognosis

Molecular classification holds great promise for assessment of optimal treatment options. DNA methylation markers have been reported as both prognostic and predictive markers, indicating the aggressiveness of tumors and the responsiveness to certain drug regimens, respectively. For example, prognostic markers for breast cancer are required to determine which patients can be spared from aggressive therapy due to very good prognosis. Recently, several reports have been published exploring the prognostic relevance of DNA methylation at the promoter of the *PLAU* gene, which codes for the matrix remodeling enzyme urokinase-type plasminogen activator (uPA) (Guo et al., 2002; Pakneshan et al., 2004; Xing and Rabbani, 1999). Interestingly, *PLAU*-promoter DNA methylation was negatively associated with tumor grade; grade 1 tumors were frequently methylated and grade 3 tumors were unmethylated. These findings are consistent with results from large studies that have demonstrated that protein expression in serum of uPA and its cell surface receptor uPAR are important determinants of distant spread of breast cancer (Look et al., 2002).

Shinozaki et al. (2005) found that in a panel of 151 primary breast tumors, hypermethylation of the *CDH1* gene was significantly associated with lymphovascular invasion, infiltrating ductal histology, and lack of estrogen receptor expression. On the other hand, *RASSF1A* and *RARB* hypermethylation were significantly more common in estrogen receptor-positive and human epidermal growth factor receptor 2-positive tumors, respectively. Other genes that have been associated with prognosis in breast cancer and which appear to be regulated via DNA methylation are *CDH3* (Paredes et al., 2005), the tuberous sclerosis (TSC) genes *TSC1* and *TSC2* (Jiang et al., 2005), and *LATS1*/*LATS2*, tumor-suppressor genes that have been implicated in the regulation of the cell cycle (Takahashi et al., 2005).

We have recently identified a methylation marker, *PITX2*, which predicts outcome in hormone receptor-positive, node-negative breast cancer patients treated with tamoxifen monotherapy (Maier et al., 2004, 2007). The marker was originally identified and validated in two independent cohorts of patients treated with adjuvant tamoxifen monotherapy. The findings were then validated in an independent study analyzing paraffin-embedded tumors of 422 node-negative patients treated with tamoxifen only after surgery. In the group with low *PITX2* methylation, 94% of the patients were metastasis-free after 10 years, compared to only 84% in the group with high *PITX2* methylation. In a multivariate model, *PITX2* methylation added significant information to conventional factors such as tumor size, grade, and age (Harbeck et al., 2005). Analysis of *PITX2* methylation in a large cohort of patients not receiving adjuvant therapy confirmed that *PITX2* methylation is a prognostic marker associated with tumor aggressiveness (Martens et al., 2005). These results provide substantial evidence that *PITX2* DNA methylation is suitable for routine clinical use in order to predict outcome in node-negative, tamoxifen-treated patients, and to identify low-risk patients who can be spared the burden of additional cytotoxic therapy.

Methylation Markers for Treatment Response Prediction

A number of studies have provided evidence that specific methylation changes are associated with responses to a variety of cancer

therapeutic agents currently used in the clinics (reviewed in Maier et al., 2005). These alterations in methylation patterns could serve as predictive markers of drug response and be used by clinicians and patients to support treatment selection. Furthermore, methylation markers could be developed together with a novel drug during clinical trials in order to better select the patients who are likely to respond. The inclusion of predictive markers in clinical trials is becoming increasingly important as more targeted therapies are being tested.

Many anti-cancer therapies disturb DNA integrity and thus interfere with DNA synthesis and successful cellular replication, ultimately inducing cell death. Tumor response to DNA damaging agents is closely linked to expression of several DNA-repair enzymes, many of which are regulated by methylation. Resistance can result from either a gain of methylation event or loss of methylation event. Methylation of *MGMT*, which codes for a DNA-repair enzyme, is associated with response to alkylating agents. In this case, tumor cells rely on the excess *MGMT* product to specifically repair the damage that occurred during therapy with the alkylating agent (Ludlum, 1990). Esteller et al. (2000) investigated the methylation status of the *MGMT* gene in 47 glioma samples from patients treated with the chloroethylating agent carmustine. The *MGMT* promoter region was methylated in 19 (40%) of the samples, 12 (63%) of which were from patients who had a response to carmustine. Of the 28 tumors with unmethylated *MGMT*, only 1 patient (4%) responded to the drug ($p < 0.001$, univariate analysis). Paz and coworkers (2004) found similar results in glioma patients. Balana and colleagues (2003) demonstrated that detection of these treatment response markers might be possible in serum for disease where affected tissue is unavailable.

MLH1 is a mismatch repair enzyme that is activated in response to DNA damage. However, MLH1 expression is not only required to repair the damage, but also seems to be linked to apoptotic signaling: Its activation induces processes leading to programmed cell death. Therefore, in contrast to *MGMT*, methylation of the *MLH1* promoter is associated with resistance to DNA damaging agents, such as temozolomide, dacarbazine, and cisplatin. Expression of *MLH1* increases the effect of the drugs, presumably by acting as a sensor detecting DNA damage caused by the drug, and by activating processes that eventually lead to apoptosis of the cell (Agarwal and Kaye, 2003; Brown et al., 1997; Karran and Bignami, 1994). Other drug classes for which methylation markers were shown to be associated with tumor response in patients or sensitivity of preclinical models are taxanes (*CHFR*, *TRAG3*, *RASSF1*, *BRCA1*), platinum compounds (*ABCB1*, *FANCF*, *RASSF1*), retinoids (*RARB*, *RBP1*), and anti-hormonal therapies (*ESR1*, *ESR2*, *AR*) (Maier et al., 2005).

2009 UPDATE

The field of epigenetics continues to expand rapidly. In early 2008, the National Institutes of Health in the U.S. introduced its Roadmap Epigenomics Initiative program (http://nihroadmap.nih.gov/epigenomics/). The program will create epigenome mapping and data coordination centers, bolster technology development and marker discovery, and drive research on the role of epigenetic regulation in health and disease. Meanwhile, support for a human epigenome project continues to grow internationally (The American Association for Cancer Research Human Epigenome Task Force and European Union, Network of Excellence, Scientific Advisory Board, 2008).

Knowledge of the strengths and limitations of various DNA methylation analysis technologies has become more sophisticated in the past few years, and researchers are beginning to experiment with "next generation" bisulfite sequencing (Taylor et al., 2007; Meissner et al., 2008). This high throughput method is likely to be more quantitative and sensitive than previous methods, and will also give researchers more qualitative information on patterns of CpG methylation across a region of the genome. As genomic methodology expands in many areas, researchers are developing computational approaches to combine information from multiple layers of genomic, epigenomic, and proteomic analyses.

Recent research continues to explore the relationship between DNA methylation and various chromatin alterations. The multiple levels of epigenetic control seem to offer an intricate system for controlling expression in cells such as embryonic stem cells and for priming the cells for response to future events (Ku et al., 2008). As we gain insight into the functioning of this system in normal cells, we are also beginning to understand how aberrant methylation events arise in neoplastic cells (Schlesinger et al., 2007). In fact, some researchers are experimenting with predicting sites of aberrant methylation based on DNA sequence motifs and polycomb binding sites (McCabe et al., 2009).

While our knowledge of DNA methylation events in oncology continues to grow, recent research has begun to elucidate the role of DNA methylation outside of cancer, including mental illness, infertility, and obesity (Houshdaran et al., 2007). For example, Mill et al. (2008) used a combined genome-wide screen and candidate gene approach to find evidence of altered DNA methylation near genes in the glutamatergic and GABAergic neurotransmission pathways in post-mortem brains from individuals with schizophrenia and bipolar disorder.

Recent research has also shown that DNA methylation marks can be long-standing, inherited through several generations, or can be quite dynamic, responding to precisely timed signals. The evidence of epigenetic inheritance comes mostly from studies of the effect of exposure of model organisms to environmental toxins (Nilsson et al., 2008), but epidemiological studies suggest a possible role in humans as well (Kaati et al.,

2007). In contrast, dynamic alterations in methylation of certain genes is involved in memory formation in experiments in which fear conditioning in rats resulted in rapid methylation changes that are reversed within 24 hours (Miller and Sweatt 2007).

Advances in our understanding of the role of DNA methylation in disease are also starting to pay off in clinical practice. The Presept trial, which began enrolling patients in 2008 (http://www.presept.net), is designed to test the promising Septin 9 methylation assay for early detection of colorectal cancer in plasma (Grützmann et al., 2008). Clinical testing for MGMT methylation, a valuable biomarker that predicts response of glioblastomas to radiation therapy and alkylating agents, became available in 2008. In coming years, we expect to see numerous other examples of the clinical benefits of the extensive work being done in the field of epigenetics.

2009 UPDATE REFERENCES

American Association for Cancer Research Human Epigenome Task Force and European Union, Network of Excellence, Scientific Advisory Board (2008). Moving AHEAD with an international human epigenome project. *Nature* 454, 711–715.

Grützmann, R., Molnar, B., Pilarsky, C., Habermann, J., Schlag, P., Saeger, H., Miehlke, S., Stolz, T., Model, F., Roblick, U. et al. (2008). Sensitive detection of colorectal cancer in peripheral blood by septin 9 DNA methylation assay. *PLoS ONE* 3, e3759.

Houshdaran, S., Cortessis, V., Siegmund, K., Yang, A., Laird, P. and Sokol, R. (2007). Widespread epigenetic abnormalities suggest a broad DNA methylation erasure defect in abnormal human sperm. *PLoS ONE* 2, e1289.

Kaati, G., Bygren, L., Pembrey, M. and Sjöström, M. (2007). Transgenerational response to nutrition, early life circumstances and longevity. *Eur J Hum Genet* 15, 784–790.

Ku, M., Koche, R., Rheinbay, E., Mendenhall, E., Endoh, M., Mikkelsen, T., Presser, A., Nusbaum, C., Xie, X., Chi, A. et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* 4, e1000242.

McCabe, M., Lee, E. and Vertino, P. (2009). A multifactorial signature of DNA sequence and polycomb binding predicts aberrant CpG island methylation. *Cancer Res* 69, 282–291.

Meissner, A., Mikkelsen, T., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B., Nusbaum, C., Jaffe, D. et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454, 766–770.

Mill, J., Tang, T., Kaminsky, Z., Khare, T., Yazdanpanah, S., Bouchard, L., Jia, P., Assadzadeh, A., Flanagan, J., Schumacher, A. et al. (2008). Epigenomic profiling reveals DNA-methylation changes associated with major psychosis. *Am J Hum Genet* 82, 696–711.

Miller, C. and Sweatt, J. (2007). Covalent modification of DNA regulates memory formation. *Neuron* 53, 857–869.

Nilsson, E., Anway, M., Stanfield, J. and Skinner, M. (2008). Transgenerational epigenetic effects of the endocrine disruptor vinclozolin on pregnancies and female adult onset disease. *Reproduction* 135, 713–721.

Schlesinger, Y., Straussman, R., Keshet, I., Farkash, S., Hecht, M., Zimmerman, J., Eden, E., Yakhini, Z., Ben-Shushan, E., Reubinoff, B.E. et al. (2007). Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for *de novo* methylation in cancer. *Nat Genet* 39, 232–236.

Taylor, K., Kramer, R., Davis, J., Guo, J., Duff, D., Xu, D., Caldwell, C. and Shi, H. (2007). Ultradeep bisulfite sequencing analysis of DNA methylation patterns in multiple gene promoters by 454 sequencing. *Cancer Res* 67, 8511–8518.

REFERENCES

Adorjan, P., Distler, J., Lipscher, E., Model, F., Muller, J., Pelet, C., Braun, A., Florl, A., Gutig, D., Grabs, G. et al. (2003). Tumor class prediction and discovery by microarray-based DNA methylation analysis. *Nucleic Acids Res* 30, e21.

Agarwal, R. and Kaye, S.B. (2003). Ovarian cancer: Strategies for overcoming resistance to chemotherapy. *Nat Rev Cancer* 7, 502–516.

Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U. and Zoqhbi, H.Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet* 23, 185–188.

Baker, S.J., Preisinger, A.C., Jessup, J.M., Paraskeva, C., Markowitz, S., Willson, J.K., Hamilton, S. and Vogelstein, B. (1990). p53 gene mutations occur in combination with 17p allelic deletions as late events in colorectal tumorigenesis. *Cancer Res* 50, 7717–7722.

Balana, C., Ramirez, J., Taron, M., Roussos, Y., Ariza, A., Ballester, R., Sarries, C., Mendez, P., Sanchez, J. and Rosell, R. (2003). O6-methyl-guanine-DNA methyltransferase methylation in serum and tumor DNA predicts response to 1,3-bis(2-chloroethyl)-1-nitrosourea but not to temozolamide plus cisplatin in glioblastoma multiforme. *Clin Cancer Res* 9, 1461–1468.

Bartolomei, M.S., Webber, A.L., Brunkow, M.E. and Tilghman, S.M. (1993). Epigenetic mechanisms underlying the imprinting of the mouse H19 gene. *Genes Dev* 7, 1663–1673.

Bastepe, M., Lane, A.H. and Jueppner, H. (2001). Paternal uniparental isodisomy of chromosome 20q – and the resulting changes in GNAS1 methylation – as a plausible cause of pseudohypoparathyroidism. *Am J Hum Genet* 68, 1283–1289.

Berlin, K., Cardon, K. and Ballhause, M. (2005). Improved bisulfite conversion of DNA. Patent WO 2005/038051.

Bird, A.P. and Southern, E.M. (1978). Use of restriction enzymes to study eukaryotic DNA methlation: I, The methylation pattern in ribosomal DNA from Xenopus laevis. *J Mol Biol* 118, 27–47.

Brown, R., Hirst, G., Gallagher, W., McIlwrath, A., Margison, G., van der Zee, A. and Anthoney, D. (1997). hMLH1 expression and cellular

responses of ovarian tumour cells to treatment with cytotoxic anticancer agents. *Oncogene* 15, 45–52.

Chen, R., Pettersson, U., Beard, C., Jackson-Grusby, L. and Jaenisch, R. (1998). DNA hypomethylation leads to elevated mutation rates. *Nature* 395, 89–93.

Colella, S., Shen, L., Baggerly, K.A., Issa, J.-P.J. and Krahe, R. (2003). Sensitive and quantitative universal Pyrosequencing methylation analysis of CpG sites. *BioTechniques* 35, 146–150.

Costello, J.F., Fruhwald, M.C., Smiraglia, D.J., Rush, L.J., Robertson, G.P., Gao, X., Wright, F.A., Feramisco, J.D., Peltomaki, P., Lang, J.C. et al. (2000). Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat Genet* 24, 132–138.

Costello, J.F., Smiraglia, D.J. and Plass, C. (2002). Restriction landmark genome scanning. *Methods* 27, 144–149.

Cottrell, S. and Laird, P. (2003). Sensitive detection of DNA methylation. *Ann NY Acad Sci* 983, 120–130.

Cottrell, S., Distler, J., Goodman, N., Mooney, S., Kluth, A., Olek, A., Schwope, I., Tetzner, R., Ziebarth, H. and Berlin, K. (2004). A real-time PCR assay for DNA-methylation using methylation-specific blockers. *Nucleic Acids Res* 32, e10.

Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P. (1994). Purification of CpG islands using a methylated DNA binding column. *Nat Genet* 6, 236–244.

Cunningham, J.M., Christensen, E.R., Tester, D.J., Kim, C.Y., Roche, P.C., Burgart, L.J. and Thibodeau, S.N. (1998). Hypermethylation of the hMLH1 promoter in colon cancer with microsatellite instability. *Cancer Res* 58, 3455–3460.

Eads, C., Danenberg, K., Kawakami, K., Saltz, L., Blake, C., Shibata, D., Danenberg, P. and Laird, P. (2000). MethyLight: A high-throughput assay to measure DNA methylation. *Nucleic Acids Res* 28, E32.

Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V.K., Attwood, J., Burger, M., Burton, J., Cox, T.V., Davies, R., Down, T.A. et al. (2006). DNA methylation profiling of human chromosomes 6, 20, and 22. *Nat Genet* 38, 1378–1385.

Ehrlich, M., Gama-Sosa, M.A., Huang, L.H., Midgett, R.M., Kuo, K.C., McCune, R.A. and Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Res* 10, 2709–2721.

El-Maarri, O., Herbiniaux, U., Walter, J. and Oldenburg, J. (2002). A rapid, quantitative, non-radioactive bisulfite-SNuPE- IP RP HPLC assay for methylation analysis at specific CpG sites. *Nucleic Acids Res* 30, e25.

ENCODE Project Consortium (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799–816.

Esteller, M. (2006). The necessity of a human epigenome project. *Carcinogenesis* 27, 1121–1125.

Esteller, M., Garcia-Foncillas, J., Andion, E., Goodman, S., Hidalgo, O., Vanaclocha, V., Baylin, S. and Herman, J. (2000). Inactivation of the DNA-repair gene MGMT and the clinical response of gliomas to alkylating agents. *N Engl J Med* 343, 1350–1354.

Feinberg, A.P. (2004). The epigenetics of cancer etiology. *Semin Cancer Biol* 14, 427–432.

Feinberg, A.P. (2007). Phenotypic plasticity and the epigenetics of human disease. *Nature* 447, 433–440.

Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Griff, G.W., Molloy, P.L. and Paul, C.L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci USA* 89, 1827–1831.

Fukasawa, M., Kimura, M., Morita, S., Matsubara, K., Yamanaka, S., Endo, C., Sakurada, A., Sato, M., Kondo, T., Horii, A. et al. (2006). Microarray analysis of promoter methylation in lung cancers. *J Hum Genet* 51, 368–374.

Gebhard, C., Schwarzfischer, L., Pham, T.H., Schilling, E., Klug, M., Andreesen, R. and Rehli, M. (2006). Genome-wide profiling of CpG methylation identifies novel targets of aberrant hypermethylation in myeloid leukaemia. *Cancer Res* 66, 6118–6128.

Gonzalgo, M.L. and Jones, P.A. (1997). Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE). *Nucleic Acids Res* 25, 2529–2531.

Grady, W.M. (2005). Epigenetic events in the colorectum and in colon cancer. *Biochem Soc Trans* 33, 684–688.

Grayson, D.R., Jia, X., Chen, Y., Sharma, R.P., Mitchell, C.P., Guidotti, A. and Costa, E. (2005). Reelin promoter hypermethylation in schizophrenia. *Proc Natl Acad Sci USA* 102, 9341–9346.

Grunau, C., Clark, S.J. and Rosenthal, A. (2001). Bisulfite genomic sequencing: Systematic investigation of critical experimental parameters. *Nucleic Acids Res* 29, e65–e65.

Hansen, R.S., Wijmenga, C., Luo, P., Stanek, A.M., Canfield, T.K., Weemaes, C.M. and Gartler, S.M. (1999). The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proc Natl Acad Sci USA* 96, 14412–14417.

Harbeck, N., Bohlmann, I., Ross, J.S., Gruetzmann, R., Kristiansen, G., Margossian, A., Hartmann, A., Cufer, T., Paradiso, A. and Maier, S. EpiBreast Group (2005). Multicenter study validates PITX2 DNA methylation for risk prediction in tamoxifen-treated, node-negative breast cancer using paraffin-embedded tumor tissue. *J Clin Oncol (ASCO Annual Meeting Proceedings)* 23, 505.

Hatada, I., Hayashizaki, Y., Hirotsune, S., Komatsubara, H. and Mukai, T. (1991). A genomic scanning method for higher organisms using restriction sites as landmarks. *Proc Natl Acad Sci USA* 88, 9523–9527.

Hayashi, H., Nagae, G., Tsutsumi, S., Kaneshiro, K., Kozaki, T., Kaneda, A., Sugisaki, H. and Aburatani, H. (2007). High-resolution mapping of DNA methylation in human genome using oligonucleotide tiling array. *Hum Genet* 120, 701–711.

Hellman, A. and Chess, A. (2007). Gene body-specific methylation on the active X chromosome. *Science* 315, 1141–1143.

Henrique, R. and Jeronimo, C. (2004). Molecular detection of prostate cancer: A role for GSTP1 hypermethylation. *Eur Urol* 46, 660–669.

Henrique, R., Jeronimo, C., Teixeira, M.R., Hoque, M.O., Carvalho, A.L., Pais, I., Ribeiro, F.R., Oliveira, J., Lopes, C. and Sidransky, D. (2006). Epigenetic heterogeneity of high-grade prostatic intraepithelial neoplasia: Clues for clonal progression in prostate carcinogenesis. *Mol Cancer Res* 4, 1–8.

Herman, J.G. and Baylin, S.B. (2003). Gene silencing in cancer in association with promoter hypermethylation. *N Engl J Med* 349, 2042–2054.

Herman, J.G., Graff, J.R., Myohanen, S., Nelkin, B.D. and Baylin, S.B. (1996). Methylation-specific PCR: A novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci USA* 93, 9821–9826.

Hitchins, M.P., Wong, J.J., Suthers, G., Suter, C.M., Martin, D.I., Hawkins, N.J. and Ward, R.L. (2007). Inheritance of a cancer-associated MLH1 germ-line epimutation. *N Engl J Med* 356, 731–733.

Holdenrieder, S., Stieber, P., Bodenmuller, H., Busch, M., Fertig, G., Furst, H., Schalhorn, A., Schmeller, N., Untch, M. and Seidel, D.

(2001). Nucleosomes in serum of patients with benign and malignant diseases. *Int J Cancer* 95, 114–120.

Hoque, M.O., Topaloglu, O., Begum, S., Henrique, R., Rosenbaum, E., Van Criekinge, W., Westra, W.H. and Sidransky, D. (2005). Quantitative methylation-specific polymerase chain reaction gene patterns in urine sediment distinguish prostate cancer patients from control subjects. *J Clin Oncol* 23, 6569–6575.

Huang, T.H., Perry, M.R. and Laux, D.E. (1999). Methylation profiling of CpG islands in human breast cancer cells. *Hum Mol Genet* 8, 459–470.

Jahr, S., Hentze, H., Englisch, S., Hardt, D., Fackelmayer, F.O., Hesch, R.D. and Knippers, R. (2001). DNA fragments in the blood plasma of cancer patients: Quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res* 61, 1659–1665.

Januchowski, R., Prokop, J. and Jagodzinski, P.P. (2004). Role of epigenetic DNA alterations in the pathogenesis of systemic lupus erythematosus. *J Appl Genet* 45, 237–248.

Jiang, W.G., Sampson, J., Martin, T.A., Lee-Jones, L., Watkins, G., Douglas-Jones, A., Mokbel, K. and Mansel, R.E. (2005). Tuberin and hamartin are aberrantly expressed and linked to clinical outcome in human breast cancer: The role of promoter methylation of TSC genes. *Eur J Cancer* 41, 1628–1636.

Jones, P.A. and Laird, P.W. (1999). Cancer epigenetics comes of age. *Nat Genet* 21, 163–167.

Jones, P.A. and Baylin, S.B. (2002). The fundamental role of epigenetic events in cancer. *Nat Rev Genet* 3, 415–428.

Jones, P.A. and Martienssen, R. (2005). A blueprint for a Human Epigenome Project: The AACR Human Epigenome Workshop. *Cancer Res* 65, 11241–11246.

Karran, P. and Bignami, M. (1994). DNA damage tolerance, mismatch repair and genome instability. *Bioessays* 16, 833–839.

Kim, J., Kim, J.Y., Song, K.S., Lee, Y.H., Seo, J.S., Jelinek, J., Goldschmidt-Clermont, P.J. and Issa, J.P. (2007). Epigenetic changes in estrogen receptor beta gene in atherosclerotic cardiovascular tissues and in-vitro vascular senescence. *Biochim Biophys Acta* 1772, 72–80.

Krueger, T., Wild, C. and Noyer-Weidner, M. (1995). McrB: A prokaryotic protein specifically recognizing DNA containing modified cytosine residues. *EMBO J* 14, 2661–2669.

Laird, P.W. (2003). The power and promise of DNA methylation markers. *Nat Rev Cancer* 3, 253–266.

Lander, E., Linton, L., Birren, B., Nusbaum, C., Zody, M., Baldwin, J., Devon, K., Dewar, K., Dewar, I., Doyle, M. et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.

Lewin, J., Plum, A., Hildmann, T., Rujan, T., Eckhardt, F., Liebenberg, V., Lofton-Day, C. and Wasserkort, R. (2007). Comparative DNA methylation analysis in normal and tumour tissues and in cancer cell lines using differential methylation hybridisation. *Int J Biochem Cell Biol* 39, 1539–1550.

Liang, G., Gonzalgo, M.L., Salem, C. and Jones, P.A. (2002). Identification of DNA methylation differences during tumorigenesis by methylation-sensitive arbitrarily primed polymerase chain reaction. *Methods* 27, 150–155.

Lofton-Day, C., Model, F., Tetzner, R., deVos, T., Schuster, M., Lesche, R., Sledziewski, A. and Day, R. (2006). A real-time PCR test for Septin 9 gene methylation identifies early stage colorectal cancer in plasma. 97th Annual Meeting of the AACR. Abstract LB-224.

Look, M.P., van Putten, W.L., Duffy, M.J., Harbeck, N., Christensen, I.J., Thomssen, C., Kates, R., Spyratos, F., Ferno, M. and Eppenberger-Castori, S. (2002). Pooled analysis of prognostic impact of urokinase-type plasminogen activator and its inhibitor PAI-1 in 8377 breast cancer patients. *J Natl Cancer Inst* 94, 116–128.

Lopatina, N., Haskell, J.F., Andrews, L.G., Poole, J.C., Saldanha, S. and Tollefsbol, T. (2002). Differential maintenance and de novo methylating activity by three DNA methyltransferases in aging and immortalized fibroblasts. *J Cell Biochem* 84, 324–334.

Ludlum, D.B. (1990). DNA alkylation by the haloethylnitrosoureas: Nature of modifications produced and their enzymatic repair or removal. *Mutat Res* 233, 117–126.

Lund, G., Andersson, L., Lauria, M., Lindholm, M., Fraga, M.F., Villar-Garea, A., Ballestar, E., Esteller, M. and Zaina, S. (2004). DNA methylation polymorphisms precede any histological sign of atherosclerosis in mice lacking apolipoprotein E. *J Biol Chem* 279, 29147–29154.

Maier, S., Nimmrich, I., Marx, A., Eppenberger-Castori, S., Jaenicke, F., Paradiso, A., Spyratos, F., Foekens, J., Schmitt, M. and Harbeck, N. (2004). DNA methylation profile predicts risk of recurrence in tamoxifen-treated, node-negative breast cancer patients. *J Clin Oncol (ASCO Annual Meeting Proceedings)* 22, 525.

Maier, S., Dahlstroem, C., Haefliger, C., Plum, A. and Piepenbrock, C. (2005). Identifying DNA methylation biomarkers of cancer drug response. *Am J Pharmacogenomics* 5, 223–232.

Maier, S., Nimmrich, I., Koenig, T., Eppenberger-Castori, S., Bohlmann, I., Paradiso, A., Spyratos, F., Thomssen, C., Mueller, V., Naehrig, J. et al. (2007). DNA-methylation of the homeodomain transcription factor PITX2 reliably predicts risk of distant disease recurrence in tamoxifen-treated, node-negative breast cancer patients – Technical and clinical validation in a multi-center setting in collaboration with the European Organisation for Research and Treatment of Cancer (EORTC) PathoBiology group. *Eur J Cancer* 43, 1679–1686.

Martens, J.W., Maier, S., Schwope, I., Koenig, T., Hartmann, O., Kluth, A., Look, M., Harbeck, N., Bolt-de Vries, J., Sieuwerts, A. et al. (2005). DNA-hypermethylation of PITX2 and risk of cancer recurrence in lymph node-negative hormone receptor-positive breast cancer patients. 96th Annual Meeting of the AACR. Abstract 2762.

Mohandas, T., Sparkes, R.S. and Shapiro, L.J. (1981). Reactivation of an inactive human X chromosome: Evidence for X inactivation by DNA methylation. *Science* 211, 393–396.

Nicholls, R.D. and Knepper, J.L. (2001). Genome organization, function, and imprinting in Prader-Willi and Angelman syndromes. *Annu Rev Genomics Hum Genet* 2, 153–175.

Oberle, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Boue, J., Bertheas, M.F. and Mandel, J.L. (1991). Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science* 252, 1097–1102.

Ogino, S., Kawasaki, T., Brahmandam, M., Cantor, M., Kirkner, G.J., Spiegelman, D., Makrigiorgos, G.M., Weisenberger, D.J., Laird, P.W., Loda, M. et al. (2006a). Precision and performance characteristics of bisulfite conversion and real-time PCR (MethyLight) for quantitative DNA methylation analysis. *J Mol Diagn* 8, 209–217.

Ogino, S., Cantor, M., Kawasaki, T., Brahmandam, M., Kirkner, G., Weisenberger, D.J., Campan, M., Laird, P.W., Loda, M. and Fuchs, C.S. (2006b). CpG island methylator phenotype (CIMP) of colorectal cancer is best characterized by quantitative DNA methylation analysis and prospective cohort studies. *Gut* 55, 1000–1006.

Okada, M., Ogasawara, H., Kaneko, H., Hishikawa, T., Sekigawa, I., Hashimoto, H., Maruyama, N., Kaneko, Y. and Yamamoto, N. (2002). Role of DNA methylation in transcription of human endogenous retrovirus in the pathogenesis of systemic lupus erythematosus. *J Rheumatol* 29, 1978–1982.

Okano, M., Bell, D.W., Haber, D.A. and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99, 247–257.

Paredes, J., Albergaria, A., Oliveira, J.T., Jeronimo, C., Milanezi, F. and Schmitt, F.C. (2005). P-cadherin overexpression is an indicator of clinical outcome in invasive breast carcinomas and is associated with CDH3 promoter hypomethylation. *Clin Cancer Res* 11, 5869–5877.

Paz, M.F., Yaya-Tur, R., Rojas-Marcos, I., Reynes, G., Pollan, M., Aguirre-Cruz, L., Garcia-Lopez, J.L., Piquer, J., Safont, M.J. and Balana, C. (2004). CpG island hypermethylation of the DNA repair enzyme methyltransferase predicts response to temozolomide in primary gliomas. *Clin Cancer Res* 10, 4933–4938.

Powell, S.M., Zilz, N., Beazer-Barclay, Y., Bryan, T.M., Hamilton, S.R., Thibodeau, S.N., Vogelstein, B. and Kinzler, K.W. (1992). APC mutations occur early during colorectal tumorigenesis. *Nature* 359, 235–237.

Richardson, B.C. (2002). Role of DNA methylation in the regulation of cell function: Autoimmunity, aging and cancer. *J Nutr* 132, 2401S–2405S.

Schuster, M. (2004). DNA methylation analysis: A powerful diagnostic tool. *J Lab Med* 28, 233–238.

Shimabukuro, M., Sasaki, T., Imamura, A., Tsujita, T., Fuke, C., Umekage, T., Tochigi, M., Hiramatsu, K., Miyazaki, T., Oda, T. et al. (2006). Global hypomethylation of peripheral leukocyte DNA in male patients with schizophrenia: A potential link between epigenetics and schizophrenia. *J Psychatr Res* e-publication.

Shinozaki, M., Hoon, D.S., Giuliano, A.E., Hansen, N.M., Wang, H.J., Turner, R. and Taback, B. (2005). Distinct hypermethylation profile of primary breast cancer is associated with sentinel lymph node metastasis. *Clin Cancer Res* 11, 2156–2162.

Stirzaker, C., Song, J.Z., Davidson, B. and Clark, S.J. (2004). Transcriptional gene silencing promotes DNA hypermethylation through a sequential change in chromatin modifications in cancer cells. *Cancer Res* 64, 3871–3877.

Takahashi, Y., Miyoshi, Y., Takahata, C., Irahara, N., Taguchi, T., Tamaki, Y. and Noguchi, S. (2005). Down-regulation of LATS1 and LATS2 mRNA expression by promoter hypermethylation and its association with biologically aggressive phenotype in human breast cancers. *Clin Cancer Res* 11, 1380–1385.

Takai, D. and Jones, P.A. (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci USA* 99, 3740–3745.

Taylor, K.H., Pena-Hernandez, K.E., Davis, J.W., Arthur, G.L., Duff, D.J., Shi, H., Rahmatpanash, F.B., Sjahputera, O. and Caldwell, C.W. (2007). Large-scale CpG methylation analysis identifies novel candidate genes and reveals methylation hotspots in acute lymphoblastic leukemia. *Cancer Res* 67, 2617–2625.

Toyota, M., Ho, C., Ahuja, N., Jair, K.W., Li, Q., Ohe-Toyota, M., Baylin, S.B. and Issa, J.P. (1999). Identification of differentially methylated sequences in colorectal cancer by methylated CpG island amplification. *Cancer Res* 59, 2307–2312.

Weksberg, R., Smith, A.C., Squire, J. and Sadowski, P. (2003). Beckwith-Wiedemann syndrome demonstrates a role for epigenetic control of normal development. *Hum Mol Genet* 12, R61–R68.

Wu, Y., Halverson, G., Basir, Z., Strawn, E., Yan, P. and Guo, S.W. (2005). Aberrant methylation at HOXA10 may be responsible for its aberrant expression in the endometrium of patients with endometriosis. *Am J Obstet Gynecol* 193, 371–380.

Yan, P.S., Chen, C.M., Shi, H., Rahmatpanah, F., Wei, S.H., Caldwell, C.W. and Huang, T.H. (2001). Dissecting complex epigenetic alterations in breast cancer using CpG island microarrays. *Cancer Res* 61, 8375–8380.

Zaina, S., Lindholm, M.W. and Lund, G. (2005). Nutrition and aberrant DNA methylation patterns in atherosclerosis: More than just hyperhomocystinemia?. *J Nutr* 135, 5–8.

Zhu, S., Goldschmidt-Clermont, P.J. and Dong, C. (2005). Inactivation of monocarboxylate transporter MCT3 by DNA methylation in atherosclerosis. *Circulation* 112, 1353–1361.

RECOMMENDED RESOURCES

Websites

http://www.epigenome.org – The website for the Human Epigenome Project.

http://www.microarrays.ca – This is the website for the microarray center at the University Health Network in Toronto. They describe a spotted array containing over 12,000 CpG island clones.

http://epigenome-noe.net/ – The website for the Epigenome Network of Excellence, a consortium of 25 European research groups.

http://www.dnamethsoc.com/ – Website for the Epigenetics Society.

Journal

Laird, P.W. (2003). The power and promise of DNA methylation markers. *Nat Rev Cancer* 3(4), 253–266. This paper provides a comprehensive review of methylation detection and measurement methods as well as a review of the literature applying these methods to clinical samples.

CHAPTER

DNA Microarrays in Biological Discovery and Patient Care

Andrew J. Yee and Sridhar Ramaswamy

INTRODUCTION

While "expression profiling" is now synonymous with DNA microarray technology, expression profiling dates back to the 1970s with techniques such as northern blotting for measuring RNA expression (Alwine et al., 1977). However, only one gene at a time could be queried with these modalities. Later developments enabled analysis of multiple genes (Liang and Pardee, 2003), including differential hybridization (Sargent, 1987), subtractive hybridization (Zimmermann et al., 1980), differential display (Liang and Pardee, 1992), and serial analysis of gene expression (SAGE) (Velculescu et al., 1995). However, these approaches required a significant amount of starting RNA and investment in labor and technical expertise, and were difficult to scale up.

DNA microarrays revolutionized the approach to gene expression profiling. Compared to prior methods, DNA microarrays were both dramatically high throughput and less cumbersome. It should be noted that the concept behind the microarray format was not new, as microarrays were originally developed as a technique for large-scale DNA mapping and sequencing (Hoheisel, 2006). However, changing the support surface from a porous membrane to a solid surface, such as glass, afforded significant improvements (Schena et al., 1998) by increasing reaction kinetics (Southern et al., 1999) and reducing background noise (Cheung et al., 1999).

As an illustration of approaches to define and develop applications of the transcriptome measured in biological samples, here we present specific issues related to the use of DNA microarrays, using applications especially in the context of cancer to illustrate the uses and challenges of gene expression profiling in genomic and personalized medicine.

MICROARRAY TECHNOLOGY

Microarray Platforms

Although the technology for DNA microarrays continues to evolve, the commonly used arrays can be divided into two main categories: (i) oligonucleotide and (ii) spotted or complementary DNA (cDNA) microarrays (see Figure 7.1). Over the years, commercial oligonucleotide platforms have gained in popularity over "homemade" spotted cDNA microarrays as costs have fallen considerably, to the point where commercial microarrays are the *de facto* platform. At the time of this writing, the dominant platforms are produced by Affymetrix and Agilent. For additional details, the reader is referred to a review by Hardiman (2004) and also to reviews collected in a series of *Nature Genetics* supplements (Bowtell, 1999; Cheung

Copyright © 2010, Elsevier Inc.
All rights reserved.

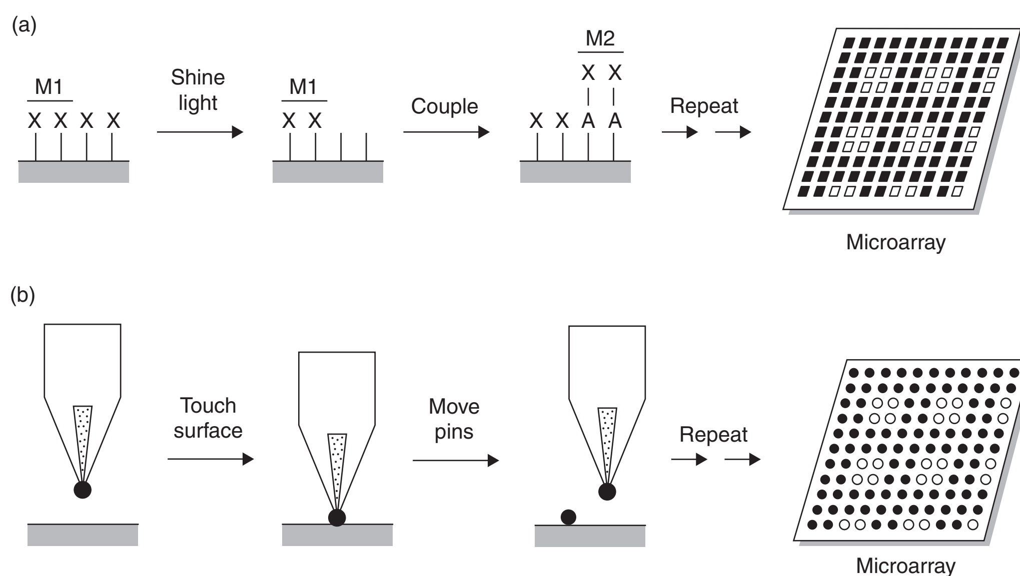


Figure 7.1 Comparison of microarray technologies. (a) Photolithography (as used by Affymetrix): a glass wafer modified with photolabile protecting groups (X) is selectively activated for DNA synthesis by shining light through a photomask (M1). The wafer is then flooded with a photoprotected DNA base (A–X), resulting in spatially defined coupling on the chip surface. A second photomask (M2) is used to deprotect defined regions of the wafer. Repeated deprotection and coupling cycles result in a highly dense microarray. (b) Spotted arrays: the gene of interest is loaded into a spotting pin by capillary action, and a small volume is transferred to a solid surface by physical contact between the pin and the solid substrate. After the first spotting cycle, the pin is washed and a second sample is loaded and deposited to an adjacent location. Robotic control systems and multiplexed printheads allow automated microarray fabrication. Adapted from Schena et al. (1998). (Reprinted with permission from Elsevier).

et al., 1999; Churchill, 2002; Duggan et al., 1999; Holloway et al., 2002; Lipshutz et al., 1999; Southern et al., 1999).

Oligonucleotide Arrays

Oligonucleotide microarrays are created either by *in situ* synthesis or deposition of presynthesized oligonucleotides ranging in size from 25- to 60-mers. Oligonucleotides can be synthesized directly *in situ* using photolithography techniques adapted from the microelectronics industry. The Affymetrix GeneChip platform is the primary example of this method, and because of its popularity, merits further discussion. Emerging from research carried out in the early 1990s (Fodor et al., 1991), this technology combines combinatorial, solid phase DNA synthetic chemistry with the benefits of photolithography. Successive photolithographic masks are used to select regions of the chip surface for exposure to light. Exposure to light deprotects photolabile groups on the nascent oligonucleotides undergoing synthesis. The light-mediated deprotection then allows these specified regions to couple with activated nucleoside monomers using standard DNA phosphoramidite synthetic chemistry. Each cycle thus extends the oligonucleotide by one base (see Figure 7.1).

Over the past 15 years, the feature size of each oligonucleotide probe or feature has been reduced 10-fold, resulting in a 1.6 cm^2 chip with 1.3 million unique features (multiple features are used to measure the expression of a particular gene). The detection sensitivity of these microarrays has been estimated at 1 in 300,000 RNAs (Lockhart et al., 1996). A practical limitation is the efficiency of each synthesis step – about 90–95%. This places a ceiling on *in situ* synthesis of an oligonucleotide to 25-mers. To control for non-specific binding, the probes on the Affymetrix chip are designed in pairs. One sequence is the exact complement of the target transcript, and the other paired sequence differs from the complement typically by one base pair near the middle of the probe. Difference in signal between the two probes can be used to control for non-specific binding and background contributions (Pease et al., 1994); 22 probes are routinely used for each expression measurement. Compared to cDNA arrays (see below), oligonucleotide arrays offer greater specificity and can distinguish single nucleotide polymorphisms (SNPs) and splice variants (Guo et al., 1994); the same technology is used in arrays for measuring DNA copy number and SNP profiling. The current generation of Affymetrix microarrays can measure the expression of over 47,000 transcripts.

Hybridization is detected by using a confocal laser scanning microscope to image the intensity of one color, fluorescently labeled samples. First, experimental mRNA is enzymatically amplified (see below for details of RNA amplification) and then labeled with biotin through the partial substitution of UTP and CTP with biotin-11-UTP and biotin-11-CTP. The labeled mRNA is then hybridized to the microarray and detected by the binding of a fluorescent compound (streptavidin–phycoerythrin) to the biotin-labeled samples. Because of the

robustness and consistency of the manufacturing process, the single-color readout provides an absolute quantitation of mRNA populations (Ishii et al., 2000). In addition, experiments can be compared with samples hybridized to different Affymetrix arrays.

Oligonucleotide microarrays may also be fabricated by using ink-jet printing technology. Nucleotide monomers are printed onto the chip and coupled using phosphoramidite chemistry (Hughes et al., 2001). Agilent has adopted this technology to create a 60-mer oligonucleotide array. Compared to the 25-mer photolithographic chip, a 60-mer array might offer increased sensitivity per probe. In contrast to the Affymetrix platform, there is only one probe per gene in this platform. The longer probes are also more tolerant of mismatches, improving analysis of polymorphic regions.

Spotted, Complementary DNA Microarrays

The first DNA microarrays were developed at Stanford University and based on cDNA samples spotted by a robotic arrayer at defined locations onto glass slides (Schena et al., 1995). These cDNA samples were generally PCR products generated from cDNA libraries or clone collections. The major benefit of spotted arrays is that they can be constructed in-house and tailored to the user's needs. The features are generally 100–300 µm in size and spaced at about the same distance. About 30,000 cDNAs can be spotted onto the surface of a microscope slide (Schulze and Downward, 2001). In the initial report, the sensitivity was estimated at 1 in 50,000 RNAs starting with 5 µg of total RNA (Schena et al., 1995). While spotted arrays offer customizability, they are better suited for smaller-scale experiments, since from a practical perspective managing large clone libraries and creating high-quality arrays consistently can be arduous. In the past, an advantage of spotted arrays was that *a priori* knowledge of the sequence being spotted was not necessary. However, this is less relevant now that complete sequence information for many experimental organisms and humans is readily accessible.

Gene expression using spotted microarrays is determined as a ratio of two samples rather than as an absolute measurement, owing to technical limitations and variability in spot size and probe concentration. Generally, cDNAs derived from experimental and reference RNA samples are labeled with deoxyribonucleotides coupled to different fluorophores (e.g., with Cy3 or Cy5) during first-strand cDNA synthesis. The two different populations are then competitively hybridized to the same array and imaged. While this method accurately compares expression levels between samples, comparing data from different platforms can be difficult, often requiring complex normalization algorithms (Dudley et al., 2002).

Sample Preparation

Tissue Sampling

For biological applications, tissue sampling is an important variable. For example, tumors are heterogeneous mixtures of different cell types, including malignant cells with varying degrees of differentiation, stromal elements, blood vessels, and inflammatory cells. Profiling of this whole sample may dilute out the "signal" profile of cancer cells with extraneous expression "noise" from surrounding tissues. Newer techniques such as laser capture microdissection (LCM) address the challenges of tissue heterogeneity by offering an approach for isolating homogeneous, individual cell populations (Emmert-Buck et al., 1996). While there are other approaches, such as fluorescence-activated cell sorting (FACS) and manual microdissection using micromanipulators (Emmert-Buck et al., 1994), LCM is faster, more versatile, and can be applied to solid tumors directly without the need for disaggregation into a cell suspension. Sgroi et al. (1999) first showed that it was feasible to use LCM to isolate subpopulations of breast cancer cells for expression profiling. Coupled with LCM has been the evolution of methodologies for purifying adequate amounts of high-quality RNA for expression profiling (see below). However, a theoretical limitation of focusing only on individual tissue components such as cancer cells relates to the growing appreciation that interactions with the tumor microenvironment (e.g., with stroma, endothelial, and immune cells) play a critical role in tumor progression. Sampling of "non" cancer cell components thus may be necessary for a profile to adequately reflect *in vivo* biology. An additional consideration is that the actual surgical manipulation itself may affect the expression profiling. Lin et al. (2006) compared expression profiles of prostate cancer from biopsies obtained *in situ* to biopsies obtained from prostates that were surgically removed. They found that 1.5% of measurable genes were altered as a result of the surgery, including transcripts for acute phase response proteins and regulators of cell proliferation.

RNA Isolation

Traditionally, microarray experiments have required roughly 50–200 µg of total RNA (Duggan et al., 1999). However, these quantities are not always easily obtainable, especially with techniques such as LCM. Protocols therefore have been developed for amplifying RNA from much smaller quantities, perhaps even down to the level of a single cell. For example, a single cell contains an estimated 0.1 pg of mRNA or 10 pg of total RNA, requiring an expansion of 10^8–10^9 fold in order to yield adequate amounts for profiling. Two commonly used amplification techniques include *in vivo* transcription and PCR amplification. *In vivo* transcription is a linear amplification approach, first described by Van Gelder et al. (1990). Messenger RNA is converted to double stranded cDNA using an oligo(dT) primer attached to a T7 promoter. T7 polymerase is then used to transcribe complementary copies of the cDNA population. One round of amplification can amplify the starting amount of mRNA by 1000-fold. A second or a third round of amplification can further increase the yield (Baugh et al., 2001). A second approach takes advantage of PCR amplification. PCR-based approaches have several advantages over linear amplification, including speed, cost, and theoretically unlimited degree of amplification. While it has been assumed that exponential amplification may lead to a biased representation of the starting

sample, results from one group suggest that this may not be as problematic as previously suggested (Iscove et al., 2002).

Standardization of Data

Critical to the success of data analysis is standardization of microarray data and annotation of the genes that comprise the microarrays. In 1999, the Microarray Gene Expression Database (MGED) group was created to establish guidelines for the format of microarray data. This led to the development of the Minimal Information About a Microarray Experiment (MIAME) standard (Brazma et al., 2001; Stoeckert et al., 2002). Most major scientific publications require compliance with MIAME standards and submission of data to public repositories.

An additional key factor is the functional annotation of the genes being studied. Biological "knowledge" is intricate, hierarchical, and interrelated. Ontologies are a formalized, standardized means of coding biological knowledge in which concepts are systematically described by both their meaning and their relationship to each other (Bard and Rhee, 2004). Adequate ontologies can thus serve as a common anchor point for systematic integration of data from various microarray sources. An example of commonly used ontology is one devised by the Gene Ontology Consortium (Harris et al., 2004).

DATA ANALYSIS

Gene expression studies pose many challenges for data organization, storage, and analysis (see Chapter 10). In concert with the development of microarray technology has been the evolution of sophisticated statistical methods for analyzing and interpreting highly dense microarray information (Ermolaeva et al., 1998; Quackenbush, 2001). These data can be one's own experimental data or existing data from other investigators' experiments, made available through public databases. Microarray data may be so rich in information that important findings may not be recognized initially, and only better appreciated in retrospect when re-analyzed or combined with newer data. Data analysis methodology is discussed further in Section 2 and has been reviewed also by Allison et al. (2006).

Unsupervised Learning

The computational analysis of gene expression data has largely centered on two approaches: unsupervised and supervised learning (see Figure 7.2). Unsupervised methods include hierarchical and *k*-means clustering (Gaasterland and Bekiranov, 2000). The goal of clustering is to organize genes and samples clusters based on relative similarity (D'Haeseleer, 2005; Eisen et al., 1998). A hierarchy of clusters can be created, which may be visualized as tree diagrams or dendrograms. Unsupervised approaches have the advantage of being unbiased and allow for the identification of structure in a complex dataset without making any *a priori* assumptions. In fact, many of the initial microarray studies on cancer samples utilized unsupervised classification, for example, lymphoma (Alizadeh et al., 2000) and breast cancer (Perou et al., 2000). However, because many different relationships are possible in a complex dataset, the predominant structure uncovered by clustering may not be clearly related to the most clinically or biologically interesting aspect of a dataset.

Supervised Learning

In contrast, supervised learning starts off with a known class distinction and uses this information upfront to select genes whose expression is best associated with this class distinction in a "training dataset" (Golub et al., 1999). This model is then applied to an independent "test" dataset to validate the accuracy of selected gene expression features in classification. Numerous supervised learning algorithms have been applied to gene expression datasets, such as *k*-nearest neighbors, neural networks, or support vector machines (Brown et al., 2000). The accuracy of supervised learning approaches heavily depends on the quality of the initial training data.

Mining

Several tools have been developed for distilling higher-level information from microarray data in order to better appreciate the "forest" among the innumerable "trees" of expression data. Two notable examples include modular analysis and Cancer Outlier Profile Analysis (COPA). Modular analysis examines the coordinated behavior of a set of genes in order to detect significant changes in modules of genes, rather than the individual genes themselves. This offers the ability to extract more "signal" from a given microarray experiment. Indeed, groups of gene expression may be more biologically interpretable and statistically robust than deciphering lists of individual genes. Mootha et al. (2003) used this approach to study changes in gene expression in the muscle of patients with type 2 diabetes. Their group introduced the gene set enrichment analysis (GSEA) and used the Kolmogorov–Smirnov test statistic, a type of non-parametric test statistic that does not make assumptions about population distribution. While the decrease in expression for individual genes was modest, around 20%, it was consistently decreased across a set of genes involved in oxidative phosphorylation, accounting for 89% of 106 genes. Thus, significant changes could be seen when looking at a group of genes, even when the expression of individual genes was not significantly different. GSEA has emerged as a useful tool for identifying biological processes from microarray data and has been refined into a robust method (Subramanian et al., 2005).

Similarly, Segal et al. (2004) created a cancer-focused compendium of expression profiles from 26 studies. They were able to extract 456 modules and found that some modules were unique to a particular type of tumor whereas other modules were shared across a range of conditions, raising the possibility that these modules reflect common tumor progression mechanisms. For example, in acute lymphoblastic leukemia (ALL), a growth-inhibitor module was specifically repressed. On the other hand, a bone osteoblastic module was found across a range of tumor types, suggesting a common mechanism for bone metastasis.

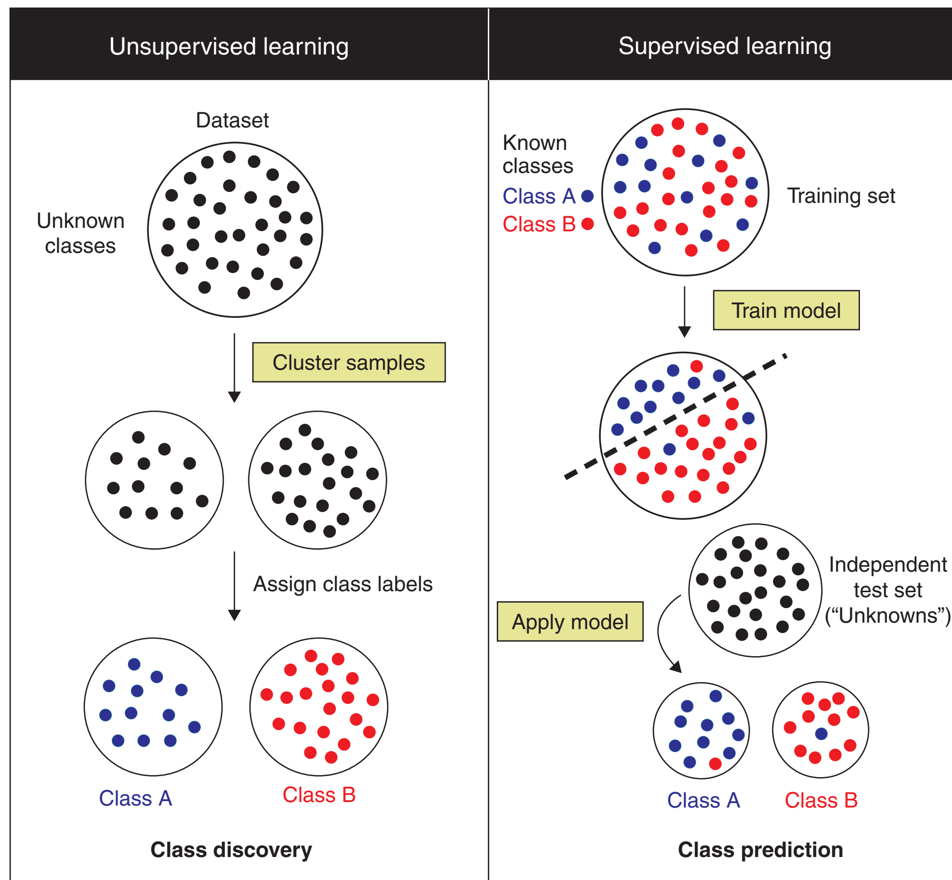


Figure 7.2 Unsupervised versus supervised learning. Unsupervised learning: multiple tumor samples are clustered into groups based on overall similarity of their gene expression profiles. This approach is useful for discovering previously unappreciated relationships. Supervised learning: multiple tumor samples from different known classes are used to train a model capable of classifying unknown samples. This model is then applied to a test set for class label assignment. Adapted from Ramaswamy and Golub (2002). (Reprinted with permission from the American Society of Clinical Oncology).

The COPA method focuses on genes characterized by marked overexpression (Tomlins et al., 2005). This strategy is used in order to efficiently process large amounts of data. COPA was applied to the Oncomine database, a compilation of 132 gene expression datasets representing 10,486 microarray experiments (Rhodes et al., 2004). This approach was instrumental for discovering two transcription factors with outlier profiles, *ERG* and *ETV1*, and subsequently shed light on recurrent gene fusions of the 5′ untranslated region of *TMPRS22* to *ERG* or *ETV1* in prostate cancer tissues. This study was notable for being the first study to identify recurrent gene re-arrangements in an epithelial cancer and suggested a novel mechanism for prostate cancer pathogenesis.

APPLICATIONS

DNA microarrays provide a panoramic, quantitative overview of a sample's expression output. The power of this is obvious, as biological processes generally result from the coordinated interaction of multiple genes.

Research with microarrays in humans can be divided into four categories:

1. normal tissue taxonomy;
2. disease diagnosis and classification;
3. disease prognostication; and
4. dissection of biological mechanisms.

Since the lion's share of research to date has been in cancer biology, in part due to the ready availability of tissue samples procured during routine diagnostic procedures, the following examples will draw largely from this field's literature.

Normal Tissue Taxonomy

Microarrays provide a new way of approaching the same fundamental question posed by Linnaeus in the 18th century: how is *A* different from and similar to *B*? An interesting example of previously unknown molecular heterogeneity was illustrated with human fibroblasts (Chang et al., 2002). Here, the authors demonstrated that the major factor responsible for differences in gene expression was the site of origin of the fibroblast. This

suggests that a detailed positional memory is encoded in a fibroblast's expression profile, and that different fibroblast subsets might be functionally distinct. Prior to the advent of microarray technology, this geographic variation was not appreciated. In addition to topographic location, the effect of aging on fibroblasts can also be measured and classified by expression profiling (Ly et al., 2000). In this study, using an Affymetrix oligonucleotide microarray, the expression profiles of fibroblasts isolated from young, middle-age, and old-age individuals and individuals with progeria were compared. The authors found consistent differences in the expression of genes between these groups in two functional classes: genes involved in cell cycle progression and genes involved in maintenance and remodeling of the extracellular matrix.

Disease Diagnosis and Classification

There have been a multitude of studies using expression profiling for disease diagnosis and classification, most notably in cancer. One seminal study used oligonucleotide microarrays to examine the expression of 6817 human genes in 72 patients with acute leukemia (Golub et al., 1999). Using unsupervised learning, gene expression was able to independently cluster leukemia samples into the known subsets of acute myelogenous leukemia (AML) and ALL (see Figure 7.3). Then, using supervised learning, gene sets that were differentially expressed in AML and ALL were used to correctly classify a group of unknown samples into the correct categories. Significantly, many markers that were both known, such as myeloperoxidase and terminal transferase, and unknown were useful for making this distinction. Although the difference between AML and ALL is often not difficult using modern histopathology, this study provided the first evidence that tumor expression profiles can be used for cancer classification.

Tumor cell lines can also be classified molecularly through expression profiling (Ross et al., 2000). The National Cancer Institute's Developmental Therapeutics Program bank of 60 cancer cell lines (NCI60) was profiled using a spotted cDNA array of 9703 cDNAs. The gene expression patterns observed in these cancer cell lines corresponded to the tissue of origin, which is remarkable since these cell lines have been selected for survival in the host and later in tissue culture. This illustrates a theme that the tissue of origin plays a major role in determining the expression profile. Similarly, a molecular taxonomy of tumor samples from 14 different tumor classes was created using a supervised learning algorithm (Ramaswamy et al., 2001). Some of these samples were from metastatic sites, and interestingly, the majority of them could be categorized from the classification information derived from the primary site, again demonstrating the importance of the tissue of origin in disease.

Breast cancer has been intensively profiled. Perou et al. (2000) reported a molecular classification system based on the expression patterns of 65 breast adenocarcinoma specimens from 42 individuals using a spotted cDNA array. Using an unsupervised method, these breast tumors were categorized into four different subtypes based on their patterns of gene expression. One clinically important subtype was already known (*Erb-B2* or *Her2/neu* overexpressing cancers), and three others were previously unknown: estrogen receptor-positive/luminal-like cancers, basal-like cancers, and normal breast. In addition to this classification system, the authors were able to show the importance of cellular lineage in determining the expression profile. This was illustrated by examining the variation in expression between primary tumors that were biopsied before and after a course of chemotherapy. They also looked at the profiles of two primary lymph node metastasis pairs. The authors found that the paired samples in the same patient were significantly more related to each other than to tumors from other patients, despite intervening chemotherapy or metastatic evolution. These initial observations were subsequently validated in a larger dataset (Sorlie et al., 2001, 2003). Importantly, these molecularly defined subtypes were clinically relevant, as patients with basal-like tumors had a significantly worse prognosis compared to patients with other breast cancer subtypes.

Outside of the realm of cancer, cDNA microarray technology has proved useful for exploring a broad range of diseases such as inflammatory diseases like rheumatoid arthritis and Crohn's disease (Heller et al., 1997), schizophrenia (Mirnics et al., 2000), multiple sclerosis (Lock et al., 2002), and differentiating between ischemic and non-ischemic cardiomyopathy (Kittleson et al., 2004). Microarrays have also been helpful for re-examining basic questions such as aging, as discussed previously (Ly et al., 2000).

Disease Prognosis

A challenge in medicine and especially in oncology is prediction–forecasting which patients will and will not benefit from therapy, particularly when the therapy, such as chemotherapy, has significant risks. DNA microarrays offer the possibility of radically transforming medicine from a traditional one-size-fits-all approach to individualized care. Intuitively, more information, when rich in content and appropriately analyzed, should lead to better prediction models. However, this has not always been borne out with various histopathologic tumor markers (e.g., breast and colon cancer), despite much initial excitement when these markers first became available (Bast et al., 2001). In contrast, there have been several microarray experiments involving different types of cancers that have found that the expression signature can serve as a robust biomarker, better than currently used criteria, for predicting prognosis. A common approach is to identify an expression profile that can subclassify a tumor type and then find correlations between this subclassification and prognosis.

The first such application was described by Alizadeh et al. (2000) in diffuse large B cell lymphoma (DLBCL). These investigators constructed the Lymphochip, a specialized spotted cDNA microarray enriched for genes preferentially expressed in lymphoid cells and for genes with known roles in immunology or cancer. Using this microarray, hierarchical clustering revealed two molecularly distinct forms of DLBCL: germinal center B-like DLBCL and activated B-like DLBCL. These subtypes were not

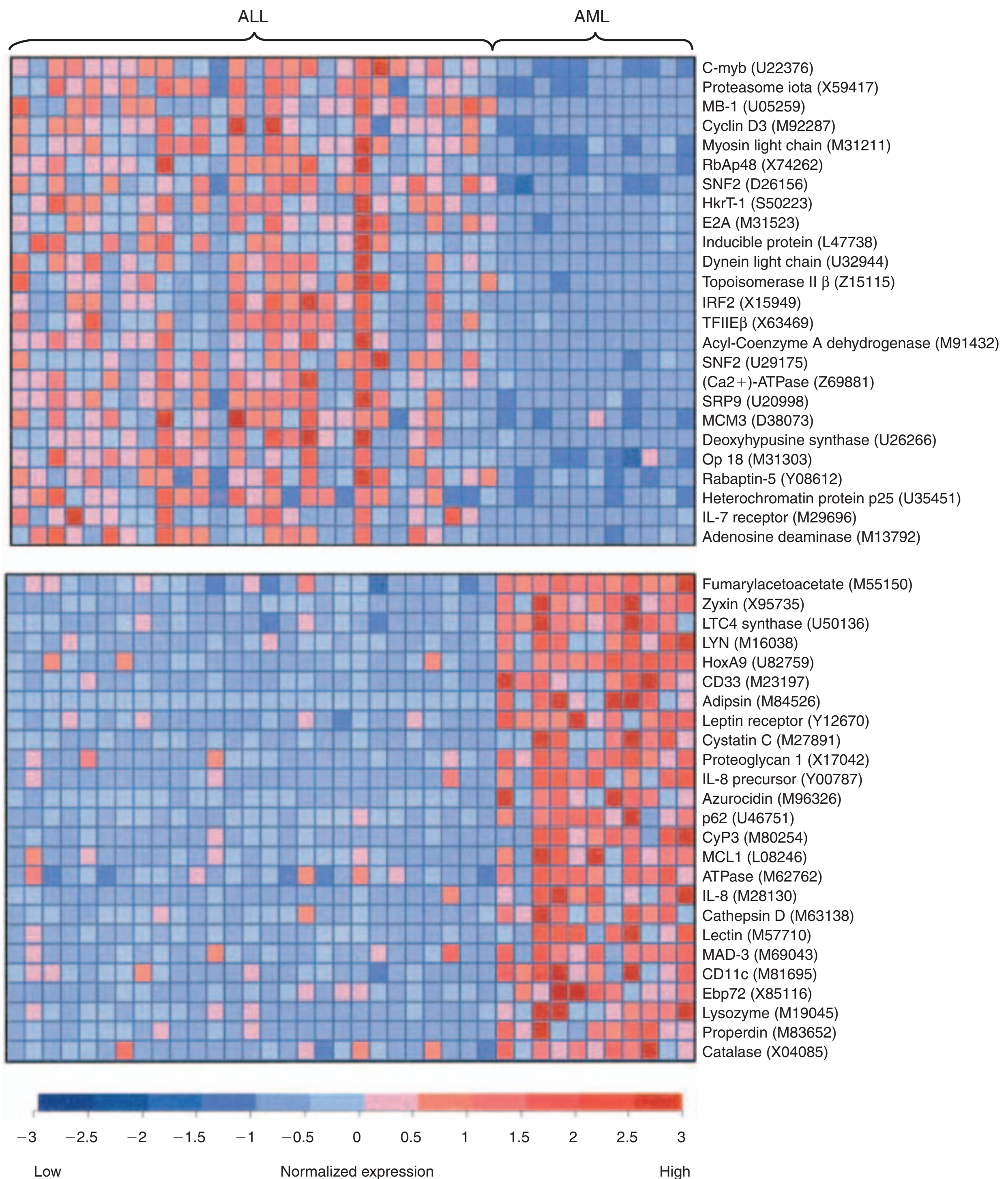


Figure 7.3 Gene expression profiling in acute leukemia. The 50 genes most highly correlated with the distinction between ALL and AML are shown. Each row corresponds to a gene, and each column corresponds to expression levels in different samples. Expression levels greater than the mean are shaded in red, and those below the mean are shaded in blue. The scale indicates SDs above or below the mean. The top panel shows genes highly expressed in ALL; the bottom panel shows genes more highly expressed in AML. Adapted from Golub et al. (1999). (Reprinted with permission from AAAS).

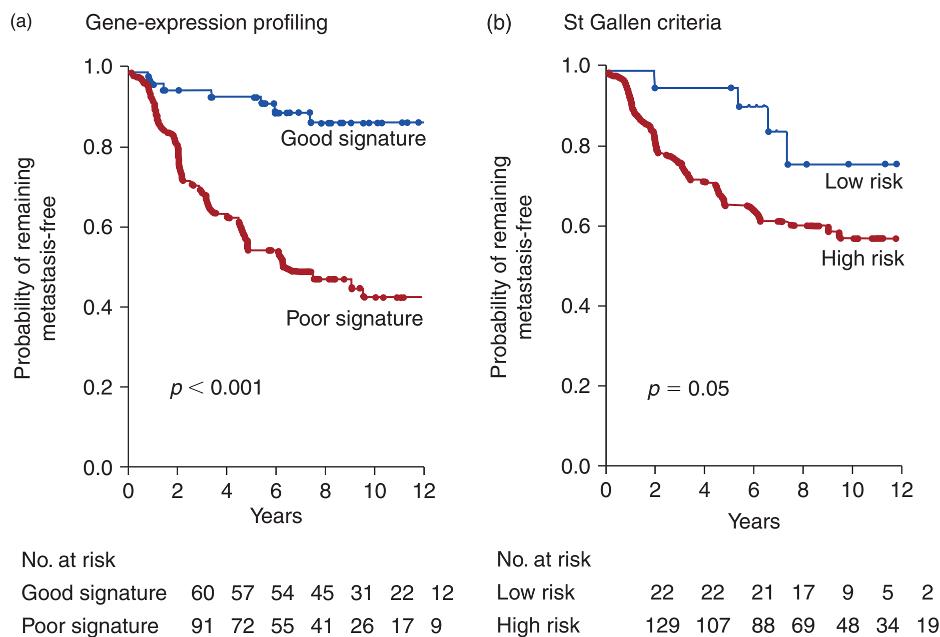


Figure 7.4 Kaplan–Meier analysis of the probability that patients would remain free of distant metastases among 151 patients with lymph node-negative breast cancer by gene expression profiling (a) and by St Gallen clinical criteria (Goldhirsch et al., 2001) (b). Adapted from Van de Vijver et al. (2002). (Copyright © 2002, Massachusetts Medical Society. All rights reserved).

appreciated using traditional histopathology and immunohistochemical techniques, yet nevertheless have clinical significance as patients with germinal center B-like DLBCL had significantly better overall survival when treated with standard chemotherapy. Indeed, as observed with other initial studies in cancer classification, cell lineage was an important determinant of the molecular phenotype. Ultimately, it remains to be seen whether or not molecular phenotyping will supplant traditional diagnostic tools, and this remains an ongoing area of research.

Similar predictive power has been exemplified in microarray studies of breast cancer. Traditional factors for predicting a patient's outcome, such as tumor size, lymph node status, and estrogen receptor status, are imperfect. Van't Veer et al. (2002) used an oligonucleotide microarray to analyze primary tumors from 117 patients with early-stage breast cancer and no lymph node involvement, who received minimal treatment after surgery, to identify a 70-gene expression signature that strongly predicted metastasis. In a subsequent validation study on 295 patients (which included the original cohort of 117 patients), this signature outperformed all currently used clinical predictors and more accurately identified patients at risk for distant metastasis (van de Vijver et al., 2002) (see Figure 7.4). For example, the profile could assign more patients into a low-risk group for metastasis (40%) compared to traditional clinical criteria (15%). These initial findings were subsequently confirmed in an independent group of patients from five European centers (Buyse et al., 2006). Taken together, these results suggested that primary tumor gene expression profiles can be used to molecularly stratify cancer patients according to risk of progression and treatment outcome.

LIMITATIONS AND CHALLENGES

While microarray technology continues to mature, as with any technology, expression profiles must be interpreted in the context of broader biological knowledge. For example, one assumption often made is that gene expression correlates well with protein quantity and therefore protein activity. This view does not take into account additional variables such as mRNA stability, protein degradation, or post-translational modifications that ultimately determine protein activity. Another supposition is that mRNA expression levels correlate well with biological activity. In contrast, small differences in expression may have a dramatic biological effect. For example, the expression level of transcription factors is often quite low even in tissues where these factors play active roles. Finally, expression profiles are not fixed; they may vary dramatically in different physiological contexts (King and Sinha, 2001).

Validation with Non-microarray Technologies

Gene expression data generated from microarray experiments can be credentialed using other methods of expression measurement such as northern blotting (Taniguchi et al., 2001), real-time PCR (Dallas et al., 2005). These studies and others demonstrate that genes found to be expressed with current generation microarrays can be consistently verified. Microarray data can also be verified at the protein expression level. A particularly useful high-throughput technology for validation using immunohistochemistry involves tissue microarrays (Hans et al., 2004).

Reproducibility of Microarray Data

An ongoing concern since the inception of the DNA microarray has been its technical reproducibility (Marshall, 2004). For a general review of these initial concerns, the interested reader is referred to a review by Chuaqui et al. (2002). While initial studies raised concerns about the reproducibility of microarray data (e.g., Tan et al. [2003]), subsequent studies have found that reproducibility is better than initially described (Bammler et al., 2005; Irizarry et al., 2005; Larkin et al., 2005). Recently, the MicroArray Quality Control (MAQC) project involved 137 participants from 51 academic, government, and commercial institutions to assess the performance of seven microarray platforms of two commercially available RNA samples (Shi et al., 2006). Reassuringly, the MAQC project showed high intra-platform consistency across test sites as well as a high level of inter-platform concordance in terms of identifying differentially expressed genes in a common set of biological samples.

FUTURE DIRECTIONS

Newer Platform Technologies

Alternatives to conventional microarray platforms are emerging. One notable example is a bead-based platform developed by Illumina, the BeadChip. This uses randomly assembled arrays of 50-mer coated 3-μm beads in order to create an extremely dense platform that is denser than spotted arrays or photolithographically made arrays (Gunderson et al., 2004). Whereas in conventional arrays, the sequence and location of each probe is known beforehand, the identity (and the corresponding probe sequence) of the beads is decoded after deposition. A DNA-based decoding algorithm is used to identify each bead. Compared to conventional platforms, the main advantage of this platform is lower per-sample cost.

A different type of platform departing from conventional technology is the CombiMatrix platform. This platform synthesizes oligonucleotides *in situ* on a semiconductor chip using routine phosphoramidite chemistry under electrochemical control (Liu et al., 2006). Each feature on the chip is digitally addressed and controlled. When activated, an electrochemical reaction generating acid occurs, ultimately leading to oligonucleotide synthesis *in situ*. The main advantage of this technology is flexibility. Custom microarrays can be created on demand with less labor, whereas technologies with "fixed" content such as the Affymetrix chip require a significant investment of resources to create a new chip (e.g., creation of a new photolithographic mask).

Formalin-Fixed, Paraffin-Embedded Samples

Traditionally, obtaining RNA from human tissues for expression profiling has required that specimens be snap-frozen in liquid nitrogen within a half hour of resection and stored at −80°C or colder to minimize RNA degradation. Samples need to be processed quickly because changes in some mRNA species have been noted even a few minutes after manipulation and devascularization of a tissue (Huang et al., 2001). However, clinical specimens are generally fixed in formalin and embedded in paraffin for routine histopathologic analysis, which causes significant RNA degradation. Only a fraction of the RNA from fixed specimens, roughly 3%, is intact enough for cDNA synthesis, posing a major barrier to obtaining expression profiles from clinical specimens (Godfrey et al., 2000; Masuda et al., 1999).

A method to maneuver around this degradation obstacle is the cDNA-mediated, annealing, selection, extension, and ligation (DASL) assay. This assay is a technology that combines the advantages of microarray-based gene expression profiling with multiplexed quantitative PCR (Fan et al., 2004). An upstream and downstream pair of oligonucleotides is designed to amplify a cDNA of interest. Each pair of oligonucleotides incorporates a universal primer sequence for PCR as well as a specific address sequence that hybridizes to a capture sequence on an array. With this methodology, cancer-specific expression profiles were generated with as little as 50 ng of total RNA from formalin-fixed tissues stored over a decade (Bibikova et al., 2004). Approaches such as the DASL assay thus have the potential to capitalize on the vast supply of archived, paraffin-embedded samples from large clinical trials or population-based cohort studies to advance expression profiling efforts.

Functional Genomics

A goal of functional genomics is to assign biological meaning to genes on a genome-wide level (Steinmetz and Davis, 2004). As illustrated by the examples described earlier, expression profiling has proved useful for assigning functional groups to genes in model organisms such as yeast (Hughes et al., 2000) and for uncovering a role for metastasis with *RhoC* (Clark et al., 2000). Expression profiling has also been used to decipher regulatory networks of transcription by discovering common sequence motifs in upstream regions of genes that have similar expression profiles (Pilpel et al., 2001). These studies, though, have been mainly been done in yeast and await application in more complex systems such as humans.

A new opportunity in functional genomics is to integrate expression profiling with RNA interference (RNAi) technologies now that genome-wide RNA interference libraries are coming to fruition (Moffat et al., 2006). One can imagine using RNAi libraries to systematically knock down individual genes and the comparing the resulting expression phenotypes. One can then look for similarities and differences between expression profiles as a starting point for inferring the function of the genes that have been ablated by RNAi.

Profiling the Transcriptome

Expression profiling has also been used to explore the transcriptome. The transcriptome represents the complete set of all the transcribed elements of the genome in contrast to the collection of transcripts that are only translated into proteins. There is a burgeoning interest in the function of the non-coding part of the transcriptome as the number of actual genes transcribed into

proteins has been substantially revised downward and increasing numbers of non-coding transcripts have been discovered. Key to their detection has been tiling microarrays, specifically those developed by Affymetrix with photolithographic technology. Rather than only using probes for known exons, tiling arrays probe across the entire genome in an unbiased fashion. Cheng et al. (2005) used a 25-mer oligonucleotide-based array spaced every 5 bp to examine ten human chromosomes. They discovered that interestingly, of all the transcribed sequences, 19% were polyadenylated, 44% were not polyadenylated, and 37% were both polyadenylated and not polyadenylated. Additionally, about 10% of the genome was transcribed as polyadenylated transcripts, compared to 1–2% of the genome corresponding to exons. These findings suggest that the expressed genome may be orders of magnitude larger than previously appreciated.

Part of the transcriptome also includes small (roughly 22 nucleotide) non-coding microRNAs (miRNAs). These miRNAs play an important role in regulating gene expression by either degrading mRNA or inhibiting its translation (Bartel, 2004). In cancer, for example, Lu et al. (2005) used a new bead-based profiling system to quantify miRNA expression because of the short size of the miRNAs. This used oligonucleotide capture probes that were coupled to beads, which in turn were differentially impregnated with variable degrees of a fluorescent mixture. Flow cytometry was then used to quantitate the beads and therefore the associated miRNA. Expression profiling of the miRNA population of various malignant and normal tissues revealed an unusually large degree of diversity of miRNA expression across cancers. Furthermore, miRNA signatures could be used to classify poorly differentiated cancers, which have traditionally been difficult to classify using standard mRNA-based profiling strategies. The interested reader is referred to a recent review by Calin and Croce (2006).

Clinical Translation

Expression Profiling as a Biomarker

The ability of expression signatures of tumor samples to predict prognosis has immediate potential clinical translation. A key lesson from these early clinical microarray studies is that appropriate sample size and careful experimental design are necessary for identifying signatures that can serve as robust biomarkers (Michiels et al., 2005; Ntzani and Ioannidis, 2003; Simon et al., 2003). However, reassuringly, a recent study of prognostic breast cancer signatures showed that four out of five reported gene signatures had significant concordance in predicting breast cancer outcome, indicating that these signatures were tracking similar biological behaviors (Fan et al., 2006), pointing to the evolving sophistication of clinically oriented profiling efforts.

The clinical application of expression profiles is most mature in breast cancer, owing to the extensive microarray data already available, and has recently been reviewed (Sotiriou and Piccart, 2007). The 70-gene-expression signature originally formulated by Van't Veer et al. (2002) has been commercialized into the MammaPrint assay for clinical use. This assay serves as the basis of the MINDACT (Microarray In Node negative Disease may Avoid ChemoTherapy) clinical trial, which aims to determine whether the expression signature can be used for making clinical decisions (Novak, 2006). It will test the question of whether or not patients with tumors that have high risk features by clinical criteria – individuals who are routinely recommended adjuvant chemotherapy – but with a low-risk molecular prognosis as determined by MammaPrint, can defer chemotherapy without affecting metastasis-free survival. In other words, it will test if the molecular prognosis generated by MammaPrint can outperform clinical criteria.

An alternative to using existing signatures is to adapt the expression profile into a RT-PCR-based assay. Compared to DNA microarray technology, RT-PCR has the potential to be less technically challenging, more cost-effective, and more easily integrated into routine clinical practice. Genomic Health International has developed the Oncotype assay which is currently available for clinical use (Paik et al., 2004). The Oncotype assay is an RT-PCR assay based on a 21-gene signature (including five control genes) derived from 250 candidate genes selected from the published microarray literature. Rather than using fresh frozen tissues, the researchers were able to exploit archived, paraffin-embedded tissues using RT-PCR as their profiling method (Cronin et al., 2004). They systematically quantified the expression of these 250 candidate genes across three historical groups of breast cancer patients. From this data, they derived a gene expression-based recurrence score for predicting breast cancer recurrence that was more powerful than standard histopathologic criteria. The reproducibility of the score was robust and was subsequently validated in a large, population-based study (Habel et al., 2006). Furthermore, Oncotype had the potential to identify patients who may be able to defer adjuvant chemotherapy treatment. Patients with low and intermediate recurrence score tumors did not appear to derive a significant benefit from chemotherapy (Paik et al., 2006). An advantage of using this RT-PCR-based approach is that it can be applied to tumor samples that have been collected and processed in routine fashion; no special handling (e.g., freezing tissues in liquid nitrogen) is required. Other examples of adapting expression profiling to an RT-PCR assay include a two-gene breast cancer signature (Ma et al., 2003) and a six-gene signature predicting survival in DLBCL (Lossos et al., 2004).

In the future, one of the challenges for translational research will be the integration of expression profiling with routine histopathology in clinical practice. As can be seen with the 21-gene breast cancer signature, expression profiling in this application offers the advantage of reproducibility and minimizes human error and variability with pathological interpretation. However, cost continues to be a major limitation to widespread adoption of profiling technologies. An ongoing prospective study, similar in concept to MINDACT, is the TAILORx (Trial Assigning Individualized Options for Treatment) trial in North America, which will help answer the question of how to integrate microarray technology (i.e., Oncotype DX) into clinical practice for patients with breast cancer. TAILORx is examining whether chemotherapy is required for the intermediate-risk group defined by Oncotype's recurrence score (Paik, 2007).

Drug Development

DNA microarrays have also facilitated drug target discovery and prediction of drug response. For general reviews, the reader is referred to a recent one by Stoughton and Friend (2005). Using the NCI60 cancer cell line panel, Scherf et al. (2000) constructed a database that correlated gene expression and drug activity patterns. They were able to find relationships between gene expression and drugs such as 5-fluorouracil and L-asparaginase. Gunther et al. (2003) employed a similar strategy with primary human neuronal precursor cells and various psychotropic drugs. Potti et al. (2006) have expanded on this approach by using the expression profiles of sensitive and resistant cancer cell lines for six different chemotherapy drugs to predict response to neoadjuvant chemotherapy for breast cancer.

Stegmaier et al. (2004) have developed a high-throughput approach to drug discovery using expression profiling: gene expression-based high-throughput screening (GE-HTS). These authors were interested in compounds that could terminally differentiate AML cells. To discover these compounds, they used 5-gene-expression signature of the differentiated neutrophil and leukemic states as a surrogate. With an RT-PCR-based readout of these five genes, they then screened a small molecule library to identify drugs that could induce differentiation of leukemic blasts and discovered an epidermal growth factor receptor kinase inhibitor with this property (Stegmaier et al., 2005). Remarkably, GE-HTS discovered this potential therapeutic without prior detailed biological knowledge.

A related approach to screening small molecules has been implemented through The Connectivity Map, an *in silico* tool recently developed at the Broad Institute that uses expression profiling to systematically discover functional connections among diseases, genetic perturbations, and drug actions (Lamb et al., 2006). Lamb et al. first created a reference collection of gene expression profiles from cultured human cancer cells treated with small molecules. They then developed pattern-matching software to mine these data. The Connectivity Map has been used to perform *in silico* discoveries of rapamycin as a new approach for treating glucocorticoid-resistant ALL (Wei et al., 2006) and the inhibition of HSP90 as a novel mechanism for inhibition of androgen receptor signaling in prostate cancer (Hieronymus et al., 2006).

2009 UPDATE

Expression profiling using microarray technology is now a standard tool in biological and medical discovery. At the technical level, recent years have seen dramatic advances in the ability to efficiently perform deep, comprehensive sequencing (Shendure and Ji, 2008).

The power of this next generation of sequencing has recently been applied to expression profiling (Wang et al., 2009). This approach, known as “RNA-seq”, typically starts with the conversion of RNA samples to cDNA fragments. These cDNA fragments are then sequenced in a massively parallel, high-throughput manner using platforms from companies such as Applied Biosystems and Illumina to obtain short sequences at great sampling depth. There are several significant advantages to evaluating RNA expression at the individual RNA level instead of the relative amounts measured by hybridization from traditional microarrays. These include lower background signal and significantly improved dynamic range (generally five orders of magnitude, which is three orders higher than microarrays). This is especially useful for quantifying genes expressed at very low or very high levels. Starting requirements for RNA are also lower for RNA-seq than for conventional microarrays. RNA-seq is beginning to be used for exploring basic biological questions, for example, characterizing mouse embryonic stem cells (Cloonan et al., 2008) and discovering new splice events, additional promoters, and untranscribed regions in the mouse transcriptome (Mortazavi et al., 2008).

There continues to be a proliferation of expression profiling data in translational and clinical research. Novel insights from basic cancer biology are being explored through microarrays in clinical settings, for example, profiles of breast cancer “stem cells” (Liu et al., 2007) and profiling of the stroma in diffuse large B-cell lymphoma (Lenz et al., 2008). Application to standard formalin-fixed, paraffin-embedded clinical specimens for microarray analysis is expanding with improving methods, for example, obtaining gene expression in hepatocellular carcinoma (Hoshida et al., 2008). Using microarray data for clinical decision making continues to be challenging. It is furthest along in breast cancer, where the Oncotype DX assay has been recently incorporated into the American Society of Clinical Oncology recommendations for use of tumor markers (Harris et al., 2007). It is gaining increasing use in the clinic for identifying patients with node-negative, estrogen-receptor positive breast cancer who may derive the most benefit from tamoxifen, sparing them from the potential toxicities of adjuvant chemotherapy. Results from the ongoing Trial Assigning IndividuaLized Options for Treatment (Rx) (TAILORx) trial will be key to determining how to effectively incorporate expression profiling information into the treatment of early stage breast cancer (Sparano, 2006).

Looking towards the future, at the time of this writing, there is excitement surrounding the application of RNA-seq technology to more clinically focused questions. Indeed, one example illustrating the power of deep sequencing, in general, was in the genomic analysis of a patient with cytogenetically normal, acute myeloid leukemia (Ley et al., 2008). In the months ahead, we anticipate that RNA-seq will change the landscape of expression profiling.

2009 UPDATE REFERENCES

Hoshida, Y., Villanueva, A., Kobayashi, M., Peix, J., Chiang, D. Y., Camargo, A., Gupta, S., Moore, J., Wrobel, M. J., Lerner, J., et al. (2008). Gene expression in fixed tissues and outcome in hepatocellular carcinoma. *N Engl J Med* 359, 1995–2004.

Lenz, G., Wright, G., Dave, S.S., Xiao, W., Powell, J., Zhao, H., Xu, W., Tan, B., Goldschmidt, N., Iqbal, J., et al. (2008). Stromal gene signatures in large-B-cell lymphomas. *N Engl J Med* 359, 2313–2323.

Ley, T.J., Mardis, E.R., Ding, L., Fulton, B., McLellan, M.D., Chen, K., Dooling, D., Dunford-Shore, B.H., McGrath, S., et al. (2008). DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 456, 66–72.

Liu, R., Wang, X., Chen, G.Y., Dalerba, P., Gurney, A., Hoey, T., Sherlock, G., Lewicki, J., Shedden, K. and Clarke, M.F. (2007). The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N Engl J Med* 356, 217–226.

Shendure, J. and Ji, H. (2008). Next-generation DNA sequencing. *Nat Biotechnol* 26, 1135–1145.

Sparano, J.A. (2006). TAILORx: trial assigning individualized options for treatment (Rx). *Clin Breast Cancer* 7, 347–350.

Wang, Z., Gerstein, M. and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10, 57–63.

REFERENCES

Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X. et al. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403, 503–511.

Allison, D.B., Cui, X., Page, G.P. and Sabripour, M. (2006). Microarray data analysis: From disarray to consolidation and consensus. *Nat Rev Genet* 7, 55–65.

Alwine, J.C., Kemp, D.J. and Stark, G.R. (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc Natl Acad Sci USA* 74, 5350–5354.

Bammler, T., Beyer, R.P., Bhattacharya, S., Boorman, G.A., Boyles, A., Bradford, B.U., Bumgarner, R.E., Bushel, P.R., Chaturvedi, K., Choi, D. et al. (2005). Standardizing global gene expression analysis between laboratories and across platforms. *Nat Methods* 2, 351–356.

Bard, J.B. and Rhee, S.Y. (2004). Ontologies in biology: Design, applications and future challenges. *Nat Rev Genet* 5, 213–222.

Bartel, D.P. (2004). MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297.

Bast, R.C., Jr., Ravdin, P., Hayes, D.F., Bates, S., Fritsche, H., Jr., Jessup, J.M., Kemeny, N., Locker, G.Y., Mennel, R.G. and Somerfield, M.R. (2001). 2000 update of recommendations for the use of tumor markers in breast and colorectal cancer: Clinical practice guidelines of the American Society of Clinical Oncology. *J Clin Oncol* 19, 1865–1878.

Baugh, L.R., Hill, A.A., Brown, E.L. and Hunter, C.P. (2001). Quantitative analysis of mRNA amplification by in vitro transcription. *Nucleic Acids Res* 29, E29.

Bibikova, M., Talantov, D., Chudin, E., Yeakley, J.M., Chen, J., Doucet, D., Wickham, E., Atkins, D., Barker, D., Chee, M. et al. (2004). Quantitative gene expression profiling in formalin-fixed, paraffin-embedded tissues using universal bead arrays. *Am J Pathol* 165, 1799–1807.

Bowtell, D.D. (1999). Options available–from start to finish–for obtaining expression data by microarray. *Nat Genet* 21, 25–32.

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C. et al. (2001). Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 29, 365–371.

Brown, M.P., Grundy, W.N., Lin, D., Cristianini, N., Sugnet, C.W., Furey, T.S., Ares, M., Jr. and Haussler, D. (2000). Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc Natl Acad Sci USA* 97, 262–267.

Buyse, M., Loi, S., van't Veer, L., Viale, G., Delorenzi, M., Glas, A.M., d'Assignies, M.S., Bergh, J., Lidereau, R., Ellis, P. et al. (2006). Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst* 98, 1183–1192.

Chang, H.Y., Chi, J.T., Dudoit, S., Bondre, C., van de Rijn, M., Botstein, D. and Brown, P.O. (2002). Diversity, topographic differentiation, and positional memory in human fibroblasts. *Proc Natl Acad Sci USA* 99, 12877–12882.

Chang, H.Y., Sneddon, J.B., Alizadeh, A.A., Sood, R., West, R.B., Montgomery, K., Chi, J.T., van de Rijn, M., Botstein, D. and Brown, P.O. (2004). Gene expression signature of fibroblast serum response predicts human cancer progression: Similarities between tumors and wounds. *PLoS Biol* 2, E7.

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., Helt, G. et al. (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* 308, 1149–1154.

Cheung, V.G., Morley, M., Aguilar, F., Massimi, A., Kucherlapati, R. and Childs, G. (1999). Making and reading microarrays. *Nat Genet* 21, 15–19.

Chuaqui, R.F., Bonner, R.F., Best, C.J., Gillespie, J.W., Flaig, M.J., Hewitt, S.M., Phillips, J.L., Krizman, D.B., Tangrea, M.A., Ahram, M. et al. (2002). Post-analysis follow-up and validation of microarray experiments. *Nat Genet* 32(Suppl), 509–514.

Churchill, G.A. (2002). Fundamentals of experimental design for cDNA microarrays. *Nat Genet* 32(Suppl), 490–495.

Clark, E.A., Golub, T.R., Lander, E.S. and Hynes, R.O. (2000). Genomic analysis of metastasis reveals an essential role for RhoC. *Nature* 406, 532–535.

Cloonan, N., Forrest, A.R., Kolle, G., Gardiner, B.B., Faulkner, G.J., Brown, M.K., Taylor, D.F,. Steptoe, A.L., Wani, S., Bethel, G., et al. (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods*. 5(7):613–619.

Cronin, M., Pho, M., Dutta, D., Stephans, J.C., Shak, S., Kiefer, M. C., Esteban, J.M. and Baker, J.B. (2004). Measurement of gene

expression in archival paraffin-embedded tissues: Development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay. *Am J Pathol* 164, 35–42.

D'Haeseleer, P. (2005). How does gene expression clustering work?. *Nat Biotechnol* 23, 1499–1501.

Dallas, P.B., Gottardo, N.G., Firth, M.J., Beesley, A.H., Hoffmann, K., Terry, P.A., Freitas, J.R., Boag, J.M., Cummings, A.J. and Kees, U.R. (2005). Gene expression levels assessed by oligonucleotide microarray analysis and quantitative real-time RT-PCR – how well do they correlate?. *BMC Genomics* 6, 59.

Dudley, A.M., Aach, J., Steffen, M.A. and Church, G.M. (2002). Measuring absolute expression with microarrays with a calibrated reference sample and an extended signal intensity range. *Proc Natl Acad Sci USA* 99, 7554–7559.

Duggan, D.J., Bittner, M., Chen, Y., Meltzer, P. and Trent, J.M. (1999). Expression profiling using cDNA microarrays. *Nat Genet* 21, 10–14.

Ebert, B.L. and Golub, T.R. (2004). Genomic approaches to hematologic malignancies. *Blood* 104, 923–932.

Eisen, M.B., Spellman, P.T., Brown, P.O. and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95, 14863–14868.

Emmert-Buck, M.R., Bonner, R.F., Smith, P.D., Chuaqui, R.F., Zhuang, Z., Goldstein, S.R., Weiss, R.A. and Liotta, L.A. (1996). Laser capture microdissection. *Science* 274, 998–1001.

Emmert-Buck, M.R., Roth, M.J., Zhuang, Z., Campo, E., Rozhin, J., Sloane, B.F., Liotta, L.A. and Stetler-Stevenson, W.G. (1994). Increased gelatinase A (MMP-2) and cathepsin B activity in invasive tumor regions of human colon cancer samples. *Am J Pathol* 145, 1285–1290.

Ermolaeva, O., Rastogi, M., Pruitt, K.D., Schuler, G.D., Bittner, M.L., Chen, Y., Simon, R., Meltzer, P., Trent, J.M. and Boguski, M.S. (1998). Data management and analysis for gene expression arrays. *Nat Genet* 20, 19–23.

Fan, C., Oh, D.S., Wessels, L., Weigelt, B., Nuyten, D.S., Nobel, A.B., van't Veer, L.J. and Perou, C.M. (2006). Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* 355, 560–569.

Fan, J.B., Yeakley, J.M., Bibikova, M., Chudin, E., Wickham, E., Chen, J., Doucet, D., Rigault, P., Zhang, B., Shen, R. et al. (2004). A versatile assay for high-throughput gene expression profiling on universal array matrices. *Genome Res* 14, 878–885.

Fodor, S.P., Read, J.L., Pirrung, M.C., Stryer, L., Lu, A.T. and Solas, D. (1991). Light-directed, spatially addressable parallel chemical synthesis. *Science* 251, 767–773.

Gaasterland, T. and Bekiranov, S. (2000). Making the most of microarray data. *Nat Genet* 24, 204–206.

Godfrey, T.E., Kim, S.H., Chavira, M., Ruff, D.W., Warren, R.S., Gray, J.W. and Jensen, R.H. (2000). Quantitative mRNA expression analysis from formalin-fixed, paraffin-embedded tissues using 5′ nuclease quantitative reverse transcription-polymerase chain reaction. *J Mol Diagn* 2, 84–91.

Goldhirsch, A., Glick, J.H., Gelber, R.D., Coates, A.S. and Senn, H.J. (2001). Meeting highlights: International Consensus Panel on the Treatment of Primary Breast Cancer. Seventh International Conference on Adjuvant Therapy of Primary Breast Cancer. *J Clin Oncol* 19, 3817–3827.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A. et al. (1999). Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* 286, 531–537.

Gunderson, K.L., Kruglyak, S., Graige, M.S., Garcia, F., Kermani, B.G., Zhao, C., Che, D., Dickinson, T., Wickham, E., Bierle, J. et al. (2004). Decoding randomly ordered DNA arrays. *Genome Res* 14, 870–877.

Gunther, E.C., Stone, D.J., Gerwien, R.W., Bento, P. and Heyes, M.P. (2003). Prediction of clinical drug efficacy by classification of drug-induced genomic expression profiles in vitro. *Proc Natl Acad Sci USA* 100, 9608–9613.

Guo, Z., Guilfoyle, R.A., Thiel, A.J., Wang, R. and Smith, L.M. (1994). Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. *Nucleic Acids Res* 22, 5456–5465.

Habel, L.A., Shak, S., Jacobs, M.K., Capra, A., Alexander, C., Pho, M., Baker, J., Walker, M., Watson, D., Hackett, J. et al. (2006). A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. *Breast Cancer Res* 8, R25.

Hans, C.P., Weisenburger, D.D., Greiner, T.C., Gascoyne, R.D., Delabie, J., Ott, G., Muller-Hermelink, H.K., Campo, E., Braziel, R.M., Jaffe, E.S. et al. (2004). Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood* 103, 275–282.

Hardiman, G. (2004). Microarray platforms–comparisons and contrasts. *Pharmacogenomics* 5, 487–502.

Harris, M.A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C. et al. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 32, D258–D261.

Harris, L., Fritsche, H., Mennel, R., Norton, L., Ravdin, P., Taube, S., Somerfield, M.R., Hayes, D.F., Bast, R.C., Jr. (2007). American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol*. 25(33), 5287–5312. (Epub 2007 Oct 22).

Heller, R.A., Schena, M., Chai, A., Shalon, D., Bedilion, T., Gilmore, J., Woolley, D.E. and Davis, R.W. (1997). Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proc Natl Acad Sci USA* 94, 2150–2155.

Hieronymus, H., Lamb, J., Ross, K.N., Peng, X.P., Clement, C., Rodina, A., Nieto, M., Du, J., Stegmaier, K., Raj, S.M. et al. (2006). Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. *Cancer Cell* 10, 321–330.

Hoheisel, J.D. (2006). Microarray technology: Beyond transcript profiling and genotype analysis. *Nat Rev Genet* 7, 200–210.

Holloway, A.J., van Laar, R.K., Tothill, R.W. and Bowtell, D.D. (2002). Options available–from start to finish–for obtaining data from DNA microarrays II. *Nat Genet* 32(Suppl), 481–489.

Huang, J., Qi, R., Quackenbush, J., Dauway, E., Lazaridis, E. and Yeatman, T. (2001). Effects of ischemia on gene expression. *J Surg Res* 99, 222–227.

Hughes, T.R., Mao, M., Jones, A.R., Burchard, J., Marton, M.J., Shannon, K.W., Lefkowitz, S.M., Ziman, M., Schelter, J.M., Meyer, M.R. et al. (2001). Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol* 19, 342–347.

Hughes, T.R., Marton, M.J., Jones, A.R., Roberts, C.J., Stoughton, R., Armour, C.D., Bennett, H.A., Coffey, E., Dai, H., He, Y.D. et al.

(2000). Functional discovery via a compendium of expression profiles. *Cell* 102, 109–126.

Irizarry, R.A., Warren, D., Spencer, F., Kim, I.F., Biswal, S., Frank, B.C., Gabrielson, E., Garcia, J.G., Geoghegan, J., Germino, G. et al. (2005). Multiple-laboratory comparison of microarray platforms. *Nat Methods* 2, 345–350.

Iscove, N.N., Barbara, M., Gu, M., Gibson, M., Modi, C. and Winegarden, N. (2002). Representation is faithfully preserved in global cDNA amplified exponentially from sub-picogram quantities of mRNA. *Nat Biotechnol* 20, 940–943.

Ishii, M., Hashimoto, S., Tsutsumi, S., Wada, Y., Matsushima, K., Kodama, T. and Aburatani, H. (2000). Direct comparison of GeneChip and SAGE on the quantitative accuracy in transcript profiling analysis. *Genomics* 68, 136–143.

King, H.C. and Sinha, A.A. (2001). Gene expression profile analysis by DNA microarrays: Promise and pitfalls. *JAMA* 286, 2280–2288.

Kittleson, M.M., Ye, S.Q., Irizarry, R.A., Minhas, K.M., Edness, G., Conte, J.V., Parmigiani, G., Miller, L.W., Chen, Y., Hall, J.L. et al. (2004). Identification of a gene expression profile that differentiates between ischemic and nonischemic cardiomyopathy. *Circulation* 110, 3444–3451.

Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.P., Subramanian, A., Ross, K.N. et al. (2006). The Connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313, 1929–1935.

Larkin, J.E., Frank, B.C., Gavras, H., Sultana, R. and Quackenbush, J. (2005). Independence and reproducibility across microarray platforms. *Nat Methods* 2, 337–344.

Liang, P. and Pardee, A.B. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 257, 967–971.

Liang, P. and Pardee, A.B. (2003). Analysing differential gene expression in cancer. *Nat Rev Cancer* 3, 869–876.

Lin, D.W., Coleman, I.M., Hawley, S., Dumpit, R., Gifford, D., Kezele, P., Hung, H., Knudsen, B.S., Kristal, A.R. and Nelson, P.S. (2006). Influence of surgical manipulation on prostate gene expression: Implications for molecular correlates of treatment effects and disease prognosis. *J Clin Oncol* 24, 3763–3770.

Lipshutz, R.J., Fodor, S.P., Gingeras, T.R. and Lockhart, D.J. (1999). High density synthetic oligonucleotide arrays. *Nat Genet* 21, 20–24.

Liu, R.H., Nguyen, T., Schwarzkopf, K., Fuji, H.S., Petrova, A., Siuda, T., Peyvan, K., Bizak, M., Danley, D. and McShea, A. (2006). Fully integrated miniature device for automated gene expression DNA microarray processing. *Anal Chem* 78, 1980–1986.

Lock, C., Hermans, G., Pedotti, R., Brendolan, A., Schadt, E., Garren, H., Langer-Gould, A., Strober, S., Cannella, B., Allard, J. et al. (2002). Gene-microarray analysis of multiple sclerosis lesions yields new targets validated in autoimmune encephalomyelitis. *Nat Med* 8, 500–508.

Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H. et al. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 14, 1675–1680.

Lossos, I.S., Czerwinski, D.K., Alizadeh, A.A., Wechser, M.A., Tibshirani, R., Botstein, D. and Levy, R. (2004). Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. *N Engl J Med* 350, 1828–1837.

Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A. et al. (2005). MicroRNA expression profiles classify human cancers. *Nature* 435, 834–838.

Ly, D.H., Lockhart, D.J., Lerner, R.A. and Schultz, P.G. (2000). Mitotic misregulation and human aging. *Science* 287, 2486–2492.

Ma, X.J., Salunga, R., Tuggle, J.T., Gaudet, J., Enright, E., McQuary, P., Payette, T., Pistone, M., Stecker, K., Zhang, B.M. et al. (2003). Gene expression profiles of human breast cancer progression. *Proc Natl Acad Sci USA* 100, 5974–5979.

Marshall, E. (2004). Getting the noise out of gene arrays. *Science* 306, 630–631.

Masuda, N., Ohnishi, T., Kawamoto, S., Monden, M. and Okubo, K. (1999). Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples. *Nucleic Acids Res* 27, 4436–4443.

Michiels, S., Koscielny, S. and Hill, C. (2005). Prediction of cancer outcome with microarrays: A multiple random validation strategy. *Lancet* 365, 488–492.

Mirnics, K., Middleton, F.A., Marquez, A., Lewis, D.A. and Levitt, P. (2000). Molecular characterization of schizophrenia viewed by microarray analysis of gene expression in prefrontal cortex. *Neuron* 28, 53–67.

Moffat, J., Grueneberg, D.A., Yang, X., Kim, S.Y., Kloepfer, A.M., Hinkle, G., Piqani, B., Eisenhaure, T.M., Luo, B., Grenier, J.K. et al. (2006). A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* 124, 1283–1298.

Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrale, M., Laurila, E. et al. (2003). PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 34, 267–273.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq.Nat Methods. 5(7), 621–628.

Novak, K. (2006). Where the chips fall. *Nat Med* 12, 158–159.

Ntzani, E.E. and Ioannidis, J.P. (2003). Predictive ability of DNA microarrays for cancer outcomes and correlates: An empirical assessment. *Lancet* 362, 1439–1444.

Paik, S. (2007). Development and clinical utility of a 21-gene recurrence score prognostic assay in patients with early breast cancer treated with tamoxifen. *Oncologist* 12, 631–635.

Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F.L., Walker, M.G., Watson, D., Park, T. et al. (2004). A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 351, 2817–2826.

Paik, S., Tang, G., Shak, S., Kim, C., Baker, J., Kim, W., Cronin, M., Baehner, F.L., Watson, D., Bryant, J. et al. (2006). Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol* 24, 3726–3734.

Pease, A.C., Solas, D., Sullivan, E.J., Cronin, M.T., Holmes, C.P. and Fodor, S.P. (1994). Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc Natl Acad Sci USA* 91, 5022–5026.

Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A. et al. (2000). Molecular portraits of human breast tumours. *Nature* 406, 747–752.

Pilpel, Y., Sudarsanam, P. and Church, G.M. (2001). Identifying regulatory networks by combinatorial analysis of promoter elements. *Nat Genet* 29, 153–159.

Potti, A., Dressman, H.K., Bild, A., Riedel, R.F., Chan, G., Sayer, R., Cragun, J., Cottrill, H., Kelley, M.J., Petersen, R. et al. (2006). Genomic signatures to guide the use of chemotherapeutics. *Nat Med* 12, 1294–1300.

Quackenbush, J. (2001). Computational analysis of microarray data. *Nat Rev Genet* 2, 418–427.

Ramaswamy, S. and Golub, T.R. (2002). DNA microarrays in clinical oncology. *J Clin Oncol* 20, 1932–1941.

Ramaswamy, S., Tamayo, P., Rifkin, R., Mukherjee, S., Yeang, C.H., Angelo, M., Ladd, C., Reich, M., Latulippe, E., Mesirov, J.P. et al. (2001). Multiclass cancer diagnosis using tumor gene expression signatures. *Proc Natl Acad Sci USA* 98, 15149–15154.

Rhodes, D.R., Yu, J., Shanker, K., Deshpande, N., Varambally, R., Ghosh, D., Barrette, T., Pandey, A. and Chinnaiyan, A.M. (2004). ONCOMINE: A cancer microarray database and integrated data-mining platform. *Neoplasia* 6, 1–6.

Ross, D.T., Scherf, U., Eisen, M.B., Perou, C.M., Rees, C., Spellman, P., Iyer, V., Jeffrey, S.S., Van de Rijn, M., Waltham, M. et al. (2000). Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet* 24, 227–235.

Sargent, T.D. (1987). Isolation of differentially expressed genes. *Methods Enzymol* 152, 423–432.

Schena, M., Heller, R.A., Theriault, T.P., Konrad, K., Lachenmeier, E. and Davis, R.W. (1998). Microarrays: Biotechnology's discovery platform for functional genomics. *Trends Biotechnol* 16, 301–306.

Schena, M., Shalon, D., Davis, R.W. and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467–470.

Scherf, U., Ross, D.T., Waltham, M., Smith, L.H., Lee, J.K., Tanabe, L., Kohn, K.W., Reinhold, W.C., Myers, T.G., Andrews, D.T. et al. (2000). A gene expression database for the molecular pharmacology of cancer. *Nat Genet* 24, 236–244.

Schulze, A. and Downward, J. (2001). Navigating gene expression using microarrays – A technology review. *Nat Cell Biol* 3, E190–E195.

Segal, E., Friedman, N., Koller, D. and Regev, A. (2004). A module map showing conditional activity of expression modules in cancer. *Nat Genet* 36, 1090–1098.

Sgroi, D.C., Teng, S., Robinson, G., LeVangie, R., Hudson, J.R., Jr and Elkahloun, A.G. (1999). In vivo gene expression profile analysis of human breast cancer progression. *Cancer Res* 59, 5656–5661.

Shi, L., Reid, L.H., Jones, W.D., Shippy, R., Warrington, J.A., Baker, S.C., Collins, P.J., de Longueville, F., Kawasaki, E.S., Lee, K.Y. et al. (2006). The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* 24, 1151–1161.

Simon, R., Radmacher, M.D., Dobbin, K. and McShane, L.M. (2003). Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *J Natl Cancer Inst* 95, 14–18.

Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S. et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 98, 10869–10874.

Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J.S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S. et al. (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci USA* 100, 8418–8423.

Sotiriou, C. and Piccart, M.J. (2007). Taking gene-expression profiling to the clinic: When will molecular signatures become relevant to patient care?. *Nat Rev Cancer* 7, 545–553.

Southern, E., Mir, K. and Shchepinov, M. (1999). Molecular interactions on microarrays. *Nat Genet* 21, 5–9.

Stegmaier, K., Corsello, S.M., Ross, K.N., Wong, J.S., Deangelo, D.J. and Golub, T.R. (2005). Gefitinib induces myeloid differentiation of acute myeloid leukemia. *Blood* 106, 2841–2848.

Stegmaier, K., Ross, K.N., Colavito, S.A., O'Malley, S., Stockwell, B.R. and Golub, T.R. (2004). Gene expression-based high-throughput screening (GE-HTS) and application to leukemia differentiation. *Nat Genet* 36, 257–263.

Steinmetz, L.M. and Davis, R.W. (2004). Maximizing the potential of functional genomics. *Nat Rev Genet* 5, 190–201.

Stoeckert, C.J., Causton, H.C. and Ball, C.A. (2002). Microarray databases: Standards and ontologies. *Nat Genet* 32(Suppl), 469–473.

Stoughton, R.B. and Friend, S.H. (2005). How molecular profiling could revolutionize drug discovery. *Nat Rev Drug Discov* 4, 345–350.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102, 15545–15550.

Tan, P.K., Downey, T.J., Spitznagel, E.L., Jr., Xu, P., Fu, D., Dimitrov, D.S., Lempicki, R.A., Raaka, B.M. and Cam, M.C. (2003). Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res* 31, 5676–5684.

Taniguchi, M., Miura, K., Iwao, H. and Yamanaka, S. (2001). Quantitative assessment of DNA microarrays–comparison with Northern blot analyses. *Genomics* 71, 34–39.

Tomlins, S.A., Rhodes, D.R., Perner, S., Dhanasekaran, S.M., Mehra, R., Sun, X.W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R. et al. (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* 310, 644–648.

van't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T. et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415, 530–536.

van de Vijver, M.J., He, Y.D., van't Veer, L.J., Dai, H., Hart, A.A., Voskuil, D.W., Schreiber, G.J., Peterse, J.L., Roberts, C., Marton, M.J. et al. (2002). A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347, 1999–2009.

Van Gelder, R.N., von Zastrow, M.E., Yool, A., Dement, W.C., Barchas, J.D. and Eberwine, J.H. (1990). Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc Natl Acad Sci USA* 87, 1663–1667.

Velculescu, V.E., Zhang, L., Vogelstein, B. and Kinzler, K.W. (1995). Serial analysis of gene expression. *Science* 270, 484–487.

Wei, G., Twomey, D., Lamb, J., Schlis, K., Agarwal, J., Stam, R.W., Opferman, J.T., Sallan, S.E., den Boer, M.L., Pieters, R. et al. (2006). Gene expression-based chemical genomics identifies rapamycin as a modulator of MCL1 and glucocorticoid resistance. *Cancer Cell* 10, 331–342.

Zimmermann, C.R., Orr, W.C., Leclerc, R.F., Barnard, E.C. and Timberlake, W.E. (1980). Molecular cloning and selection of genes regulated in Aspergillus development. *Cell* 21, 709–715.

RECOMMENDED RESOURCES

Resources for expression profiling experiments

Web site	URL	Resources
National Human Genome Research Institute	http://research.nhgri.nih.gov/microarray/main.html	Protocols, web links
Broad Institute, Cancer Genomics Group	http://www.broad.mit.edu/cancer/	Gene expression analysis software, web links
Stanford Genomics	http://genome-www.stanford.edu/	Software, web links, cDNA microarray protocols
Jackson Laboratory, Statistical Genomics Group	http://www.jax.org/staff/Churchill/labsite/	Advice about the design of microarray experiments
Microarray Gene Expression Data Society (MGED)	http://www.mged.org/Workgroups/MIAME/miame_checklist.html	Minimal requirements for the publication of microarray data
Gene Expression Omnibus database	http://www.ncbi.nlm.nih.gov/geo/	Repository of gene expression data
Array Express database	http://www.ebi.ac.uk/arrayexpress/	Repository of gene expression data
Affymetrix	http://www.affymetrix.com	Commercial website with useful resources

This listing is adapted from Ebert and Golub (2004).

CHAPTER

Proteomics: The Deciphering of the Functional Genome

Li-Rong Yu, Nicolas A. Stewart and Timothy D. Veenstra

INTRODUCTION

The continuing development of novel analytical technologies and instrumentation has brought a dramatic new dimension to how biological research can be pursued. This new dimension has enabled discovery-driven research to be conducted at par with hypothesis-driven research. In a hypothesis-driven approach, the understanding of a particular biological entity – whether it be a cell, gene, or protein – is used to develop studies that are designed to answer a specific question about a single gene, transcript, or protein, for example. Hypothesis-driven studies utilize technologies that focus on one specific entity per experiment. Discovery-driven studies, however, are designed around broad questions and focus on global characteristics of a cell or organism. These studies utilize advanced technologies that are able to gather information on thousands of different genes, transcripts, proteins, and metabolites in a relatively rapid and comprehensive fashion.

While these globally directed studies have seemingly been thrust upon the scientific community overnight, in reality it has taken decades of technological and instrumental development to bring this field of science to fruition. The first giant step in discovery-driven science is the completion of the Human Genome Project. This monumental effort could only have been accomplished with the many developments in cloning, amplification, and high-throughput gene sequencing (Yager et al., 1991). The analysis of thousands of gene transcripts using mRNA arrays has been made possible by the ability to synthesize mRNA probes and coupling them to solid surfaces (Gilham, 1970). Shortly after the development of these genome – and transcriptome-scanning technologies, many scientists shifted their focus onto proteins. Proteomics – the analysis of the entire protein complement of a cell, tissue, or organism under a specific, defined set of conditions – in its present state has also been dependent on decades of technological and instrumental developments. These developments have included advances in mass spectrometry (MS) technology, protein fractionation techniques, and bioinformatics, to name a few.

One of the difficult challenges in proteomics is the uncertainty of its complexity. Let us assume that as many as 15,000 genes are expressed within a given human cell at any single time. Let us then assume that these genes give rise, on average, to four different mRNA transcripts each, or 60,000 different transcripts in total. This estimate is very conservative given the presence of splice variants, single nucleotide polymorphisms, etc. that can arise from expressed genes. Now consider all of the events that can happen to a single transcript during and after translation into a mature protein. The protein can be processed from a pre-protein into a mature protein. The most common event that happens to change a protein is post-translational modification. There are over 300 known modifications, however; some of the more common are (in no particular order) phosphorylation, glycosylation, methylation, and acetylation. A single protein that may have four potential phosphorylation sites will, in itself, give rise to 16 different isoforms ranging from all the sites phosphorylated to none of them being phosphorylated. If each of the 60,000

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

possible protein products contained only four possible sites of modification, these would give rise to as many as 960,000 different proteins. Obviously, these numbers are extremely conservative and suggest that the human proteome is likely to contain well over 1,000,000 distinct species. This number is well beyond the current capacity of state-of-the-art technologies, resulting in a consistent undersampling of the proteome regardless of the analysis method chosen. It is this uncertainty in the overall measurement that results in the question that hovers around many global proteomic studies: "Was the protein absent in the sample or did we simply fail to detect it?"

GEL-BASED AND SOLUTION-BASED PROTEOMICS

Proteomics relies on three basic technological cornerstones: a method to fractionate complex protein or peptide mixtures, MS to acquire the data necessary to identify individual proteins, and bioinformatics to analyze and assemble the MS data. While the MS and bioinformatic components are somewhat similar in most applications, there are two very distinct methods to separate out complex protein samples in proteomics. The mere mention of the word "proteomics" conjures up an image of stained protein spots that have been fractionated by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE). This separation technology, which was developed over 30 years ago, is still the most common separation technique used in proteomics today (O'Farrell, 1975). Advances have been made in this technology that enables thousands of proteins to be resolved in a single gel.

While the primary role of 2D-PAGE is to separate complex mixtures of proteins, it also allows a means to compare the relative abundances of proteins from different proteomes. In a typical 2D-PAGE-based study, the proteomes from two distinct cell populations (e.g., control versus treated cells) are extracted and fractionated on separate gels (Figure 8.1) (Van den Bergh and Arckens, 2005). Since most proteins are colorless, the separated proteins need to be visualized (Westermeier and Marouga, 2005). This visualization is routinely accomplished using colorimetric stains, such as coomassie blue or silver stain. The resulting images are typically quite complex; therefore the spots observed on both gels are aligned so that the relative staining intensity of individual proteins can be compared between gels. Protein spots that are stained more intensely on one gel compared to the other are excised from the gel. The protein within the gel is then enzymatically digested by introducing trypsin into the gel piece. The resultant tryptic peptides are extracted from the gel and analyzed by MS or tandem MS (MS/MS) to acquire the raw data necessary to identify the protein(s) within the gel spot. This identification is accomplished using software designed to compare this type of data against databases containing large amounts of genomic or protein sequence information.

While 2D-PAGE fractionation has been the subject of many different criticisms over the years, it still remains a cornerstone technology in proteomics (Van den Bergh and Arckens, 2005). Many of the criticisms, such as its inability to resolve membrane proteins and gel-to-gel irreproducibility have been addressed through the development of better reagents, equipment, and gel alignment software. For instance, over the past several of years, 2D-DIGE reagents have been developed that allow comparative proteomes to be labeled with different fluorophores (i.e., Cy2, Cy3, and Cy5) (Hoorn et al., 2006). This differential labeling allows for the relative abundance of proteins from different proteomes to be compared by their combination prior to their run on a single gel, obviating the separation irreproducibility that can be observed between gels.

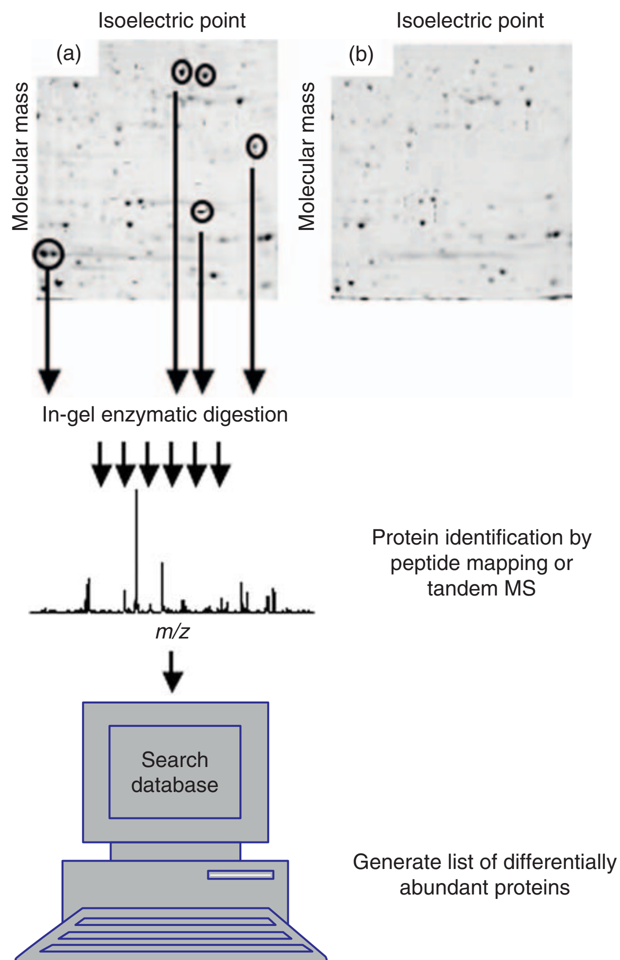


Figure 8.1 Quantitative proteomics using two-dimensional polyacrylamide gel electrophoresis (2D-PAGE). In this method, comparative proteome samples are separated on distinct 2D-PAGE gels. After staining, protein spots that are more abundant on one gel compared to the other are excised from the gel. The protein(s) within the gel is then subjected to in-gel tryptic digestion and the resultant peptides are extracted and analyzed by mass spectrometry (MS). The MS data are then searched against the appropriate database to identify the protein(s) with the gel spot.

No single chromatographic or electrophoretic procedure, including 2D-PAGE, can completely resolve a mixture as complex as that of a proteome. As a complement to 2D-PAGE methods,

solution-based chromatographic techniques have been developed that permit direct on-line MS analysis of proteomes (Issaq et al., 2005). Analogous to 2D-PAGE, these solution-based methods typically use at least two different fractionation schemes to separate peptides/proteins prior to their entering the mass spectrometer. While many different combinations have been proposed, the most effective and popular has been the use of strong cation exchange (SCX) followed by reversed-phase high-performance liquid chromatography (RPLC). John Yates III first demonstrated the effectiveness of combining this multidimensional separation with MS (Washburn et al., 2001). This method, termed "MudPIT" (multidimensional protein identification technology), has shown the capability of identifying thousands of proteins per proteome study. In this scheme, the proteins extracted from a cell lysate are first digested into peptides. There are two major advantages in working with peptides instead of proteins. Peptides are more soluble and easier to separate than intact proteins, especially hydrophobic and membrane proteins, and peptides are more amenable to identification by MS than are proteins (as described in more detail later). The challenge with this strategy is that the number of species that have to be resolved has dramatically increased. The digestate is initially separated by SCX and fractions collected off this column are then loaded onto a RP column that is coupled directly on-line with the mass spectrometer. The mass spectrometer can then dynamically select and acquire the data necessary to identify peptides as they elute from the RP column.

A limitation of MudPIT when compared to 2D-PAGE is in its inability to directly measure the relative abundance of proteins in different proteome samples. To meet this need, many different approaches have been developed, all of which use the data provided by the mass spectrometer for quantitation (Ong and Mann, 2005). Many of the methods rely on the use of differential stable-isotope labeling (Julka and Regnier, 2005). Analogous to DIGE, which uses differently fluorescent molecules with identical electrophoretic properties, stable-isotope labeling of proteomes uses chemically identical reagents that differ in their isotope content (i.e., mass). For example, the most popular labeling method termed isotope-coded affinity tags (ICAT) utilizes thiol-modifying reagents that differ in their carbon-13 (^{13}C) content (Gygi et al., 1999). In this procedure, one proteome sample is labeled with the "light" ICAT reagent while a comparative sample is labeled with the "heavy" ICAT reagent that contains nine ^{13}C atoms in place of nine ^{12}C atoms found in the "light" version (Figure 8.2). The label is attached to the proteins after they have been separately extracted from the samples to be compared. Once the two samples are differentially modified, they can be combined. The proteins are digested into peptides and the modified peptides are extracted by virtue of a biotin tag that is present at the termini of the ICAT reagents. This enriched peptide mixture is finally analyzed using MudPIT and the relative abundance and identity of the peptides are determined from measurements provided by the mass spectrometer. While ICAT illustrates a post-extraction/chemical modification method, stable-isotope labeling can also be performed at the metabolic level by culturing cells in growth media that is highly enriched

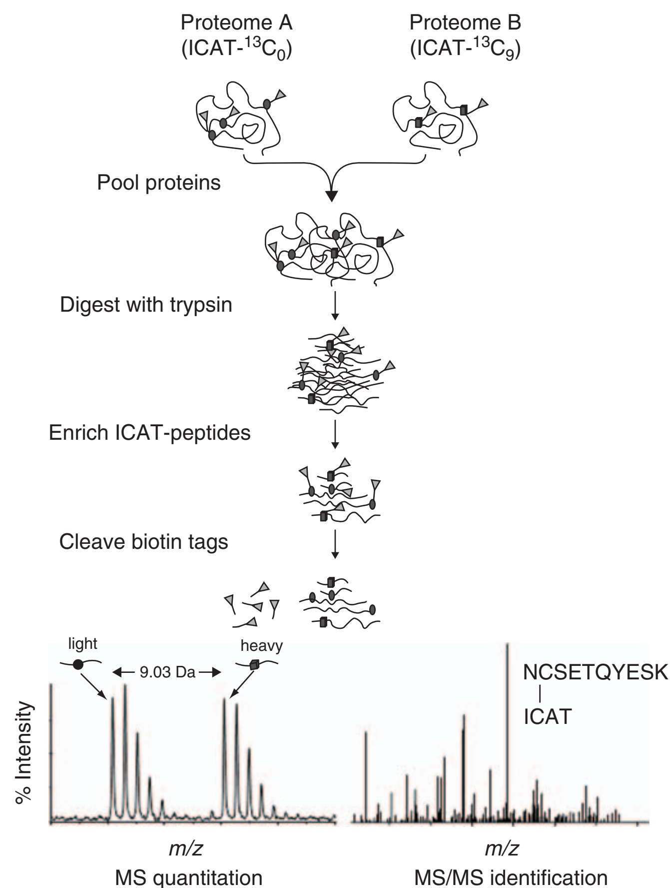


Figure 8.2 Quantitative proteomics using isotope-coded affinity tags (ICAT). In this method, comparative proteome samples are labeled with chemically identical reagents that differ only by their carbon isotope content (i.e., $^{12}C_9$ for the light reagent and $^{13}C_9$ for the heavy reagent). After the proteins are modified, the proteome samples are combined and digested into tryptic peptides. The ICAT-modified peptides are extracted using avidin chromatography by virtue of the biotin moiety on the terminus of the ICAT reagents. After removal of the biotin portion, these ICAT-peptides are analyzed by reversed-phase liquid chromatography coupled directly on-line with a mass spectrometer. The mass spectrometer is operated in a data-dependent tandem mass spectrometry (MS/MS) mode, enabling the relative quantitation of the peptide in the two samples to be measured in the MS mode as well as its identity be discerned from the data acquired by MS/MS.

in a particular heavy isotope (e.g., ^{15}N in place of ^{14}N) or a "heavy" amino acid (e.g., $^{13}C_6$-lysine) (Ong et al., 2003). A list of various sample preparation used in proteomics today, along with their advantages and disadvantages is provided in Table 8.1.

MASS SPECTROMETRY

While high-resolution separations are critical for complex protein mixture analysis, MS is inarguably the technology that

TABLE 8.1 List of advantages and disadvantages of popular sample preparation methods used in proteomics

Method	Advantages	Disadvantages
2D-PAGE	• Staining provides relative quantitation • Only sequence proteins of interest	• Laborious • Limited sensitivity • Irreproducibility a factor
2D-DIGE	• Can compare multiple proteomes on single gel • Reproducible separation	• Laborious • Requires fluorescent scanner
MudPIT	• Highly sensitive • Broad proteome coverage • High throughput	• Limited individual protein coverage • Unable to target specific proteins Lacks inherent ability for quantitative comparison
ICAT	• Reduces complexity prior to MS analysis	• Low throughput • Limited to a binary comparison • Unable to target specific proteins
iTRAQ	• Allows 4 samples to be simultaneously compared	• Low throughput • Unable to target specific proteins

has driven the proteomics revolution. Although, it has been around for approximately a century, MS has gained increased prominence with the development of techniques to rapidly identify proteins. There are many different reasons that have propelled the MS to its prominent position within the field of proteomics. Among these are the sensitivity afforded using MS, allowing proteins and peptides present in the low femtomole (fmol, 10^{-15} mol) to high attomole (amol, 10^{-18} mol) range to be successfully identified. The mass measurement accuracy available using current MS technology, which can routinely be in the range of 1–5 ppm, also increases the confidence in the identification provided by the bioinformatic search of the raw data.

While these and other technology specifications have been important, the development of MS/MS (McLafferty, 1981) and the coupling of on-line protein/peptide separations with MS (Henion, 1978) have been the key determinants that have enabled proteomics. Tandem MS allows tryptic peptides obtained from an enzymatic digestion of a complex proteome mixture, to be fragmented in such a way that sufficient sequence information can be obtained for its unambiguously identification. While proteins are typically identified through MS/MS identification of their peptide surrogates, methods to characterize intact proteins by MS/MS are becoming more and more popular. The coupling of on-line separations, most commonly RPLC, has enabled complex mixtures to be fractionation prior to MS analysis so that thousands of peptides can be identified within a few hours, as shown in Figure 8.3 (Washburn et al., 2001). The throughput by which proteins in complex mixtures are identified by LC-MS is unparalleled using any other technology: a critical parameter for discovery science in which large amounts of data are collected in order to find the important factors.

There are many different types of commercially available mass spectrometers, and it can be confusing as to which type is most suitable for a specific application. For instance, ion-trap mass spectrometers equipped with electrospray ionization (ESI) sources are the "work-horse" instruments for the characterization of complex proteomes when solution-based separations are used. Time of flight (TOF) instruments are popular for identifying proteins separated using 2D-PAGE, primarily because of their speed and mass accuracy. The choice of what instrument to be utilized needs to be weighed against the type of research that is being focused on.

How is MS able to identify so many peptides in a complex mixture? A schematic of the process is shown in Figure 8.3. The mass spectrometer is operated in what is referred to as a "data-dependent" mode of operation. In this mode, the instrument takes a series of "snapshots" of peptides eluting from a LC column into the mass analyzer. This scan is referred to as the MS scan. In between these snapshots, a series of MS/MS scans of peptides that are isolated within the instrument for CID are acquired. The instrument selects peptides for MS/MS analysis based on their intensity in the preceding MS scan. A data-dependent experiment can perform 10 MS/MS scans between every MS scan. The mass spectrometer continues this cycle of MS scan followed by ten MS/MS scans throughout the entire LC/MS/MS experiment. Depending on the length of the LC separation, upwards of 7000 MS/MS sequencing attempts are made on peptides eluting from the column into the mass spectrometer. Bioinformatic analysis will result in 10–25% (i.e., 700–1750) of these sequencing attempts resulting in successful peptide identifications. A list of popular mass analyzers used in proteomics and their performance characteristics is given in Table 8.2.

BIOINFORMATICS

Probably no area of proteomics evolves faster than bioinformatics. Proteomics is critically dependent on bioinformatics to process the raw mass spectral data into protein data. While routinely used by every proteomic laboratory, the most critical software programs are those that take peptide mapping and/or tandem MS results and determine the protein or peptide sequence that most closely matches the experimental data. The two most popular software packages for matching experimental MS data to peptide/protein sequences are MASCOT

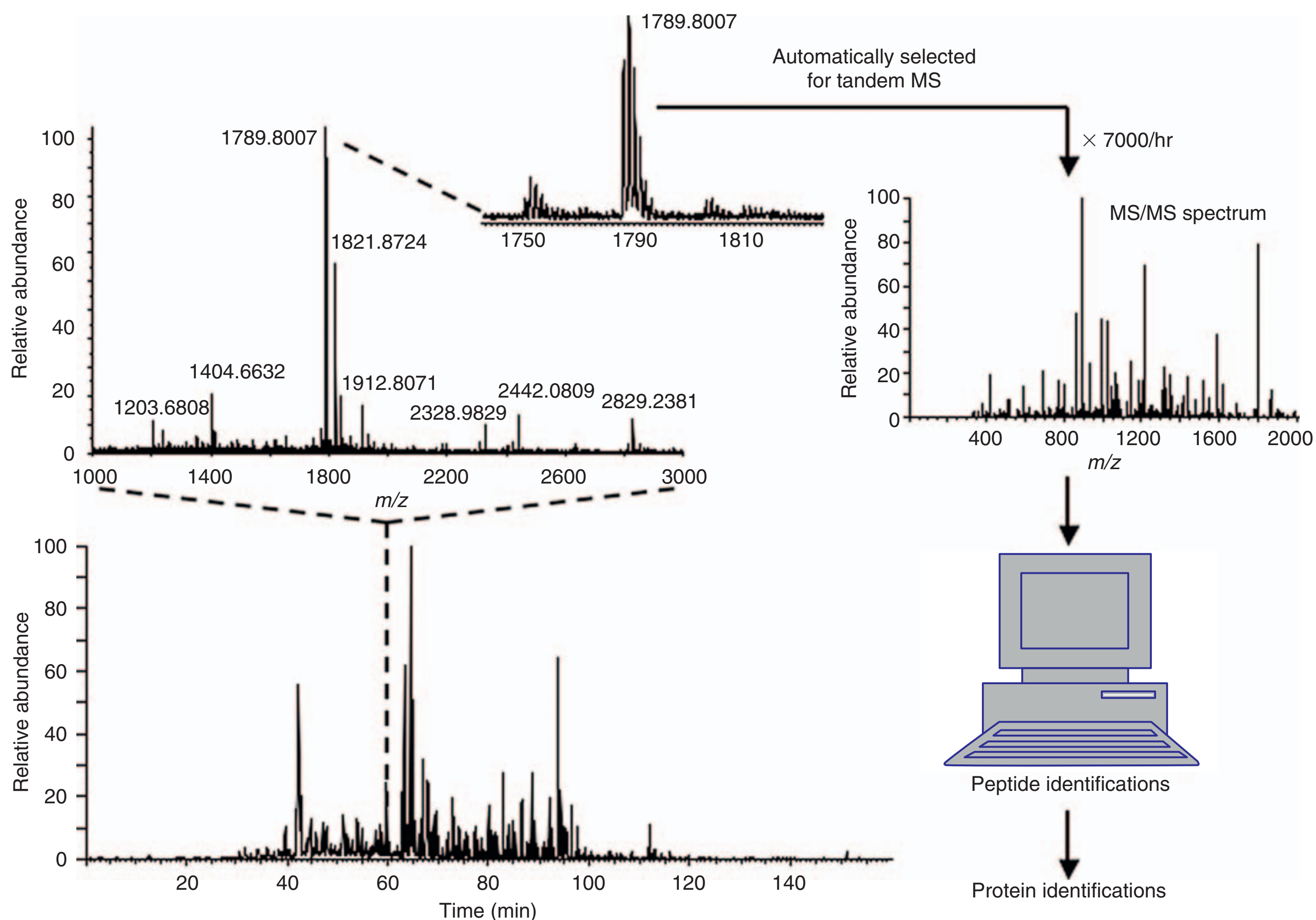


Figure 8.3 Principles of data-dependent tandem mass spectrometry (MS/MS). In data-dependent MS/MS, the mass spectrometer conducts an MS scan to record peptide masses that can be detected during a reversed-phase liquid chromatography (RPLC)-MS analysis of a complex proteome sample. The mass spectrometer isolates the peptide producing the most intense signal (i.e., *m/z* 1789.8007) and subjects it to collisional induced dissociation (CID). The resultant fragment ions are detected by the mass spectrometer and are subsequently used to identify the peptide in a bioinformatic analysis of the data. The mass spectrometer then isolates the next most intense peptide and collects its MS/MS spectrum. Linear ion-trap mass spectrometers may sequentially select on the order of 10 peptides for CID prior to recording the next MS scan. In the following MS scan, the instrument looks for new peptides that may have eluted from the LC column and selects these for CID and possible identification.

Table 8.2 Performance characteristics of various mass analyzers commonly used in proteomic research

Analyzer	Sensitivity	Resolution	Mass accuracy
Ion-trap	Good	Low	Low
Linear ion-trap	Excellent	Low	Low
Triple-quadrupole	Good	Good	Good
TOF	Good	High	High
FTICR	Excellent	High	High

(Lubec et al., 2005) and SEQUEST (Yates et al., 1995). As with most software algorithms, they both do what can also be done manually but at a much faster speed. For instance, SEQUEST is capable of transforming thousands of raw tandem MS files into potential peptide identifications within minutes, whereas manually assigning a single raw file would take on the order of hours. These software algorithms are also capable of identifying modified peptides when the mass added to the peptide as a result of some covalent modification (e.g., phosphorylation) is known. While not in widespread distribution, software packages, such as Scoring ALgorithm for Spectral Analysis (SALSA) developed by Daniel Leibler, provides a means of identifying unanticipated modifications, by identifying spectra that display characteristic

product ions, neutral or charged losses of groups of signals that are indicative of a particular modification (Hansen et al., 2001). Taken together, these varieties of software packages have been critical to the enabling of data analysis on the scale needed for discovery-driven proteomic studies.

Throughout this chapter, the theme has been to use proteomic technologies to measure the relative abundances of proteins in different samples. Software has been developed to do this both at the gel-staining level and the mass spectral data level (i.e., when stable-isotope labeling quantitation is employed). Where the major lack has been is how this relative abundance data is presented. It is still typically displayed as lists of proteins with their relative abundance between two samples being given as a fold-difference. Unfortunately, with the ability to generate abundance data for thousands of proteins, these lists become an "unsolvable puzzle" when trying to identify major effects on a cell's proteome as a result of a specific treatment. Recently, software packages such as Ingenuity™ have been developed to distill abundance data into annotated protein pathways enabling the probability that any specific signaling pathway has been affected can be statistically measured (www.ingenuity.com/products/pathways_analysis.html). Software packages such as these also allow transcriptomic and proteomic data to be directly compared at both the single biomolecule and pathway levels.

IMPACT OF PROTEOMICS ON UNDERSTANDING DISEASES

It is impossible to find a common characteristic that threads every disease together. One prevalent characteristic, primarily in the area of cancer, is that the earlier it is detected the better the chance of survival. To detect cancer at earlier stages is going to require novel biomarkers that possess high sensitivity and specificity. This need has driven much of the proteomic research over the past 3–5 years. Much of the search for novel biomarkers falls within discovery-driven research, and therefore MS has played a major role. The aim, in most studies, is to leverage the ability of MS to identify proteins in biological samples from patients with specific cancers and compare that to lists of proteins identified in the same sample types obtained from healthy-matched controls. Proteins that are found to be more prevalent in samples from disease-afflicted patients are considered as potential biomarkers.

While on the surface the premise sounds simple enough, in reality there are several daunting challenges, both technological and physiological. Even with the speed at which MS is capable of identifying proteins, it still undersamples any complex proteome mixture. When such entities as post-translational modifications and splice variants are taken into account, it is impossible to know exactly the number of distinct proteins within a cell or biofluid proteome. Therefore, when a protein is detected in a diseased sample but not in a matched control, there is always that uncertainty as to whether the protein is not really present or it just was not detected in the present experiment. As is commonly quoted in science, "the absence of evidence is not the evidence of absence". On the physiological side, the fact that the proteome of any cell or biofluid is dynamic results in a constantly changing background that occurs both between different patients and within the same patient. For instance, the serum proteome of an individual is different before and after lunch. Compound this with the genetic and lifestyle variability that is observed between different patients and it is easy to grasp how difficult it is to confidently identify a disease-specific change in a protein in this shifting matrix.

Regardless, the promise of technologies such as MS that can scan proteomes at levels never before possible propels investigators to search for new diagnostic markers. While the success in discovering new markers has been limited, they do however exist. For instance, a biomarker for the bladder disorder interstitial cystitis (IC) was found using a combination of protein fractionation and MS techniques (Keay et al., 2004). Previous to this report, no useful diagnostic marker for this disorder existed and patients were typically diagnosed by the exclusion of other maladies. This study resulted in the discovery of a novel nine-residue sialoglycopeptide that was shown to be present only in the supernatant of cultured bladder epithelial cells obtained from patients with IC and not in healthy controls or patients with bacterial cystitis.

One of the more intriguing MS-based proteomic techniques to identify signatures of diseases in blood involves the detection of peptides that signify exo- and endo-protease activities (Villanueva et al., 2006). In this study, low molecular weight peptides were extracted from serum and analyzed using matrix-assisted laser desorption ionization (MALDI) TOF MS. The study showed that a subset of the serum peptides were able to provide a signature allowing patients with three types of solid tumors to be classified separately from controls without cancer. Sixty-one of the signature peptides were identified and shown to fall into several tight clusters (Villanueva et al., 2006). Most of the identified peptides were generated by cancer-type specific exopeptidase activities that enabled highly accurate classification of a set of prostate cancer samples used to validate the markers. While this method does not identify a novel diagnostic or therapeutic protein marker, it does use MS to reveal a cancer-specific activity found within serum of cancer-affected patients.

2009 UPDATE

As any useful field of science does, mass spectrometry (MS)-based proteomics has continued to evolve in capabilities and direction. While many of the fundamental aims of MS-based proteomics – such as identifying differentially abundant proteins between different systems – remain, the instrumental capabilities available to achieve these aims have increased in the recent past. In addition, many novel experimental approaches have been designed to better interrogate specific proteins and post-translational modifications.

The complexity of the proteome, even from a simple organism, is such that fractionation remains critical prior to MS analysis. Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) (O'Farrell, 1975) remains a cornerstone separation method with the use of 2D-DIGE (differential in-gel electrophoresis) becoming increasingly popular due to its comparative reproducibility, quantitative capability, and ease of use (Hoorn et al., 2006). Multidimensional protein identification technology (popularly known as MudPIT) continues to dominate the separation of complex proteomes in cases where investigators wish to avoid the use of gels (Washburn et al., 2001). This combination of strong cation exchange and reversed phase liquid chromatography (LC), combined with the continuing development of mass spectrometers with higher sensitivity and faster duty cycles, has enabled laboratories to routinely identify >5,000 proteins within complex proteomes.

The chief development in mass spectrometry technology in the recent past is the Orbitrap mass spectrometer with electron transfer dissociation (ETD) (McAlister et al., 2007). This instrument combines high sensitivity, mass accuracy, and resolution in an ion-trap configuration that uses ETD to fragment proteins and peptides. Conventional fragmentation methods, such as collisional induced dissociation (CID), are most effective when dealing with small peptides with molecular masses in the range of 500-3500. ETD has the ability to effectively fragment much larger peptides, thereby enabling top-down approaches in which peptide and intact proteins with molecular masses up to 20,000 can be effectively identified. Post-translation modifications are also preserved during fragmentation using ETD enabling phosphorylation sites, in particular, to be more effectively characterized.

Quantitative proteomics, in which global differences in protein abundances are measured, continues to be an important area in proteomics in both cell biology and biomarker discovery. While this area has been dominated by the use of stable-isotope labeling approaches (e.g. ICAT, iTRAQ, etc.), recent developments in label-free quantitation are changing the way quantitation is conducted (Du et al., 2008). Label-free methods rely on the measured intensity of a peptide ion and compare this to its intensity as recorded within other samples. Theoretically, quantitative proteomics methods enable cross comparison of a large number of samples, instead of the small number enabled by isotope-labeling approaches. In addition, label-free methods are higher throughput and require fewer sample manipulation steps. Some of the difficulties in routine application of these methods, however, have centered on correct matching of peptides from run-to-run due to variation in LC separations of the mixtures prior to MS analysis.

While MS-based proteomics has probably not yet had the impact on understanding human diseases that was anticipated, the progress has still been substantial. Many investigators are now using multiple-reaction monitoring (MRM) of specific peptides within very biofluid samples to quantitate the absolute abundance of proteins in clinical samples (Issaq and Veenstra, 2008). This method combined with development of MS technology continues to push proteomics closer to routine use within clinical laboratories, where its impact on public health will be substantial.

2009 UPDATE REFERENCES

Du, X., Callister, S.J., Manes, N.P., Adkins, J.N., Alexandridis, R.A., Zeng, X., Roh, J.H., Smith, W.E., Donohue, T.J., Kaplan, S., Smith, R.D. and Lipton, M.S. (2008). A computational strategy to analyze label-free temporal bottom-up proteomics data. *J Proteome Res* 7, 2595–2604.

Hoorn, E.J., Hoffert, J.D. and Knepper, M.A. (2006). The application of DIGE-based proteomics to renal physiology. *Nephron Physiol* 104, 61–72.

Issaq, H.J. and Veenstra, T.D. (2008). Would you prefer multiple reaction monitoring or antibodies with your biomarker validation? *Expert Rev Proteomics* 5, 761–763.

McAlister, G.C., Phanstiel, D., Good, D.M., Berggren, W.T. and Coon J.J. (2007). Implementation of electron-transfer dissociation on a hybrid linear ion trap-orbitrap mass spectrometer. *Anal Chem* 79, 3525–3534.

O'Farrell, P.H. (1975). High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 250, 4007–4021.

Washburn, M.P., Wolters, D. and Yates, J.R., III (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 19, 242–247.

ACKNOWLEDGEMENTS

This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract N01-CO-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organization imply endorsement by the United States Government.

REFERENCES

Gilham, P.T. (1970). RNA sequence analysis. *Annu Rev Biochem* 39, 227–250.

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H. and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 17, 994–999.

Hansen, B.T., Jones, J.A., Mason, D.E. and Liebler, D.C. (2001). SALSA: A pattern recognition algorithm to detect electrophile-adducted peptides by automated evaluation of CID spectra in LC-MS-MS analyses. *Anal Chem* 73, 1676–1683.

Henion, J.D. (1978). Drug analysis by continuously monitored liquid chromatography/mass spectrometry with a quadrupole mass spectrometer. *Anal Chem* 50, 1687–1693.

Hoorn, E.J., Hoffert, J.D. and Knepper, M.A. (2006). The application of DIGE-based proteomics to renal physiology. *Nephron Physiol* 104, 61–72.

Issaq, H.J., Chan, K.C., Janini, G.M., Conrads, T.P. and Veenstra, T.D. (2005). Multidimensional separation of peptides for effective proteomic analysis. *J Chromatogr B Analyt Technol Biomed Life Sci* 817, 35–47.

Julka, S. and Regnier, F.E. (2005). Recent advancements in differential proteomics based on stable isotope coding. *Brief Funct Genomic Proteomic* 4, 158–177.

Keay, S.K., Szekely, Z., Conrads, T.P., Veenstra, T.D., Barchi, J.J., Jr., Zhang, C.O., Koch, K.R. and Michejda, C.J. (2004). An antiproliferative factor from interstitial cystitis patients is a frizzled 8 protein-related sialoglycopeptide. *Proc Natl Acad Sci USA* 101, 11803–11808.

Lubec, G., Afjehi-Sadat, L., Yang, J.W. and John, J.P. (2005). Searching for hypothetical proteins: Theory and practice based upon original data and literature. *Prog Neurobiol* 77, 90–127.

McLafferty, F.W. (1981). Tandem mass spectrometry. *Science* 214, 280–287.

O'Farrell, P.H. (1975). High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 250, 4007–4021.

Ong, S.E. and Mann, M. (2005). Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol* 1, 252–262.

Ong, S.E., Foster, L.J. and Mann, M. (2003). Mass spectrometric-based approaches in quantitative proteomics. *Methods* 29, 124–130.

Van den Bergh, G. and Arckens, L. (2005). Recent advances in 2D electrophoresis: An array of possibilities. *Expet Rev Proteomics* 2, 243–252.

Villanueva, J., Shaffer, D.R., Philip, J., Chaparro, C.A., Erdjument-Bromage, H., Olshen, A.B., Fleisher, M., Lilja, H., Brogi, E., Boyd, J. et al. (2006). Differential exoprotease activities confer tumor-specific serum peptidome patterns. *J Clin Invest* 116, 271–284.

Westermeier, R. and Marouga, R. (2005). Protein detection methods in proteomics research. *Biosci Rep* 25, 19–32.

Yager, T.D., Nickerson, D.A. and Hood, L.E. (1991). The Human Genome Project: Creating an infrastructure for biology and medicine. *Trends Biochem Sci* 16, 454–458.

Yates, J.R., Eng, J.K., McCormack, A.L. and Schieltz, D. (1995). Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal Chem* 67, 1426–1436.

RECOMMENDED RESOURCES

Books

Proteomics for Biological Discovery by Timothy D. Veenstra and John R. Yates

Introduction to Proteomics: Tools for the New Biology by Daniel C. Liebler

The Proteomics Protocols Handbook by John M. Walker

Proteomics in Cancer Research by Daniel C. Liebler

Proteins and Proteomics: A Laboratory Manual by Richard Simpson

Journals

Molecular and Cellular Proteomics (mcponline.org)

Journal of Proteome Research (jpr.org)

Proteomics (proteomics.org)

Proteomics – Clinical Applications

Websites

http://expasy.org

http://www.hupo.org

http://www.eupa.org

http://www.asms.org

CHAPTER

Comprehensive Metabolic Analysis for Understanding of Disease

Christopher B. Newgard, Robert D. Stevens, Brett R. Wenner, Shawn C. Burgess, Olga Ilkayeva, Michael J. Muehlbauer, A. Dean Sherry and James R. Bain

INTRODUCTION

Despite decades of research, the molecular etiology of chronic and indolent diseases and conditions such as diabetes, obesity, cancer, and cardiovascular disease remains obscure, in part because they represent slowly developing phenotypes caused by the interplay of a complex mixture of genetic and environmental effectors. For similar reasons, it has been very difficult to detect and diagnose these diseases at a stage that allows sufficient time for effective intervention. Comprehensive measurement of intermediary metabolites and changes in metabolic activity (metabolic flux) could lead to improved detection and understanding of these diseases.

As will be discussed in this chapter, instrumentation for measurement of large groups of intermediary metabolites as well as computational methods for analyzing such data are emerging. This surge in development of "metabolomics" technologies confers a number of potential advantages to scientists interested in disease research. First, it is estimated that human beings contain approximately 5000 discrete small molecule metabolites, far less than the estimate of 25,000 genes and >100,000 proteins. Thus, metabolomics may ultimately be the most tractable of the "omics" sciences. Second, metabolomics measures changes in metabolic or chemical milieu that are downstream of genomic and proteomic alterations, therefore potentially providing the most integrated picture of biological status. Third, identification of metabolic fingerprints for specific diseases may have particular practical utility for development of therapies, because metabolic changes immediately suggest enzymatic drug targets. And fourth, metabolomics is likely to be a powerful and precise tool for discerning mechanisms of action and possible toxicological effects of drug therapies.

The purpose of this chapter is to describe current methodologies for metabolic profiling and flux analysis and to provide selected examples of insights into biological and disease mechanisms gained from these approaches. Studies involving integration of metabolomics with genomic and transcriptomic profiling methods are just beginning to emerge, and early examples of this exciting area will also be highlighted. It is important

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

to emphasize from the outset that metabolic profiling tools are in a rapid phase of development, such that our description of currently useful methods will likely represent a "snapshot in time" that will evolve rapidly over the next several years.

CURRENT METABOLOMICS PLATFORMS: BASIC TOOLS AND GENERAL FEATURES

As described in other chapters in this volume, genome-wide analysis of genetic variation ("genomics") and surveys of changes in mRNA abundance by microarray technologies ("transcriptomics") are examples of robust and relatively mature technologies for investigation of mechanisms underlying biological variability. For example, mRNA profiling has developed to a point that microarray cores have become common in both the academic and private sectors, with many thousands of RNA profiling experiments performed on a yearly basis. This is not yet the case for laboratories that focus on comprehensive metabolic analysis. One reason for this is the complexity inherent in attempting to measure small molecule metabolites (intermediary metabolites as discussed in this chapter are defined as non-peptidic and with a molecular weight of <3000) in a quantitatively rigorous fashion. Underlying issues include the variable lability of metabolites (e.g., amino acids are often quite stable, whereas organic acid or lipid species can be quite labile), the chemical diversity of different metabolite classes, problems encountered in efficient extraction of metabolites from different biological matrices (e.g., tissues, blood, urine, etc.), and the wide dynamic concentration range of metabolite species (ranging from sub-nanomolar to milimolar). Thus, it is not surprising that there is currently no single technology for measurement of all of the metabolites in the "metabolome". This means that the small number of existing laboratories that engage in comprehensive metabolic analysis have usually been constructed on an "*ad hoc*" basis, often driven by the direct technical experience and/or scientific goals and philosophies of the founding scientists. This leads to varying choices in instrument platforms, and places an even greater burden on individual laboratories for rigor in quantitative analysis, since standardized methods do not yet exist for the majority of analytes.

The two major instrument platforms for measuring metabolite levels in biological samples are nuclear magnetic resonance (NMR) and mass spectrometry (MS) (Dettmer et al., 2007; German et al., 2005; Hollywood et al., 2006; Lindon et al., 2007; Watson, 2006). Both technologies can also be used for metabolic flux analysis following provision of stable isotope-labeled metabolic fuels (e.g., U-^{13}C glucose) to cells and organisms, as discussed later. In general, when applying these technologies to assays of metabolite levels, research groups in the field tend to fall into two camps that adopt either "unbiased"/"top-down" approaches or "targeted"/"bottom-up" methods as their core technologies. Unbiased profiling involves use of either NMR

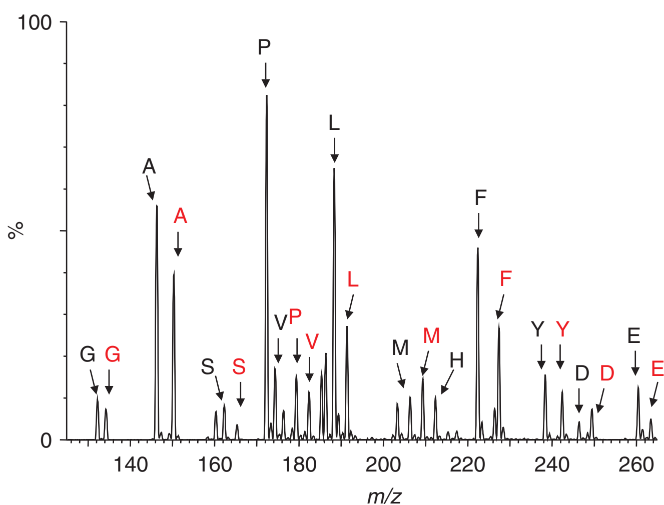


Figure 9.1 Tandem mass spectrum of butyl esters of acidic and neutral amino acids in serum. Data were obtained using a neutral loss of 102 Da scan function. The figure illustrates the use of internal standards for "targeted" quantitative profiling of metabolites. Each amino acid in serum (shown by black letters) is measured against its own stable-isotope-labeled internal standard (shown in red letters). G, glycine, A, alanine, S, serine, P, proline, V, valine, L, leucine, M, methionine, H, histidine, F, phenylalanine, Y, tyrosine, D, aspartate, E, glutamate.

or MS for measurement of as many metabolites as possible in a biological specimen simultaneously, regardless of the chemical class of the metabolites. When applied to unbiased profiling, both NMR and MS have advantages and limitations, as discussed below, and neither technology can presently be used for surveying all of the metabolites in a sample in a quantitative fashion. In contrast, targeted profiling focuses on quantification of discrete clusters of chemically related metabolites ("modules") using various combinations of chromatographic separations technologies and MS instruments that are most compatible with the analyte class. For example, electrospray ionization (ESI) and tandem mass spectrometry (MS/MS) are used for quantitative measurements of multiple amino acids and urea cycle intermediates or acylcarnitines (byproducts of mitochondrial oxidation of fatty acids, amino acids and glucose) of varying chain length and degree of saturation, whereas gas chromatography-MS (GC-MS) is used routinely for measurement of multiple species of free fatty acids and organic acids.

For each separate module, accurate quantification of the analytes is facilitated by addition of a cocktail of stable isotope labeled standards to the biological sample prior to the extraction and derivatization steps that may be necessary for the particular MS approach (see below and example shown in Figure 9.1). Ideally, each unknown analyte will be paired with its labeled cognate (usually an M+3 heavy isotope, e.g., methionine

TABLE 9.1 Summary of mass spectrometry (MS) methods used for metabolic profiling

Technique	Mass Analyzer	Reference	Metabolites
GC/MS	Quadrupole	Fiehn et al. (2000)	Fatty acids, alcohols, sterols, aliphatic and
		Roessner et al. (2001)	aromatic acids, amino acids, sugars, sugar
	TOF	Gullberg et al. (2004)	alcohols, phosphates, polyamines
		Pope et al. (2007)	
		Denkert et al. (2006)	
FIA or Infusion	QqQ	An et al. (2004)	Acylcarnitines, amino acids, coenzyme-As,
		Haqq et al. (2005)	
	TOF	Allen et al. (2003)	
	Q-TOF	Scholz et al. (2004)	
	FT-ICR	Oikawa et al. (2006)	
LC/MS	QqQ	Bajad et al. (2006)	Phoshpholipids, sphingolipids, eicosanoids,
		Sabatine et al. (2005)	amino acids, sulfoconjugates, carboxylic acids,
		Han and Gross (2003)	amines, glucuronides, glycosides, glutathione
		Murphy et al. (2005)	conjugates
	QIT	Lafaye et al. (2003)	
		Lafaye et al. (2004)	
	TOF	Plumb et al. (2005)	
	Q-TOF	Lenz et al. (2007)	
		Rainville et al. (2007)	
CE/MS	Quadrupole	Sato et al. (2004)	Amino acids, sugar nucleotides, carboxylic acids,
		Williams et al. (2007)	sugar phosphates, amines
	QqQ-IT	McNally et al. (2006)	
		McNally et al. (2007)	
	TOF	Soga et al. (2006)	
	Q-TOF	Soga et al. (2006)	
Direct Ionization	MALDI-TOF	Yu et al. (2006)	
		Fraser et al. (2007)	Amino acids, fatty acids, carotenoids,
	MALDI-TOF/TOF	Sun et al. (2007)	nucleosides, carboxylic acids
	DESI-QIT	Cooks et al. (2006)	
		Pan et al. (2007)	

Abbreviations: GC, Gas Chromatography; MS, Mass Spectrometry; LC, Liquid Chromatography; CE, Capillary Electrophoresis; TOF, Time of Flight; FT-ICR, Fourier Transform Ion Cyclotron Resonance;Q, Quadrupole; QqQ Triple Quadrupole; QIT Quadrupole Ion Trap; MALDI, Matrix Assisted Laser Desorption Ionization; DESI, Desorption Electrospray Ionization.

with methyl-D_3 methionine, pyruvate with $^{13}C_3$-pyruvate, etc.) to control for differences in analyte loss during chemical extraction and/or derivatization of individual analyte species and to compensate for ionization suppression effects. In practice, the limited range of stable isotope-labeled metabolites available from commercial suppliers and their relatively high cost constitute a current impediment to more broad-scale development and deployment of these technologies. To overcome this problem in the future, it may be necessary for metabolomics core laboratories to import chemical synthesis capabilities for production of comprehensive libraries of internal standards.

Table 9.1 provides a summary of various MS-based metabolic profiling technologies that are currently used for measurement of different metabolite classes.

COMPARISON OF NMR AND MS TECHNOLOGIES FOR UNBIASED METABOLIC PROFILING

Nuclear magnetic resonance spectroscopy is theoretically an excellent tool for unbiased metabolic profiling of all small

molecule metabolites, since the method is based on detection of any molecules that contain carbon or hydrogen (German et al., 2005; Lindon et al., 2007). Moreover, the method is non-destructive, in that analyses can be conducted directly in bodily fluids, cells, and even in intact tissues without the need for chemical extraction or derivatization of the individual analytes. These advantages are off-set by significant technical challenges, including poor sensitivity (it is generally not possible to detect metabolites at sub-micromolar levels with NMR) and the difficulty of deconvoluting spectra of complex metabolite mixtures such as typically found in biological matrices like plasma, urine, or tissue extracts. When NMR is applied to such samples, all metabolites within the detection limits of the technology resolve as hundreds to thousands of proton magnetic resonance peaks. The complexity of these data and the absence of comprehensive libraries of NMR spectra make it very difficult to assign the different resonances to specific metabolites. Thus, although NMR spectra are rich with information, the full harvesting of these datasets is not yet possible. Instead, peaks within NMR datasets are often treated as independent statistical variables, and the data analyzed by methods such as principal components analysis (PCA) to identify clusters of resonances that characterize different biological or disease states.

Mass spectrometry has the immediate advantage of much higher sensitivity compared to NMR, and the most advanced MS platforms such as Fourier transform ion cyclotron mass spectrometers (FT-ICR-MS) have the capacity to detect metabolites at concentrations as low as the femtomole range (Dettmer et al., 2007). Moreover, modern MS platforms have very high-mass resolution (>100,000), can resolve literally thousands of individual small molecules without the need for chromatography, and can achieve mass accuracies of <1 ppm. Seemingly, current MS platforms could therefore provide a means of circumventing problems encountered when using NMR methods for unbiased metabolic profiling of complex biological matrices, as described earlier. In practice, however, MS techniques usually require a sample extraction step that can cause metabolite losses. Moreover, different kinds of MS instruments use varying methods for the generation and detection of ions, and in concert, different kinds of metabolites have discrete ionization properties and are therefore variably resolved and detected in instruments that rely on mass and charge for analyte separation (Dettmer et al., 2007). Thus, while several current MS instrument platforms can be used for resolution and detection of large numbers of metabolites (e.g., high-performance liquid chromatography-MS, Q-TOF-MS, and FT-ICR-MS), the variable lability, solubility, recovery, and ionization of the different analyte species, coupled with the lack of broad-coverage spectral libraries makes these unbiased methods inherently non-quantitative and useful mainly for detecting gross changes in spectral peaks in samples of different biological origin. Similar to NMR, masses in these complex MS spectra are often not assigned to a specific analyte, and the data are instead analyzed by PCA or other multi-factorial statistical tools. However, an advantage of MS-based unbiased profiling methods is that when specific peaks or clusters of peaks consistently differentiate biological states (e.g., a metastatic tumor from a benign tumor), their detection by high-resolution MS platforms provides an immediate clue for analyte identification by providing an accurate mass estimate of the analyte. These mass estimates can be used to query databases such as KEGG, HMP, and ChemSpider for chemical formulae that match the estimated mass, thereby usually providing a strong starting point for identifying the analyte, and ultimately facilitating a return to more mechanistically oriented studies. The number of possible formulae is also significantly reduced by high-mass accuracy; for example, the number of possibilities for a mass ~500 at a mass accuracy 5 ppm with an elemental palate CHNOS is ~23. A mass accuracy of 2 ppm would reduce the number to 10 and an accuracy of 1 ppm would yield only 4 possibilities.

MS METHODS FOR TARGETED METABOLIC PROFILING

For scientists with core interests in biological mechanisms rather than biomarkers, knowledge about the identity of metabolites being surveyed and their exact concentration greatly enhances the value of the datasets. This means that some metabolic profiling laboratories are built in a bottom-up fashion, with emphasis on targeted and quantitative MS-based analyses. Although these approaches tend to utilize simpler tools such as ESI-MS/MS and GC-MS, significant challenges must still be overcome in building rigorous and fully vetted analysis modules for groups of metabolites. For example, GC-MS, when coupled with the use of stable isotope standards, can provide sufficient chromatographic resolution and analyte-specific detection to allow quantification of groups of metabolites in a class (e.g., fatty acids of different change length and degree of saturation or organic acids/TCA cycle intermediates). However, because the chromatography is performed in the gas phase at high-temperature, analytes must be volatile and have sufficient thermal stability to survive the analysis. To help stabilize the analytes under study, reactive carboxyl, sulfhydryl, amine, or hydroxyl groups are derivatized by akylation, oximation, acylation, or silylation (Dettmer et al., 2007; German et al., 2005). These methods, while effective, add complexity beyond that already introduced by sample extraction to the analytical procedures, and can result in batch-to-batch variation if not properly controlled. Similarly, analysis of acylcarnitines or amino acids/urea cycle intermediates by ESI-MS/MS requires different derivatization steps for each module – for example, treatment with acidic methanol for acylcarnitines and n-butanol for amino acids/urea cycle intermediates (Haqq et al., 2005). Nevertheless, with care, these modules can provide quantitative data for highly informative clusters of analytes (e.g., fatty acids, fatty acylCoAs, and acylcarnitines of different chain length and degree of saturation, amino acids/urea

cycle intermediates, and organic acids), with coefficients of variation in replicate assays in the range of 10%.

EXAMPLES OF NMR-BASED METABOLIC PROFILING IN DISEASE RESEARCH

The most common application of ^{1}H-NMR-based metabolic profiling in mammals and humans to date has been for evaluation of toxic effects of xenobiotic drugs. Work in this area has been reviewed by others (Lindon et al., 2007), so we focus instead on newer applications in the realm of disease research.

Coronary artery disease (CAD) is the leading cause of death in the United States and, in concert with the epidemic of obesity and diabetes, is rapidly becoming the leading cause of death in many developing countries. The genetic predilection of CAD is well established; family history has repeatedly been shown to be a robust, independent risk factor for CAD. Many of the commonly accepted risk factors for CAD are metabolic, such as lipid abnormalities and diabetes, making this disease a particularly logical target for comprehensive metabolic profiling technologies. Indeed, ^{1}H NMR was recently reported to detect metabolic profiles that predicted the presence and severity of CAD (Brindle et al., 2002). The study involved comparison of a relatively small group of patients with triple vessel coronary heart disease ($n = 36$) to angiographically normal subjects ($n = 30$). Partial least squares-discriminant analysis was used to identify peaks in the major lipid regions of the spectra that appeared to provide separation between the groups. The specific lipid species involved were not identified by this analysis, although it was suggested that choline-containing metabolites were particularly diagnostic. However, a subsequent study using very similar techniques demonstrated that the predictive value of the NMR-based metabolic profiles was weak when other factors such as gender and use of medical interventions such as statins were taken into account (Kirschenlohr et al., 2006). This second group of authors demonstrated that the ^{1}H NMR technique could identify male versus female subjects with 100% accuracy, but was much less able to identify statin users or subjects with CAD, despite expectations of substantial changes in lipid profile in the former group. This study makes it clear that ^{1}H NMR is currently not a substitute for the more invasive procedure of angiography in detection of CAD.

An intriguing and more promising recent application of ^{1}H NMR-based metabolic analysis has been to study the influence of intestinal bacteria (microbiota) on pathophysiologic outcomes of high-fat (HF) feeding (Dumas et al., 2006). This involved feeding of a HF diet to a mouse strain (129S6) known to be susceptible to HF-induced hepatic steatosis and insulin resistance, and for comparison, to a strain (BALBc) with relative resistance to these effects of diet. Metabolic profiling of plasma and urine samples revealed low circulating levels of plasma phosphatidylcholine (PChol) and high levels of methylamines in urine in the 129S6 strain. Since the conversion of choline to methylamines is accomplished by gut microbiota rather than mammalian enzymes (Zeisel et al., 1983), and choline deficient diets are associated with hepatic steatosis (Buchman et al., 1995), the authors propose that the increased propensity of the 129S6 strain for hepatic steatosis and insulin resistance could be due to a reduced PC pool necessary for the assembly of VLDL particles, leading to deposition of triglycerides in liver. As with its application to diagnosis of CAD, the value of NMR for detecting the contributions of bacterial metabolism to whole animal metabolic profiles and disease development remains to be confirmed or refuted by an independent study. However, in this instance, other studies suggest that this could indeed be an important pathway. Thus, treatment of germ-free mice with microbiota from the cecum of conventionally raised animals produced an increase in body fat content and appearance of insulin resistance within 14 days of transfer (Backhed et al. 2004). In addition, more recent studies have revealed that genetically obese rodents or obese humans have changes in relative abundance of two main intestinal bacterial groups, the Bacteroidetes and the Firmicutes, relative to lean controls, and that the "obese microbiome" has an increased capacity to harvest energy from the diet (Turnbaugh et al., 2006). The mechanism by which the obese microbiome influences peripheral metabolism of the host seems to involve changes in circulating levels of fasting-induced adipose factor (FIAF), a circulating lipoprotein lipase inhibitor, as well as changes in activity of the important metabolic regulatory proteins PGC-1α and 5′ AMP-activated kinase (AMPK) in skeletal muscle and liver (Backhed et al., 2007). An intriguing area for future study is to attempt to define metabolic profiles for specific gastrointestinal bacteria such that their relative contributions to changes in overall metabolic profile can be monitored and fully understood in the context of fuel homeostasis in the host.

EXAMPLES OF TARGETED MS-BASED METABOLIC PROFILING FOR UNDERSTANDING OF DISEASE MECHANISMS

Development of Tools

The past decade has seen a rapid increase in the use of MS-based analysis platforms for metabolic profiling. Many of the early applications of MS technologies were in the area of plant biology (Fiehn et al., 2000) or for detecting inborn errors of metabolism in newborn children (Frazier et al., 2006; Shekhawat et al., 2005). Application of tandem mass spectrometry (MS/MS) to newborn screening has allowed detection of more than 40 different genetic diseases of lipid and amino acid metabolism. In newborn screening, less emphasis is placed on absolute quantification of multiple analytes in a module, since the screen only requires detection of changes in a single

or discrete cluster of analytes with respect to established laboratory norms. For example, a defect in HMG-CoA lyase results in large and specific increases in C5-OH-carnitine and methylglutaryl-carnitine species detected by MS/MS, whereas defects in long-chain L-3-hydroxyacyl CoA dehydrogenase (LCHAD) or mitochondrial trifunctional protein (MTP) are associated with increases in C16-OH and C18-OH-carnitine metabolites (Shekhawat et al., 2005). Similarly, defects in amino acid catabolizing enzymes such as phenylalaine hydroxylase to cause phenylketonuria, or the branched-chain α-ketoacid dehydrogenase complex to cause Maple Syrup urine disease are readily detected by dramatic increases in phenylalanine or branched-chain amino acid (BCCA) levels, respectively.

In more recent years, many of the MS-based targeted metabolic profiling techniques developed for diagnosing inborn errors of metabolism have been adopted and refined for studies of metabolic regulatory mechanisms, disease detection, and mechanisms of disease pathogenesis. Selective examples, meant to be illustrative rather than comprehensive, are provided.

One approach has been to combine several targeted MS-based assay modules as an assemblage of tools that can report on several critical metabolic pathways (An et al., 2004; Haqq et al., 2005; Koves et al., 2008). Using a combination of GC/MS and MS/MS, we are currently able to perform quantitative analysis of approximately 150 metabolites in six groups:

1. free fatty acids of varying chain length and degrees of saturation;
2. total fatty acids (free+esterified);
3. acylcarnitines (representing products of mitochondrial fatty acid, amino acid, and glucose oxidation);
4. acyl CoAs of varying chain length and degree of saturation;
5. organic acids (TCA cycle intermediates and related metabolites); and
6. amino acids, including urea cycle intermediates.

Although the total number of analytes measured with these tools is small relative to estimates of 5000 total metabolites in the "metabolome", they are nevertheless highly useful for understanding changes in metabolic fuel selection under different physiologic and pathophysiologic circumstances. Moreover, expansion of the platform to include a broader range of analytes of interest in disease pathogenesis would appear possible in the near term, via adoption of published methods for analysis of sphingolipids (including ceramides), phospholipids, and prostaglandins and their metabolites (eicosanoids) (Dettmer et al., 2007; Han and Gross, 2005; Merrill et al., 2005; Watson, 2006). However, the reader should appreciate that development of new modules is not a trivial or inexpensive undertaking, as it requires acquisition or synthesis of stable isotope-labeled internal standards to cover most if not all of the individual analytes in the module, development of extraction procedures that are efficient for multiple analytes in a class, and demonstration of quantitative reproducibility of the method. The advantage gained is that such tools can be applied to understanding of metabolic regulatory mechanisms in isolated cells, animal models of disease, and human disease states, as summarized in the following specific examples.

Metabolic Profiling for Studies of Metabolic Regulatory and Signaling Mechanisms in Cultured Cells

An example of integration of NMR- and MS-based tools for understanding of metabolic regulation and signaling mechanisms comes from recent work on the process of glucose-stimulated insulin secretion (GSIS) in pancreatic islet β-cells (Boucher et al., 2004; Jensen et al., 2006; Joseph et al., 2006; Joseph et al., 2007; Lu et al., 2002; Ronnebaum et al., 2006; Ronnebaum et al., 2008). Stimulation of islet β-cells with glucose causes increases in insulin secretion within seconds to minutes, and this response is mediated by signals that are generated by β-cell glucose metabolism. A commonly accepted idea is that increases in the rate of glucose metabolism in β-cells leads to increases in ATP:ADP ratio, which causes inhibition of ATP-sensitive K^+ channels (K_{ATP} channels), membrane depolarization, and activation of voltage-gated calcium channels. The resulting influx of extracellular Ca^{2+} is then thought to stimulate exocytosis of insulin containing secretory granules. However, this model clearly does not provide a complete description of signals that regulate GSIS, since pharmacologic or molecular inhibition of K_{ATP}-channel activity does not abrograte glucose regulation of insulin secretion (Henquin et al., 2003; Nenquin et al., 2004).

To investigate non-K_{ATP}-channel related signals for GSIS, a set of insulinoma (INS-1)-derived cell lines with varying capacities for GSIS were developed (Hohmeier et al., 2000). Application of ^{13}C NMR-based isotopomer analysis and MS-based profiling of intermediary metabolites led to the discovery of a critical link between pyruvate carboxylase (PC)-mediated pyruvate exchange with TCA cycle intermediates ("pyruvate cycling") and GSIS, and demonstration that these pathways are dysregulated in lipid-cultured and dysfunctional β-cells (Boucher et al., 2004; Jensen et al., 2006; Joseph et al., 2006; Joseph et al., 2007; Lu et al., 2002; Ronnebaum et al., 2006). More recent studies have focused on identification of the specific pyruvate cycling pathways that may be involved in generation of signals for insulin secretion. One important pathway appears to involve export of citrate and/or isocitrate from the mitochondria via the citrate/isocitrate carrier (CIC), and subsequent conversion of isocitrate to α-ketoglutarate (α-KG) by the cytosolic NAPD-dependent isoform of isocitrate dehydrogenase (ICDc) (Joseph et al., 2006; Ronnebaum et al., 2006). siRNA-mediated suppression of CIC or ICDc cause substantial impairment of GSIS in both insulinoma cell lines and primary rat islets. These studies suggest that a metabolic byproduct of pyruvate/isocitrate cycling may be an important amplifying signal for control of GSIS. Possible mediators that are under investigation include NADPH, α-KG or its metabolites, or GTP generated by the succinyl CoA dehydrogenase reaction, all of which are direct or downstream products of the ICDc reaction.

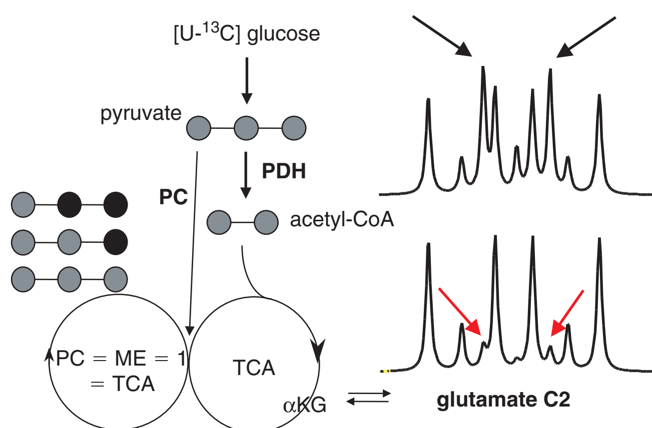


Figure 9.2 Use of ^{13}C NMR for measurement of metabolic flux in cells. To measure flux of pyruvate into the TCA cycle via the oxidative enzyme pyruvate dehydrogenase (PDH) versus the anaplerotic enzyme pyruvate carboxylase (PC), [U-^{13}C] glucose is administered to cells in culture. After 3 hours of incubation, cells are extracted and the NMR spectra of different carbons of glutamate are obtained. If all of the ^{13}C-labeled glucose is converted to pyruvate to enter the TCA cycle through PDH, the spectrum of carbon 2 of glutamate will appear as shown at the bottom right of the figure. In contrast, if the labeled pyruvate enters the TCA cycle both through PDH and PC in equal proportions, the glutamate C2 spectrum will appear as shown at the top right of the figure. Arrows show significant difference in peak height in the two spectra. The change in the spectra under these two scenarios occurs because pyruvate that enters the TCA cycle via PC can engage in pyruvate recycling pathways that results in dilution of ^{13}C (black circles) with natural abundance ^{12}C (gray circles), thus changing the pool of glutatmate mass isotopomers. Spectra of other glutamate carbons are obtained in the same fashion, and the aggregate data are deconvoluted to derive estimates of flux through PDH versus PC (Boucher et al., 2004; Lu et al., 2002; Sherry et al., 2004).

In the foregoing studies, measurement of pyruvate cycling flux was accomplished by incubation of β-cells in low (3 mM) or high (12 mM) concentrations of U-^{13}C glucose for several hours, followed by extraction of cells and analysis of the glutamate spectrum by NMR (see e.g., Figure 9.2). Specific resonances for each of the carbons of glutamate are affected by the population of mass isotopomers (glutamate with varying mixtures of ^{13}C and ^{12}C at each of the 5 carbons of the molecule), and this information can be used to calculate flux through the oxidative (PDH) and anaplerotic (PC) entry points of the tricarboxylic acid (TCA) cycle (Boucher et al., 2004; Lu et al., 2002; Sherry et al., 2004). When coupled to measurements of oxygen consumption (respiration), these rates can be expressed in absolute rather than relative terms (Boucher et al., 2004; Sherry et al., 2004). These methods revealed that the capacity for GSIS in variously glucose responsive INS-1-derived cell lines was tightly correlated with PC-catalyzed pyruvate cycling activity, but not PDH-catalyzed glucose oxidation.

To gain deeper understanding of the specific pyruvate cycling pathway involved in GSIS, these NMR-based methods of flux anlaysis were combined with MS/MS and GC/MS methods for profiling of acylcarnitines and organic acids, respectively. These experiments revealed that siRNA-mediated suppression of PC activity has little influence on pyruvate cycling rates or GSIS due to a compensatory rise in C2-acylcarnitine, which reflects the acetyl CoA pool; acetyl CoA is an allosteric activator of PC, which offsets the decrease in PC protein. These studies therefore uncovered a compensatory mechanism for preservation of pyruvate cycling activity and GSIS even when PC levels are reduced (Jensen et al., 2006). In another study, organic acid profiling was used to show that suppression of CIC caused a fall in cytosolic citrate levels, as would be expected if a major pathway for export of citrate from the mitochondria to the cytosol is blocked (Joseph et al., 2006). Finally, suppression of ICDc caused a decrease in pyruvate cycling flux, as measured by NMR, coupled with a rise in lactate levels measured by GC/MS, suggesting that in the face of suppressed pyruvate cycling activity, pyruvate is converted to lactate by default (Ronnebaum et al., 2006). These examples illustrate two important application of MS-based metabolic profiling in cellular research: (1) use in validation of specific genetic engineering or pharmacologic manipulations of cellular systems; (2) integration with flux analysis to provide a complete picture of changes in metabolic pathways under varying experimental conditions.

Application of MS-Based Metabolic Profiling for Understanding of Human Disease Pathogenesis

Targeted MS-based metabolic profiling has also been increasingly applied to studies of human diseases and conditions. Specific examples of applications to the areas of obesity and cardiovascular disease (CAD) are discussed here, but these tools are also being applied to diverse areas such as cancer (Fan et al., 2006) and mental disorders (Kaddurah-Daouk et al., 2007).

It has been appreciated for many years that obesity is associated with impaired metabolic regulation, including development of insulin resistance in muscle, liver, and adipose tissue, as well as loss of mass and function of pancreatic islet β-cells (Muoio and Newgard, 2008). However, the specific mechanisms that link overnutrition and obesity to global metabolic dysfunction remain incompletely defined. We recently completed a metabolic analysis of 74 obese (median BMI of 36.6 kg/m^2) and 67 lean (median BMI of 23.2 kg/m^2) human subjects using the MS-based metabolic, hormonal and physiologic profiling tools described earlier (Haqq et al., 2005; Newgard et al., 2008). PCA was used to detect a metabolic signature comprised of changes in amino acid, acylcarnitine, and organic acid metabolites that suggests that pathways of BCAA catabolism are overloaded in obese humans. To test the significance of this for development of obesity-related disease, we fed normal rats on HF, high-fat with supplemented BCAA (HF/BCAA) or standard chow (SC) diets. Despite a low rate of body weight gain equivalent to the SC group, HF/BCAA rats were equally insulin resistant as HF rats, demonstrating a contribution of BCAA to development of

insulin resistance independent of body weight. However, the effects of BCAA on body weight and insulin sensitivity seem to require the presence of fat, since animals fed on SC diet with supplemented BCAA gain weight normally and have normal insulin sensitivity. These findings demonstrate a synergistic interaction of dietary protein and fat in the development of insulin resistance, and provide a demonstration of how broad-based metabolic profiling tools can identify unanticipated contributions of dietary and metabolic factors to disease development.

MS-based metabolic profiling has also been applied to cardiovascular disease. In one study, liquid chromatography/triple quadrupole high-sensitivity MS analysis was applied to plasma samples of 18 subjects with exercise-inducible ischemia compared to 18 normal controls (Sabatine et al., 2005). Blood samples were taken immediately before and immediately after exercise, and levels of 173 known and several more minor intermediary metabolites were measured. A number of metabolites were found to be discordant between the two groups, including lactate, byproducts of AMP metabolism, metabolites of the citric acid cycle, and several of the unknowns. Using the six most discordantly regulated metabolites, a metabolic ischemia risk score was created and found to be statistically correlated to the probability of ischemia. However, it should be noted that the number of subjects was quite small in this study, necessitating follow up studies that either confirm or refute these interesting initial findings.

In another recent study, MS/MS and GC/MS-based profiling of fatty acids, acylcarnitines, and amino acids was performed on plasma samples from 117 subjects from eight multiplex families with familial early-onset CAD (Shah et al., 2008). Using SOLAR and adjusting for variables such as diabetes, hypertension, dyslipidemia, body-mass-index (BMI), age and sex, high heritabilities were found for several amino acids (arginine, glutamate, alanine, ornithine, valine, leucine/isoleucine, h^2 0.50–0.85, $P = 0.00004 - 7.2 \times 10^{-15}$), free fatty acids (arachidonic, linoleic, h^2 0.57–0.59, $P = 0.0004 - 0.00005$), and acylcarnitines (h^2 0.54–0.86, $P = 0.002 - 0.0000002$). Moreover, PCA was used to identify several metabolite clusters, with several of these factors found to be highly heritable. Interestingly, GENECARD families showed two distinct metabolite profiles, tracking with their clinical characteristics and reflective of factors identified from PCA, suggesting different genetic backgrounds and consequent variation in control of key metabolic pathways that converge on CAD. However, once again, the utility of these profiles for predicting CAD in other high-risk families or in the general population will require further study.

2009 UPDATE

Since the original publication of this chapter, a number of interesting advances have been made in the area of application of metabolomics tools to understanding of disease mechanisms and susceptibility.

Integration of metabolomics tools with other forms of "omics" technologies is at a very early stage, but has great potential in chronic disease research. As an example, diabetes-resistant B6-ob/ob and diabetes-susceptible BTBR-ob/ob mice were crossed to create an F2 cohort that was analyzed by microarray analysis and targeted mass spectrometry-based metabolic profiling (Ferrara et al., 2008). Application of advanced computational methods defined correlations and inferred causality between metabolite and transcript levels (Chaibub Neto et al., 2008; Ferrara et al., 2008) and unveiled a novel network by which the amino acid glutamine controls expression of the key gluconeogenic enzyme phosphoenolpyruvate carboxykinase (PEPCK). These studies show how integration of transcriptomics and metabolomics can lead to identification of novel pathways for diabetes pathogenesis.

Metabolomics has also recently been integrated with genotyping for understanding the biochemical impact of common genetic polymorphisms (Geiger et al., 2008). These authors performed ESI-MS/MS to measure 363 metabolites in serum of 284 male participants in the KORA study (a general population study from Germany). Significant associations were observed between frequent single nucleotide polymorphisms (SNPs) and changes in specific metabolites. Moreover, polymorphisms in four specific genes encoding metabolic enzymes (FADS1, LIPC, SCAD, and MCAD) were linked to perturbations in the metabolic pathways in which the enzymes are known to reside. The authors suggest that the combination of genotyping and metabolic phenotyping may provide new roadmaps for application of personalized medicine.

Other studies have recently emerged in the realm of applying metabolomics to better understand metabolic lesions in heart failure and myocardial infarction (MI). In one study, LC-MS-based profiling of approximately 200 metabolites was performed on subjects undergoing planned MI via alcohol-mediated septal ablation, or on subjects with spontaneous MI or undergoing elective coronary angiography as positive and negative controls, respectively (Lewis et al., 2008). Five metabolites were altered in both spontaneous and planned MI, and this metabolic signature may become useful for early detection of myocardial injury with further validation.

A second study applied targeted ESI-MS to measure 63 metabolites in arterial and coronary sinus blood obtained during cardiac surgery, before and after ischemia/reperfusion (I/R) (Turer et al., 2009). This work demonstrates that the preexisting ventricular state [left ventricular dysfunction (LVD), coronary artery disease, or neither condition] is associated with clear differences in myocardial fuel uptake, both at baseline and following I/R. In particular, LVD was associated with global

suppression of metabolic fuel intake (glucose, fatty acid, and amino acids) and limited myocardial metabolic reserve and flexibility following I/R. Moreover, altered metabolic profiles following I/R were associated with postoperative hemodynamic course. These findings help to focus future studies on optimal design of perioperative treatment regimens based on the particular form of cardiovascular disease and the metabolic status of the heart.

2009 UPDATE REFERENCES

Chaibub Neto, E., Ferrara, C.T., Attie, A.D. and Yandell, B.S. (2008). Inferring causal phenotype networks from segregating populations. *Genetics* 179(2), 1089–1100.

Ferrara, C.T., Wang P., Stevens, R.D., Neto E.C., Bain J.R., Choi Y., Keller M.P., Kendziorski C.M., Yandell, B.S., Wenner, B.R. et al. (2008). Genetic networks of liver metabolism revealed by integration of metabolic and transcriptomic profiling. *PLoS Genet* 4, e1000034.

Geiger, C., Geistlinger, L., Altmaier, E., Hrabe de Angelis, M., Kronenberg, F., Meitinger, R., Mewes, H.-W., Wichmann, H.-E., Weinberger, K.M., Adamski, J. et al. (2008). Genetics meets metabolomics: A genome-wide association study of metabolite profiles in human serum. *PLoS Genet* 4, e1000282.

Lewis, G.D., Wei, R., Liu, E., Yang, E., Shi, X., Martinovic, M., Farrell, L., Asnani, A., Cyrille, M., Ramanathan, A. et al. (2008). Metabolite profiling of blood from individuals undergoing planned myocardial infarction reveals early markers of myocardial injury. *J Clin Invest* 118, 3502–3512.

Turer, A.T., Stevens, R.D., Bain, J.R., Muehlbauer, M.J., van der Westhuizen J., Mathew, J.P., Schwinn, D.A., Glower, D.D., Newgard, C.B. and Podgoreanu, M.V. (2009). Metabolomic profiling reveals distinct patterns of metabolic substrate utilization in humans with coronary artery disease or left ventricular dysfunction during surgical ischemia-reperfusion. *Circulation*. 119(13), 1736–1746.

REFERENCES

Allen, J., Davey, H.M., Broadhurst, D., Rowland, J.J., Oliver, S.G. and Kell, D.B. (Jun 2003). High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nat Biotechnol* 21(6), 692–696.

An, J., Muoio, D.M., Shiota, M., Fujimoto, Y., Cline, G.W., Shulman, G.I., Stevens, R., Millington, D. and Newgard, C.B. (2004). Hepatic expression of malonyl CoA decarboxylase reverses muscle, liver, and whole-animal insulin resistance. *Nat Med* 10, 268–274.

Backhed, F., Ding, H., Wang, T., Hooper, L.V., Koh, G.Y., Nagy, A., Semenkovich, C.F. and Gordon, G.I. (2004). The gut microbiota as an environmental factor that regulates fat storage. *Proc Natl Acad Sci USA* 101, 15718–15723.

Backhed, F., Manchester, J.K., Semenkovich, C.F. and Gordon, J.I. (2007). Mechanisms underlying the resistance to diet-induced obesity in germ-free mice. *Proc Natl Acad Sci USA* 104, 979–984.

Bajad, S.U., Lu, W., Kimball, E.H., Yuan, J., Peterson, C. and Rabinowitz, J.D. (2006). Separation and quantitation of water soluble cellular metabolites by hydrophilic interaction chromatography-tandem mass spectrometry. *J Chromatogr* 1125, 76–88.

Boucher, A., Lu, D., Burgess, S., Telemaque-Potts, S., Jensen, M., Mulder, H., Wang, M-Y., Unger, R.H., Sherry, A.D. and Newgard, C. B. (2004). Mechanism of lipid-induced impairment of glucose-stimulated insulin secretion and its reversal by an analogue of malate. *J Biol Chem* 279, 27263–27271.

Brindle, J.T., Antti, H., Holmes, E., Tranter, G., Nicholson, J.K., Bethell, H.W.L., Clarke, S., Schofield, P.M., McKilligin, E., Mosedale, D.E. et al. (2002). Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics. *Nat Med* 8, 1439–1444.

Buchman, A.L., Dubin, M.D., Moukarzel, A.A., Jenden, D.J., Roch, M., Rice, K.M., Gornbein, J. and Ament, M.E. (1995). *Hepatology* 22, 1399–1403.

Cooks, R.G., Ouyang, Z., Takats, Z. and Wiseman, J.M. (2006). Detection Technologies. Ambient mass spectrometry. *Science* 311, 1566–1570.

Denkert, C., Budczies, J., Kind, T., Weichert, W., Tablack, P., Schouli, J., Niesporek, S., Kōnsgen, D., Dietel, M. and Fiehn, O. (Nov 15 2006). Mass spectrometry-based metabolic profiling reveals different metabolite patterns in invasive ovarian carcinomas and ovarian borderline tumors. *Cancer Res* 66(22), 10795–10804.

Dettmer, K., Aronov, P.A. and Hammock, B.D. (2007). Mass spectrometry-based metabolomics. *Mass Spectrom Rev* 26, 51–78.

Dumas, M.-E., Barton, R.H., Toye, A., Cloarec, O., Blancher, C., Rothwell, A., Fearnside, J., Tatoud, R., Blanc, V., Lindon, J.C. et al. (2006). Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. *Proc Natl Acad Sci USA* 103, 12511–12516.

Fan, T.W., Higashi, R.M. and Lane, A.N. (2006). Integrating metabolomics and transcriptomics for probing SE anticancer mechanisms. *Drug Metab Rev* 38, 707–732.

Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R.N. and Willmitzer, L. (2000). Metabolite profiling for plant functional genomics. *Nat Biotechnol* 18, 1157–1161.

Fraser, P.D., Enfissi, E.M., Goodfellow, M., Eguchi, T. and Bramley, P.M. (2007). Metabolite profiling of plant carotenoids using the matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Plant J* 49, 552–564.

Frazier, D.M., Millington, D.S., McCandless, S.E., Koeberl, D.D., Weavil, S.D., Chaing, S.H. and Muenzer, J. (2006). The tandem mass spectrometry newborn screening experience in North Carolina: 1997–2005. J Inherit Metab Dis 29, 76–85.

German, J.B., Hammock, B.D. and Watkins, S.M. (2005). Metabolomics: Building on a century of biochemistry to guide human health. *Metabolomics* 1, 3–9.

Gullberg, J., Jonsson, P., Nordstrom, A., Sjostrom, M. and Moritz, T. (2004). Design of experiments: An efficient strategy to identify factors influencing extraction and derivatization of Arabidopsis thaliana samples in metabolomic studies with gas chromatography/mass spectrometry. *Anal Biochem* 331, 283–295.

Han, X. and Gross, R.W. (2003). Global analyses of cellular lipidomes directly from crude extracts of biological samples by ESI mass spectrometry: A bridge to lipidomics. *J Lipid Res* 44, 1071–1079.

Han, X. and Gross, R.W. (2005). Shotgun lipidomics: Electrospray ionization mass spectrometric analysis and quantitation of cellular lipodomes directly from crude extracts of biological samples. *Mass Spectrom Rev* 24, 367–412.

Haqq, A., Lien, L.F., Boan, J., Arlotto, M., Slentz, C.A., Muehlbauer, M.J., Rochon, J., Gallup, D., McMahon, R.L., Bain, J.R. et al. (2005). The Study of the Effects of Diet on Metabolism and Nutrition (STEDMAN) Project: Rationale and Design. *Contem Clin Trials* 26, 616–625.

Henquin, J.C., Ravier, M.A., Nenquin, M., Jonas, J.C. and Gilon, P. (2003). Hierarchy of the beta-cell signals controlling insulin secretion. *Eur J Clin Invest* 33, 742–750.

Hohmeier, H.E., Mulder, H., Chen, G., Henkel-Rieger, R., Prentki, M. and Newgard, C.B. (2000). Isolation of INS-1-derived cell lines with robust ATP-sensitive K+ channel-dependent and -independent glucose-stimulated insulin secretion. *Diabetes* 49, 424–430.

Hollywood, K., Brison, D.R. and Goodacre, R. (2006). Metabolomics: Current technologies and future trends. *Proteomics* 6, 4716–4723.

Jensen, M.V., Joseph, J., Ilkayeva, O., Burgess, S., Lu, D., Ronnebaum, S., Odegaard, M., Becker, T.C., Sherry, A.D. and Newgard, C.B. (2006). Compensatory responses to pyruvate carboxylase suppression in islet beta-cells: Preservation of glucose-stimulated insulin secretion. *J Biol Chem* 281, 22342–22351.

Joseph, J.W., Jensen, M.V., Ilkayeva, O., Palmieri, F., Alarcon, C., Rhodes, C.J. and Newgard, C.B. (2006). The mitochondrial citrate carrier plays a regulatory role in glucose-stimulated insulin secretion. *J Biol Chem* 281, 35624–35632.

Joseph, J.W., Odegaard, M.L., Ronnebaum, S.M., Burgess, S.C., Muehlbauer, J., Sherry, A.D. and Newgard, C.B. (2007). Normal flux through ATP-citrate lyase or fatty acid synthase is not required for glucose-stimulated insulin secretion. *J Biol Chem*. 282, 31592–31600.

Kaddurah-Daouk, R., McEvoy, J., Baillie, R.A., Lee, D., Yao, J.K., Doraiswamy, P.M. and Krishnan, K.R. (2007). Metabolomic mapping of atypical antipsychotic effects in schizophrenia. *Mol Psychiatry* 12, 934–945.

Kirschenlohr, H.L., Griffin, J.L., Clarke, S.C., Rhydwen, R., Grace, A.A., Schofield, P.M., Brindle, K.M. and Metcalfe, J.C. (2006). Proton NMR analysis of plasma is a weak predictor of coronary artery disease. *Nat Med* 12, 705–710.

Koves, T.R., Ussher, J.R., Noland, R.C., Slentz, D.S., Mosedale, M., Ikayeva, O.J., Bain, J., Stevens, R., Dyck, J.R., Newgard, C.B., Lopaschuk, G. and Muoio, D.M. (2008). Mitochondrial overload and incomplete fatty acid oxidation contribute to skeletal muscle insulin resistance. *Cell Metab* 7, 45–56.

Lafaye, A., Junot, C., Ramounet-Le Gall, B., Fritsch, P., Tabet, J.C. and Ezan, E. (2003). Metabolite profiling in rat urine by liquid chromatography/electrospray ion trap mass spectrometry. Application to the study of heavy metal toxicity. *Rapid Commun Mass Spectrom* 17, 2541–2549.

Lafaye, A., Junot, C., Ramounet-Le Gall, B., Fritsch, P., Ezan, E. and Tabet, J.C. (2004). Profiling of sulfoconjugates in urine by using precursor ion and neutral loss scans in tandem mass spectrometry. Application to the investigation of heavy metal toxicity in rats. *J Mass Spectrom* 39, 655–664.

Lenz, E.M., Williams, R.E., Sidaway, J. et al. (2007). The application of microbore UPLC/oa-TOF-MS and 1H NMR spectroscopy to the metabonomic analysis of rat urine following the intravenous administration of pravastatin. *J Pharm Biomed Anal* 44, 845–852.

Lindon, J.C., Holmes, E. and Nicholson, J.K. (2007). Metabonomics in pharmaceutical R & D. *FEBS J* 274, 1140–1151.

Lu, D., Mulder, H., Zhao, P., Burgess, S.C., Jensen, M.V., Kamzolova, S., Newgard, C.B. and Sherry, A.D. (2002). ^{13}C NMR isotopomer analysis reveals a connection between pyruvate cycling and glucose-stimulated insulin secretion (GSIS). *Proc Natl Acad Sci USA* 99, 2708–2713.

McNally, D.J., Hui, J.P., Aubry, A.J. et al. (2006). Functional characterization of the flagellar glycosylation locus in Campylobacter jejuni 81–176 using a focused metabolomics approach. *J Biol Chem* 281, 18489–18498.

McNally, D.J., Aubry, A.J., Hui, J.P. et al. (2007). Targeted metabolomics analysis of Campylobacter coli VC167 reveals legionaminic acid derivatives as novel flagellar glycans. *J Biol Chem* 282, 14463–14475.

Merrill, A.H., Jr., Sullards, M.C., Allegood, J.C., Kelly, S. and Wang, E. (2005). Sphingolipidomics: high-throughput, structure-specific, and quantitative analysis of sphingolipids by liquid chromatography tandem mass spectrometry. *Methods* 36, 207–224.

Muoio, D.M. and Newgard, C.B. (2008). Mechanisms of disease: Molecular and metabolic mechanisms of insulin resistance and β-cell failure in type 2 diabetes. *Nat Rev Mollar Cell Biol* 9, 193–205.

Murphy, R.C., Barkley, R.M., Zemski, Berry, K., Hankin, J., Harrison, K., Johnson, C., Krank, J., McAnoy, A., Uhlson, C. and Zarini, S. (2005). Electrospray ionization and tandom mass spectrometry of eicosanoids. *Anal Biochem* 346, 1–42.

Nenquin, M., Szollosi, A., Aguilar-Bryan, L., Bryan, J. and Henquin, J.C. (2004). Both triggering and amplifying pathways contribute to fuel-induced insulin secretion in the absence of sulfonylurea receptor-1 in pancreatic beta-cells. *J Biol Chem* 279, 32316–32324.

Newgard, C.B., Bain, J.R., Muehlbauer, M.J., McMahon, R.L., Rochon, J., Gallup, D., Stevens, R., Ilkayeva, O., Millington, D., Butler, M.D. et al. (2007). Comprehensive metabolic profiling of obese versus lean humans reveals a signature of overloaded branch chain amino acid catabolism. *Submitted for publication*.

Oikawa, A., Nakamura, Y., Ogura, T. et al. (2006). Clarification of pathway-specific inhibition by Fourier transform ion cyclotron resonance/mass spectrometry-based metabolic phenotyping studies. *Plant Physiol* 142, 398–413.

Pan, Z., Gu, H., Talaty, N. et al. (2007). Principal component analysis of urine metabolites detected by NMR and DESI-MS in patients with inborn errors of metabolism. *Anal Bioanal Chem* 387, 539–549.

Plumb, R.S., Granger, J.H., Stumpf, C.L. et al. (2005). A rapid screening approach to metabonomics using UPLC and oa-TOF mass spectrometry: Application to age, gender and diurnal variation in normal/Zucker obese rats and black, white and nude mice. *Analyst* 130, 844–849.

Pope, G.A., MacKenzie, D.A., Defernez, M. et al. (2007). Metabolic footprinting as a tool for discriminating between brewing yeasts. *Yeast* 24, 667–679.

Rainville, P.D., Stumpf, C.L., Shockcor, J.P., Plumb, R.S. and Nicholson, J.K. (2007). Novel application of reversed-phase UPLC-oaTOF-MS for lipid analysis in complex biological mixtures: A new tool for lipidomics. *J Proteome Res* 6, 552–558.

Ronnebaum, S., Joseph, J.W., Burgess, S.C., Lu, D., Ilkayeva, O., Stevens, R., Becker, T.C., Muehlbauer, J., Sherry, A.D., Newgard, C.B. et al. (2006). A pyruvate cycling pathway involving cytosolic NAPD-dependent isocitrate dehydrogenase regulates glucose-stimulated insulin secretion. *J Biol Chem* 281, 31041–31049.

Roessner, U., Luedemann, A., Brust, D. et al. (2001). Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell* 13, 11–29.

Sabatine, M.S., Liu, E., Morrow, D.A., Heller, E., McCarroll, R., Wiegand, R., Berriz, G.F., Roth, F.P. and Gerszten, R.E. (2005). Metabolomic identification of novel biomarkers of myocardial ischemia. *Circulation* 112, 3868–3875.

Sato, S., Soga, T., Nishioka, T. and Tomita, M. (Oct 2004). Simultaneous determination of the main metabolites in rice leaves using capillary electrophoresis mass spectrometry and capillary electrophoresis diode array detection. *Plant J* 40(1), 151–163.

Shah, S.H., Hauser, E.R., Bain, J.R., Muehlbauer, M.J., Stevens, R.D., Wenner, B.R., Sketch, M., Dowdy, E., Granger, C.B., Ginsburg, G.S. et al. (2007). High heritability of metabolomic profiles in families burdened with early-onset cardiovascular disease. *Submitted for publication*.

Shekhawat, P.S., Matern, D. and Strauss, A.W. (2005). Fetal fatty acid oxidation disorders, their effect on maternal health and neonatal outcome: Impact of expanded newborn screening on their diagnosis and management. *Ped Res* 57, 78R–86R.

Scholz, M., Gatzek, S., Sterling, A., Fiehn, O. and Selbig, J. (2004). Metabolite fingerprinting: detecting biological features by independent component analysis. *Bioinformatics* 20, 2447–2454.

Sherry, A.D., Jeffrey, F.M. and Malloy, C.R. (2004). Analytical solutions for (13)C isotopomer analysis of complex metabolic conditions: Substrate oxidation, multiple pyruvate cycles, and gluconeogenesis. *Metab Eng* 6, 12–24.

Soga, T., Baran, R., Suematsu, M. et al. (2006). Differential metabolomics reveals ophthalmic acid as an oxidative stress biomarker indicating hepatic glutathione consumption. *J Biol Chem* 281, 16768–16776.

Sun, G., Yang, K., Zhao, Z., Guan, S., Han, X. and Gross, R.W. (2007). Shotgun metabolomics approach for the analysis of negatively charged water-soluble cellular metabolites from mouse heart tissue. *Anal Chem* 79, 6629–6640.

Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R. and Gordon, J.I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 444, 1027–1031.

Vaidyanathan, S., Kell, D.B. and Goodacre, R. (2002). Flow-injection electrospray ionization mass spectrometry of crude cell extracts for high-throughput bacterial identification. *J Am Soc Mass Spectrom* 13, 118–128.

Vaidyanathan, S., Gaskell, S. and Goodacre, R. (2006). Matrix-suppressed laser desorption/ionisation mass spectrometry and its suitability for metabolome analyses. *Rapid Commun Mass Spectrom* 20, 1192–1198.

Watson, A.D. (2006). Lipidomics: A global approach to lipid analysis in biological systems. *J. Lipid Res* 47, 2101–2111.

Williams, B.J., Cameron, C.J., Workman, R., Broeckling, C.D., Sumner, L.W. and Smith, J.T. (2007). Amino acid profiling in plant cell cultures: An inter-laboratory comparison of CE-MS and GC-MS. *Electrophoresis* 28, 1371–1379.

Yu, H., Lopez, E., Young, S.W., Luo, J., Tian, H. and Cao, P. (2000). Quantitative analysis of free fatty acids in rat plasma using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry with meso-tetrakis porphyrin as matrix. *Anal Biochem* 354, 182–191.

Zeisel, S.H., Wishnok, J.S. and Blusztajn, J.K. (1983). *J Pharmacol Exp Ther* 225, 320–324.

This page intentionally left blank

Informatics

Section 2

10. Bioinformatic and Computational Analysis for Genomic Medicine

11. Systems Biology and Systems Medicine

12. Electronic Medical Records in Genomic Medicine Practice and Research

13. Online Health Information Retrieval by Consumers

This page intentionally left blank

CHAPTER

Bioinformatic and Computational Analysis for Genomic Medicine

Atul J. Butte

INTRODUCTION

The past 10 years have seen development and deployment of a variety of measurement tools in molecular science that enable the large-scale parallel quantitative assessment of molecular state. The premier example of such measurement tools is the RNA expression detection microarray, which provides quantitative measurements of expression of over 40,000 unique RNAs within cells (Chee et al., 1996; DeRisi et al., 1996). Yet, as illustrated in previous chapters in this volume, microarrays represent just one of at least 30 or so measurement or experimental modalities available to investigators in molecular biology and genomics, including metabolite quantification (Edwards et al., 2001), DNA polymorphism measurements (Johnson et al., 2001), protein quantification (Espina et al., 2003; Ghaemmaghami et al., 2003; Gygi et al., 1999; Liotta et al., 2003), protein activity (Hestvik et al., 2003), protein interactions with small molecules (Jessani et al., 2002) and DNA (Lee et al., 2002; Odom et al., 2004; Ren et al., 2000; Wang et al., 2001; Weinmann et al., 2002), and many others. Though these molecular technologies are commonly described as being "high-throughput" or "genome-wide" in nature, we will use the term *high-bandwidth* to describe them in this chapter. An analogy explains our reasoning: *high-bandwidth* is to a *supermarket*, which serves a large variety of different foods irrespective of speed, while *high-throughput* is to a *fast-food restaurant*, serving a few types of foods quickly. *Genome-wide* technologies are biased towards genes, while we are covering measurements of many other types besides genes.

The service role for bioinformatics in research using these new tools is undeniable. Each of these high-bandwidth modalities yields a sizeable amount of raw data, so distilling these raw data and filtering out measurement noise through the proper use of bioinformatics methods is crucial. Bioinformatics clearly plays a role in the storage, retrieval, and sharing of these measurements from local and international repositories and in relating these measurements to clinical outcomes. However, what we will argue in this chapter is that the role for bioinformatics in genomic and personalized medicine is well beyond that of providing a service and lies more broadly in enabling new and interesting questions to be asked in biomedical, translational, and clinical research.

In this chapter, we will present several examples that illustrate how the use of specific bioinformatics methods has enabled a revolution of discovery in medical diagnostics, prognostics, therapeutics, anatomy, and nosology. We will then review a few specific analytic methods and standardized vocabularies useful in studying and identifying high-bandwidth measurements for genomic and personalized medicine. After covering a few freely available software tools valuable for conducting analyzes, we will end with some thoughts about where the future discoveries in genomic medicine, enabled by bioinformatics, may come from.

Copyright © 2010, Elsevier Inc.
All rights reserved.

VIGNETTES: HOW SPECIFIC BIOINFORMATICS METHODS CAN CHANGE THE PRACTICE OF MEDICINE

While many high-bandwidth measurement technologies have been used to enable basic discoveries in the life sciences, certainly the most intriguing and provocative use of these measurement modalities has been in enabling novel types of medical discoveries. Here, we will present case examples of how these discoveries were enabled by bioinformatics techniques and have demonstrably altered our abilities in four distinct areas of medicine: diagnosis, therapeutics, histopathology, and nosology (Table 10.1). For these vignettes, we will particularly focus on the use of microarrays, as this is currently the most commonly used high-bandwidth measurement modality.

Diagnosis

Diagnosis is a process in which a physician, when presented with a patient holding a chief complaint, gathers information through questioning, history taking, physical examination, and testing, compares the results of those queries with patterns established for known diseases, and assigns the disease with the highest likelihood. High-bandwidth testing in molecular biology has impacted every step of the diagnostic process. For example, those being evaluated for a disease may not yet even have a complaint, as DNA polymorphisms are discovered and used to predict future disease (Kohane et al., 2006). Medical history taking becomes problematic when considering the genetic context of a patient, which is increasingly known; for example, many infants now have one or more genes sequenced at birth as part of state-mandated newborn screening (Khoury et al., 2003). History taking has potential to become even more convoluted as more adults have genes (and even whole genomes) sequenced. The bioinformatics challenge in storing, providing, and teaching patients and parents how to use this type of medical information, while protecting against others finding this information, has only just started to be addressed (Adida and Kohane, 2006; 2004; Centers for Disease Control and Prevention, 2004; Henneman et al., 2006; Roche and Annas, 2001, 2006).

The nature of the physical examination itself has changed. Positive and negative physical findings (e.g., *acanthosis nigricans*, or darkening of the skin of the neck) are traditionally considered as probes into the process of disease (e.g., type 2 diabetes mellitus). With the democratization of genomics throughout all fields of medicine, physical findings can now be viewed collectively as phenotype and as a probe into the inner workings of the genome (Freimer and Sabatti, 2003). Two different diseases with similar sets of physical findings may indeed appear quite similar when viewed from the perspective of genes and gene products (Brunner and van Driel, 2004).

But certainly the greatest impact on diagnosis from high-bandwidth molecular testing has been felt on our ability to redefine our patterns of disease using these new measurements. One

TABLE 10.1 Bioinformatics techniques have demonstrably altered our abilities in four distinct areas of medicine

Area of the practice of medicine	Example of bioinformatics altering practice
Diagnosis	
Chief complaint	Predicting future disease from polymorphisms before complaint
Medical history	Genetic screening at birth should become part of a patient's life-long electronic medical record
Medical record keeping	Patient-owned Internet-based genomic records
Physical examination	Findings collectively viewed as phenotype and matched against patterns
Differential diagnosis	Molecular measurements can be compared against case-examples
Disease sub-type	Hierarchical clustering to determine novel subtypes from molecular measurements
Therapeutics	
Finding novel effects for existing drugs	Linking drug effects to molecular measurements
Side effects	Predicting drug adverse effect from pre-clinical molecular measurements
Finding novel therapies for untreatable conditions	Matching molecular profiles of disease with molecular profiles of drug effect
Histopathology	
Metastasis of unknown origin	Matching molecular measurements from metastasis with profiles of known primary cancers
Pathological mechanism of action of disease	Identifying differentially expressed genes and linking back to known biological pathways
Nosology	
Finding differences between diseases	Principal components of variance of disease sample microarrays
Finding similarities between diseases	Common signatures across microarray measurements of disease
Linking cellular models to human diseases	Comparing cellular and disease microarray measurements

of the earliest examples of this impact was in the work of Golub and colleagues published in 1999 (Golub et al., 1999). Acute lymphomas were first divided into acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML) in the 1960s based on histochemical testing, and more recently by antibody-based testing against cell-surface molecules, but there was no single established test to make this diagnosis. Golub was the first to show that gene expression microarrays could be used to measure gene expression levels from samples from acute lymphomas, and using a bioinformatics technique known as *supervised machine learning*, trained a prediction system to find and use those genes that uniquely distinguished ALL from AML (Figure 10.1). He then used this predictor to accurately classify additional new leukemia cases. Many others have used alternative computational and bioinformatics techniques to build predictors since this landmark study, while many additional diagnostic dilemmas have similarly

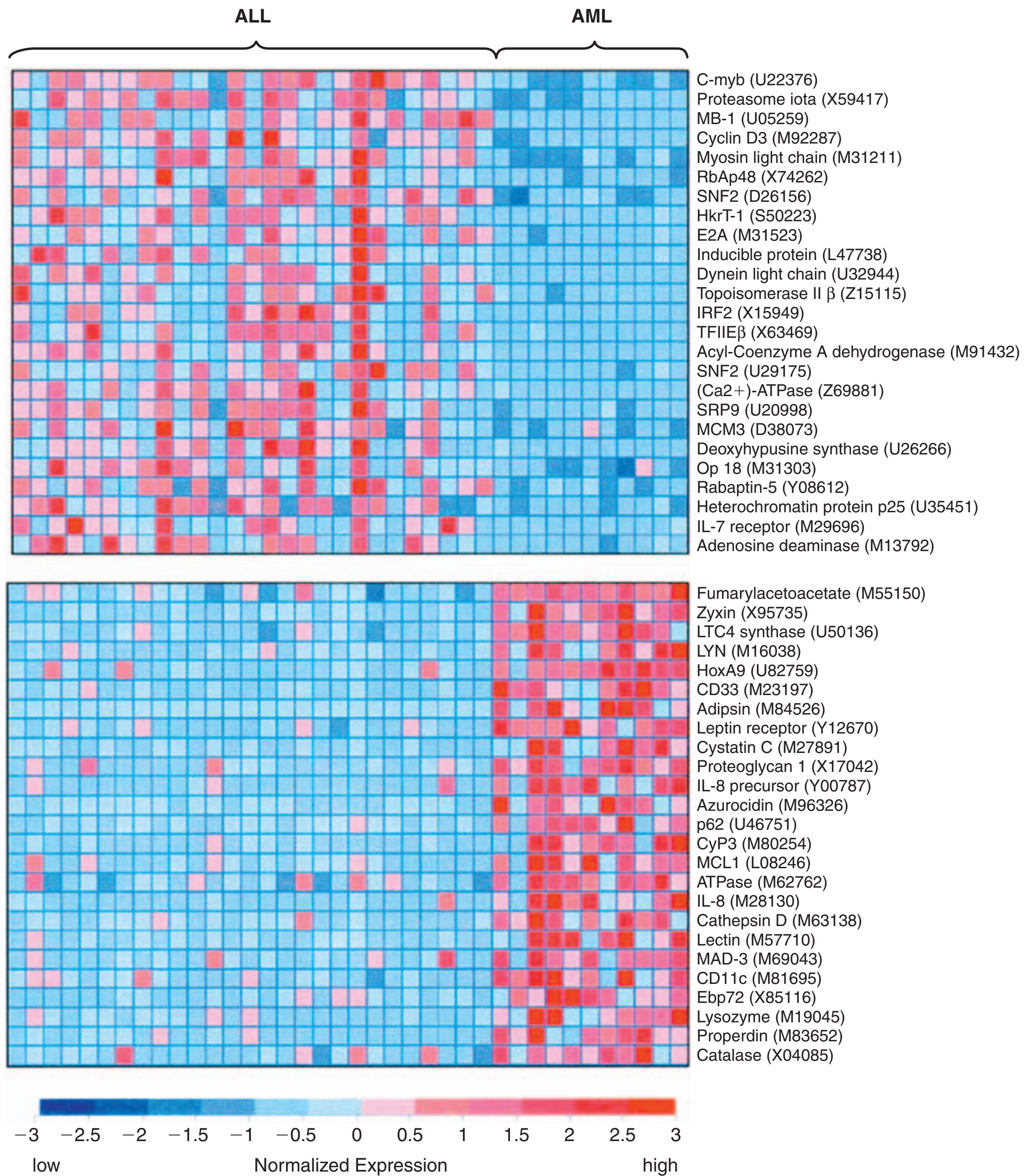


Figure 10.1 Todd Golub and colleagues measured genome-wide expression levels in samples from patients with ALL and AML. The top 50 genes distinguishing ALL and AML are displayed, with genes elevated in expression shown in pink, and genes decreased in expression shown in blue. From Golub et al., 1999 *Science* 286:531. Reprinted by permission from AAAS.

been addressed using gene expression microarrays, including melanoma (Bittner et al., 2000) and prostate cancer (Singh et al., 2002). Other types of high-bandwidth modalities have been used to assist in other diagnostic questions, including proteomics (Petricoin et al., 2002; Robinson et al., 2002) and metabolomics (Clayton et al., 2006) (see also Chapters 8 and 9).

While enabling physicians to make more efficient and accurate decisions between diseases itself is remarkable, the potential for high-bandwidth molecular testing in diagnosis is even greater. Beyond just improving on older tests for established diagnostic dilemmas, high-bandwidth studies have also been used to redefine and even divide diagnoses into novel sub-types. The earliest example of this was in the work of Alizadeh and colleagues in 2000 (Alizadeh et al., 2000). Alizadeh used gene expression microarrays to measure gene expression in samples from patients diagnosed with diffuse large B-cell lymphoma, a cancer of B-lymphocytes. This type of cancer was previously divided into low-, intermediate-, and high-grade categories based on growth pattern and immunohistochemistry. After applying a bioinformatics technique called *hierarchical clustering* (described later in this chapter) to the microarray data, Alizadeh discovered that the patient samples of B-cell lymphoma could be divided into an equal split of two sub-types (Figure 10.2). Most importantly, patients with these two different sub-types of lymphoma retrospectively demonstrated significant differences in survival by *Kaplan–Meier analysis*. This was a landmark study because it showed how high-bandwidth molecular measurements could provide a next level of resolution to resolve sub-types of disease not otherwise discernable by physicians. This type of study has been replicated for other types of cancers, including lung cancer (Bhattacharjee et al., 2001) and breast cancer (Sorlie et al., 2001), has been replicated using other types of measurements, including proteomics (Welsh et al., 2003), and is starting to enter clinical trials (Potti et al., 2006).

Therapeutics

Bioinformatics, coupled with the exponential growth in data collected using high-bandwidth measurement modalities, has the potential to revolutionize the way patients are treated with pharmaceuticals. Much has already been written on how high-bandwidth measurements have generally altered the process of drug discovery (Debouck and Goodfellow, 1999). Here, we will cover specific case examples on how the conceptualization of pharmacology can be radically changed through development and use of biomedical informatics methods.

In 2000, Butte and colleagues focused on generating hypotheses of functional relationships between pairs of genes and pharmaceuticals, with the hope that specific genes could be found that altered a cancer's ability to respond to specific drugs. To do this, they started with two databases, both related to the NCI60, a standardized set of 60 human cancer cell lines used by the National Cancer Institute Developmental Therapeutics Program (Weinstein et al., 1997). One database contained the baseline microarray measurements of the expression level of 6701 genes measured in these cell lines. The other database contained a set of drug susceptibility measurements for the same cell lines, across

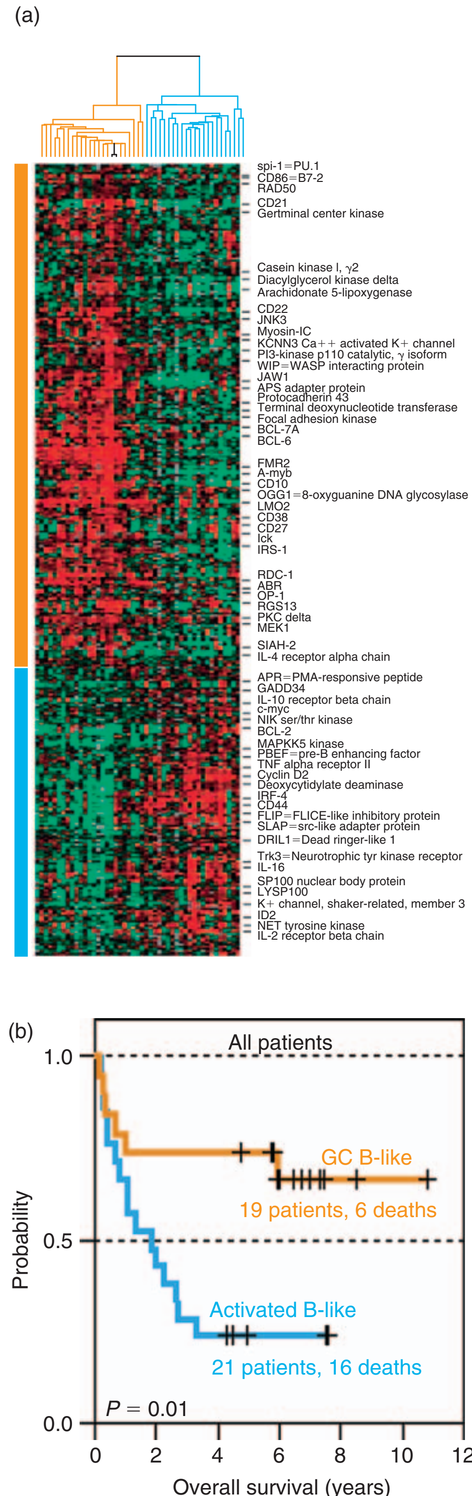


Figure 10.2 (a) Ash Alizadeh and colleagues measured genome-wide expression levels in samples from patients with diffuse large B-cell lymphoma and found samples clustered into two sub-types based on expression patterns. (b) Patients in these two sub-types have a significant difference in survival (Alizadeh et al., 2000). Reprinted by permission from Macmillan Publishers Ltd: *Nature* 403:503, copyright 2000.

nearly 5000 anti-cancer agents. They then used a database technique known as *join* to combine these two databases through the cell lines in common. After joining the two tables, they searched for baseline RNA expression levels in the cell lines that correlated with the inhibition of growth of these same cell lines by thousands of anti-cancer agents. They found only one network containing an association between a gene expression and a measure of anti-cancer agent susceptibility (Figure 10.3). The association suggested that increased expression of lymphocyte cytosolic protein-1 (*LCP1*) is associated with increased susceptibility to the anti-cancer agent NSC 624044, a thiazolidine carboxylic acid derivative. Though a specific role for *LCP1* in tumorogenicity had been postulated and though other thiazolidine carboxylic acid derivatives were known to inhibit tumor cell growth, there was no known relationship between this specific anti-cancer agent and gene in the biomedical literature. The significance of this work was that drug efficacy measurements could be joined to microarray measurements and potentially other types of genomic measurements, enabling the discovery of hidden associations between drugs and genes.

Fliri and colleagues joined databases in a similar manner in 2005, but instead of joining using cell lines, they used the drugs themselves (Fliri et al., 2005). Specifically, they took a database of known side-effects of drugs and joined it to a database of ligand binding assay results as measured in the presence of those same drugs. With this method, they were able to show that drugs with similar side-effect profiles actually had similar ligand-binding assays. The significance of this finding is that pre-clinical studies, such as ligand-binding assays and others, could be used to predict clinical side-effects, which, when severe enough, can lead to withdrawal of a drug from the market.

Lamb and colleagues in 2006 built a database from microarrays measured across the genome after the application of 164 different small molecules to a breast cancer cell line (Lamb et al., 2006). They then queried this database of drug responses using responses measured from other experiments for which a detailed

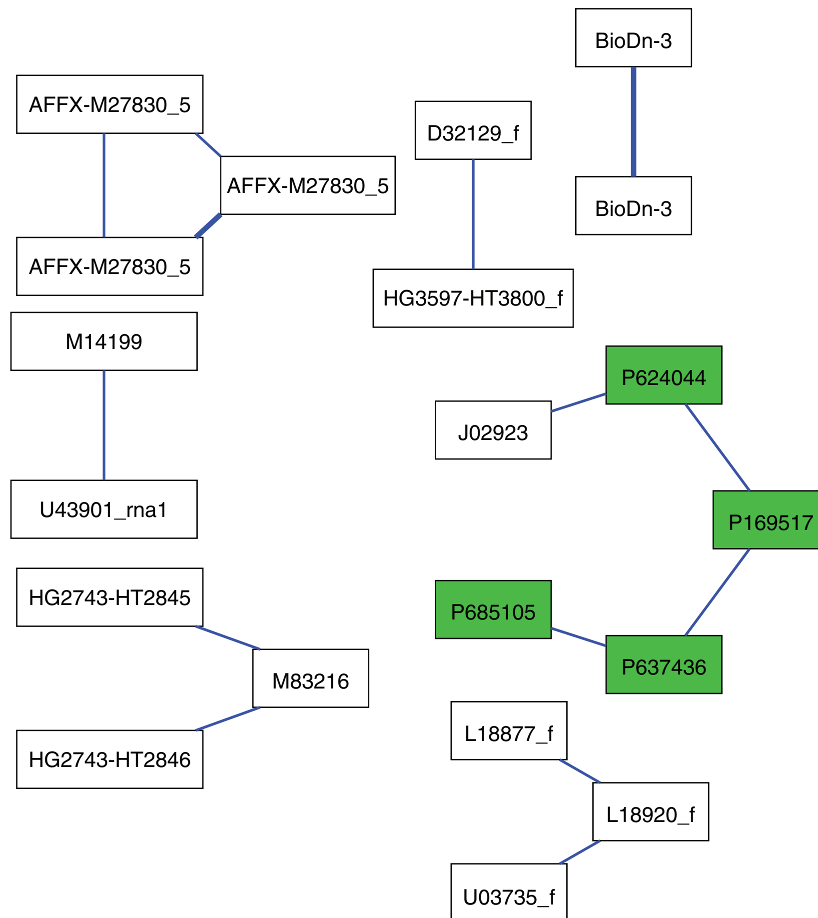


Figure 10.3 Relevance networks calculated by joining drug activity data with gene microarray measurements. White rectangles indicate genes, and green rectangles indicate anti-cancer agents. The only discovered network involving both genes and drugs is shown in the middle. The activity of four anti-cancer agents against multiple cancer cell lines correlates with the expression pattern of a single gene across those same cancer cell lines. *J02923* indicates the gene *LCP1*, while drug identifiers correspond to National Cancer Institute NSC identifiers (Butte et al., 2000).

mechanism had not been known, using a measure of similarity known as a *Kolmogorov-Smirnov statistic* (Figure 10.4). For example, when queried with the gene expression differences between dexamethasone-treatable and dexamethasone-resistant ALL, they found that the genome-wide difference between these two types of leukemia most closely matched that of the genome-wide response to the drug sirolimus. The significance of this approach is that existing drugs could be associated with unexplained disease-related processes, and might even be used to treat those diseases. But most importantly, this work highlights the importance of building structured searchable repositories of experimental data.

Histopathology

Histopathology is the study of diseased tissue by sectioning, staining, and multi-resolution microscopy. Given high-bandwidth measurements from samples of potentially diseased tissue, bioinformatics methods can be applied to determine whether such samples do quantitatively resemble a disease (van de Rijn and Gilks, 2004). For example, Ramaswamy and colleagues tackled the problem of finding the original source of a cancer, given just a metastatic sample (Ramaswamy et al., 2001). They used a method called *support vector machines*, each of which was designed to recognize one specific type of cancer (Figure 10.5). This work was significant in that it was one of the first to show how multiple types of cancers could be distinguished using high-bandwidth measurements, and that unknown samples could be accurately assigned a primary diagnosis. Similar work has been published using alternative bioinformatics methods (Su et al., 2001).

Fibroblasts are cells in the connective tissue that create much of the extracellular matrix between tissues and proliferate during wound healing. Though fibroblasts from different parts of the body may have similar properties and resemble each other under a microscope, Chang and colleagues demonstrated that when fibroblasts are considered by high-bandwidth measurements of gene expression, significant differences between fibroblasts are seen depending on where in the body they were originally obtained. In other words, cells that appear exactly the same by microscope appear quite differently by microarray, consistent with their original source (Chang et al., 2002).

Microarrays have also been used to uncover histopathological findings that might have otherwise been missed. Sarwal and colleagues used microarrays to study the process of acute rejection after kidney transplantation (Sarwal et al., 2003). After *hierarchically clustering* their samples, they noticed a set of B-cell specific genes that corresponded to one sub-type of acute rejection. This was particularly interesting because B-cells had not previously been thought to be involved in acute rejection. Staining for a B-cell specific cell surface protein showed large aggregations of B-cells in samples of acute rejection. The significance of this finding was the realization that while histochemistry can offer a broad survey of the cellular makeup in a sample, these cells will still remain invisible unless illuminated with the right stain. The proper stain can be found using gene expression microarrays coupled with the right bioinformatics techniques, especially the prior annotation of genes as being specific for a certain cell type.

Nosology

In the mid 1700s, Carl Linnaeus promoted the binomial nomenclature to classify living things into a hierarchy, or taxonomy. Though we still organize species into taxonomies based on anatomic and physiological similarities, our modern-day use of DNA sequencing and genome sequence has enabled the reorganization of the position of species within established taxonomical trees.

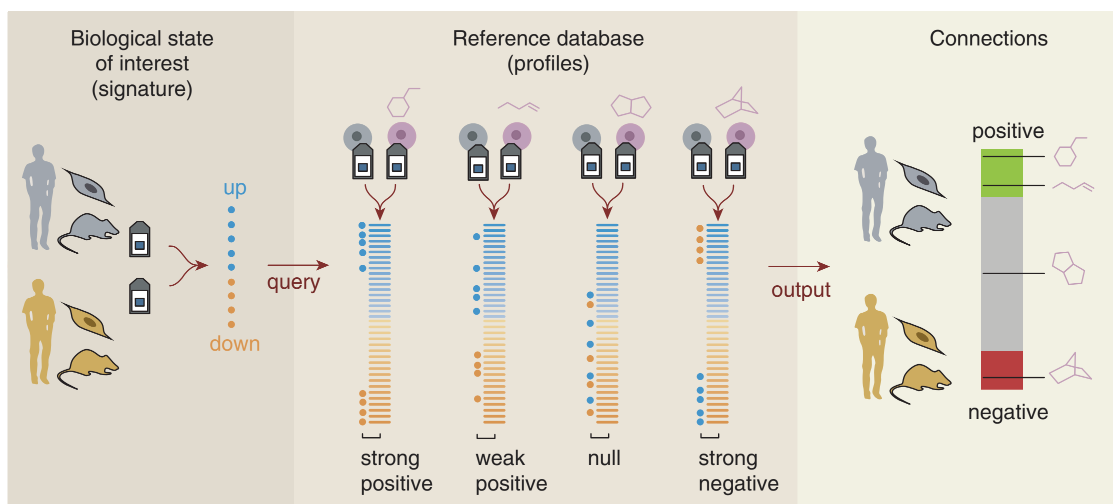


Figure 10.4 A reference database of high-bandwidth molecular measurements can be made after the application of various drugs, then queried with an unknown biological response, yielding those drugs sharing a similar response. From Lamb et al., 2006, *Science* 313:1929. Reprinted by permission from AAAS.

This is because DNA sequence similarity analysis provides a quantitative measurement of how different two species are (Baldauf et al., 2000). Linnaeus was also a co-founder of systematic nosology, or the classification of disease, for which he is not as well known (Linnâe and Schrèoder, 1763). Genomic data and bioinformatics methods have now advanced to the point that we can begin to modernize the classification of disease itself, similar to how DNA sequencing has modernized taxonomy.

Crescenzi and Giuliani were one of the first to demonstrate this principle in 2001 (Crescenzi and Giuliani, 2001). They gathered samples representative of 60 cancer cell lines and used a statistical method known as *principal components analysis* to find combinations of genes that capture the significance variance across cancers. They then showed how analyzes using just the top five principal components could reproduce the findings from the entire dataset. In other words, Crescenzi and Giuliani were able to show that when measured using high-bandwidth measurements, the 60 varied types of cancers differed in only a few ways that could be modeled.

Instead of modeling the significant differences between cancers, Rhodes and colleagues searched for commonalities in 2004 (Rhodes et al., 2004). After collecting 40 published microarray datasets totaling more than 3700 samples of cancer, Rhodes determined a genome-wide signature representative of neoplastic transformation. As expected, most of the genes in this cancer signature were known to participate in the cell cycle, the biological process involved in cell replication. A similar study by Bild and colleagues found commonly deregulated biological pathways across cancers that correspond with worsening survival (Bild et al., 2006). This type of work is significant because it suggests that a quantitative classification of cancers, or one based on actual measurements, might be more useful than the traditional anatomic or even histopathological classification of cancer (Figure 10.6).

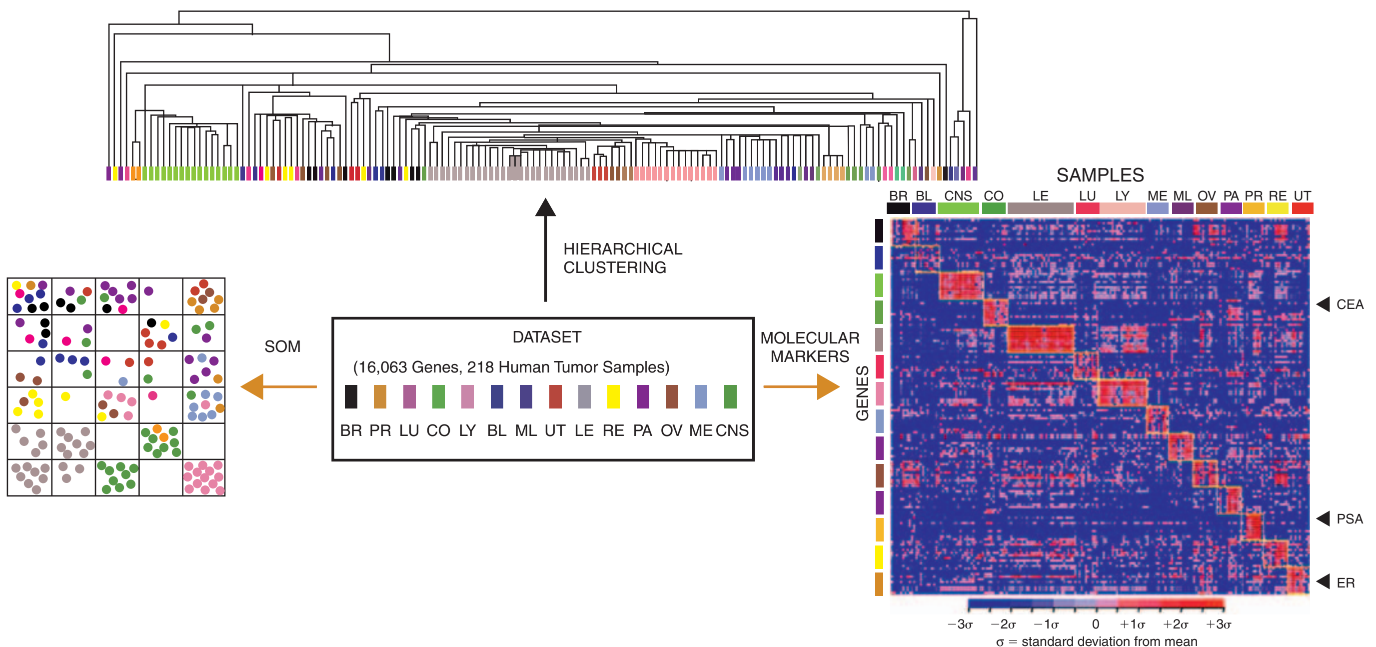


Figure 10.5 Over 200 samples of 14 different types of tumors were clustered to indicate similarities in tumor types and then classified to find genes uniquely implicating each type of tumor (Ramaswamy et al., 2001).

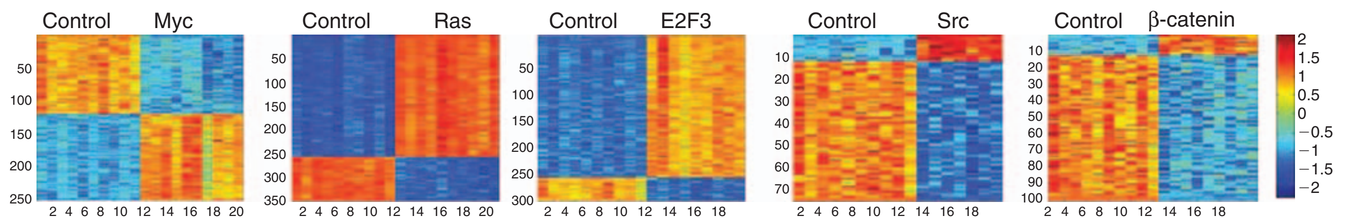


Figure 10.6 Cells made to over-express particular oncogenes were measured using gene expression microarrays, and a "signature" or pattern of gene expression changes for each oncogene was then computed and used to probe samples of cancer (Bild et al., 2006). Reprinted by permission from Macmillan Publishers Ltd: *Nature* 439:353, copyright 2006.

ANALYTIC METHODS

The field of computer science involved in the search for patterns and the induction of rules from data is called *machine learning*. Current methodologies to analyze high-bandwidth measurements can broadly be divided into two categories: *supervised machine learning* approaches, or analysis to determine genes or proteins that fit a predetermined pattern; and *unsupervised machine learning* approaches, or analysis to characterize of the components of a dataset, without the *a priori* input or knowledge of a right answer (Butte, 2002).

Supervised machine learning methods are commonly used to find those individual or sets of measured elements with measurements that (1) are significantly different between defined groups of samples, and (2) accurately predict a characteristic of the sample. An example of an application of a supervised machine learning method is the vignette described above in which Golub and colleagues found genes that distinguish ALL and AML.

There are many published supervised methods that find measured elements that accurately predict characteristics of samples, such as distinguishing one type of disease from another, or a malignant disease from a benign one (Dudoit et al., 2002). There are methods that find individual measured elements, such as the nearest-neighbor approach (Golub et al., 1999), and those that find sets of multiple elements, such as decision trees (Quinlan, 1992), neural networks (Rumelhart et al., 1986), and support vector machines (Brown et al., 2000; Chow et al., 2001; Furey et al., 2000).

Unsupervised machine learning methods are used to find internal structure or relationships within a dataset, instead of trying to determine how best to predict a "correct" answer. Two of the example vignettes provided above – in which Crescenzi determined those genes that best explained the differences between cancers and in which Alizadeh found sub-types of B-cell lymphoma – are good examples of the use of unsupervised machine learning methods.

Within unsupervised learning, there are three classes of techniques, including (1) measured element discovery, or finding elements (e.g., genes) with interesting properties without specifically searching for a specific predefined pattern, such as by principal component analysis (Alter et al., 2000; Fiehn et al., 2000; Hilsenbeck et al., 1999; Raychaudhuri et al., 2000; Wen et al., 1998); (2) cluster discovery, or finding groups of measured features or samples with similar patterns, such as by nearest-neighbor clustering (Ben-Dor et al., 2000; Golub et al., 1999), self-organizing maps (Tamayo et al., 1999; Toronen et al., 1999), k-means clustering, and one and two-dimensional hierarchical clustering (Eisen et al., 1998; Ross et al., 2000); and (3) network discovery, or finding graphs representing associations between measured elements, such as by using Boolean networks (Liang et al., 1998; Szallasi and Liang, 1998; Wuensche, 1998), Bayesian networks (Friedman et al., 2000), and relevance networks (Butte and Kohane, 1999; Butte and Kohane, 2000; Butte et al., 2000).

Several recent articles have reviewed many of the available methods (Butte, 2002; Quackenbush, 2004, 2006). Here we will consider two of the most commonly used supervised and unsupervised methods used in those publications with the greatest impact for medicine.

Case-Control Studies

The majority of experiments performed using high-bandwidth measurements still typically use only a handful of cases in two or three conditions, and the experimenter typically has the goal of finding those measured elements (e.g., genes) that are significantly different between these groups. Significance has and continues to be evaluated in a multitude of different methods, including parametric (Tusher et al., 2001), non-parametric (Butte et al., 2001), Bayesian (Baldi and Long, 2001), and many others. Analysis of variance has been used for measurements from samples with multiple characteristics (Pavlidis and Noble, 2001). Signal processing (Butte et al., 2002), trend detection (Tseng et al., 2005), Bayesian (Ramoni et al., 2002), non-parametric (Reis et al., 2001), and many other methods have been used for measurements from samples across time (Simon et al., 2005).

The challenge in using any of these methods across the set of measured elements (e.g., genes) is compensating for the high number of measurements made. For example, even a standard t-test could be used on every gene measured in two sets of microarrays from two conditions. The null hypothesis for a particular gene would be that the gene expression level is not significantly different across the two conditions, and a p-value can be calculated to assess the ability to reject that null hypothesis for each gene. But in such a case, it would be incorrect to use the "standard" threshold of $p \leq 0.05$ to determine those genes that are significantly different. Traditionally, such a threshold indicates the likelihood of calling a truly negative finding positive, otherwise known as the *false positive rate* or a type 1 error in statistics. For a single test, a 5% false positive rate may not be unreasonable, but when multiplied across 40,000 transcripts, such a false positive rate would be expected to yield 2000 falsely positive genes, too many to validate.

With high-bandwidth measurements, one is typically more interested in the *false discovery rate*, rather than the false positive rate. The false discovery rate is the expected proportion of accepted type 1 errors within a list of tested hypotheses, and controls for the number of tests being made. A false discovery rate can actually be calculated for every measured element. In essence, such a rate would indicate the likelihood of false discoveries if that element were considered positive. In practice, this is called the q-value for that element (Storey and Tibshirani, 2003). Those elements with low p-values also have low q-values.

Though q-values and false discovery rates are gaining in popularity, it is important to note that there have been many other proposed methods to compensate for multiple-hypothesis testing (Dudoit et al., 2003b).

Sub-Type Discovery

The discovery of disease sub-types using high-bandwidth measurements has been the prime motivator for personalized disease

therapeutics, where a sample of disease from an individual could be potentially measured and the most appropriate set of therapies potentially prescribed.

Typically, the discovery of sub-types involves a two-step process. First, an unsupervised machine learning method is applied yielding subsets of samples (e.g., patients) with similar measurements. Many choices are available in this step, including the specific clustering method used (e.g., hierarchical clustering (Eisen et al., 1998), self-organizing maps (Tamayo et al., 1999), or others), the mathematical definition of similarity used (e.g., Pearson correlation coefficient, Spearman rank correlation, Euclidean distance, or others), and the threshold similarity at which clusters are demarcated.

While this first step will yield potential sub-types of disease, the utility of these sub-types still needs to be determined. Typically, utility is defined using a marker of clinical outcome, such as survival, disease-free survival, or response duration. The two or more sub-types of disease identified in the first step are then tested statistically against the marker, using a Kaplan–Meier estimator or Cox Proportional Hazards Model.

The most important point to note with these types of analyzes is the censored nature of the samples. A particular patient that is lost to follow-up can only "contribute" his or her survival up to the length of follow-up. Recent methodological work has provided numerous new methods that combine both of these steps, enabling the single-step discovery of sets of measured elements (e.g., genes) that best discriminate for censored survival or other outcomes (Bair and Tibshirani, 2004; Nguyen and Rocke, 2002; Park et al., 2002).

WHERE DATA FOR STUDIES MAY BE FOUND

Corresponding with the successful application of high-bandwidth molecular measurement modalities, the amount of data in international repositories has grown exponentially, because top-tier journals require the public availability of such data and because of increased calls for transparency of raw data in publications (Nature, 2002; Ball et al., 2004; Perou, 2001). For example, the Gene Expression Omnibus (GEO) is an international repository for gene expression data, developed and maintained by the National Library of Medicine (Wheeler et al., 2004). As of this writing, GEO holds 103,000 samples (i.e., microarrays) from over 4300 experiments involving 189 species, across over 2100 types of microarrays, with 2.6 billion individual gene measurements (Figure 10.7). More impressively, GEO has been gaining data at 300% per year.

GEO is not the only international repository for microarray data. ArrayExpress is a similar database of gene expression measurements supported by the European Bioinformatics Institute (EBI) (Parkinson et al., 2005). As of this writing, ArrayExpress contains 42,424 samples from 1437 experiments. The Stanford Microarray Database (SMD) was the first international repository

Figure 10.7 Screen shot of the home page for the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO).

for microarray data (Ball et al., 2005). As of this writing, SMD holds 61,573 samples involving 37 species, with a total of 1.84 billion individual gene measurements. It is not clear how much overlap there is in experiments with GEO, but it is likely that over 150,000 unique samples are available for study.

Recently, international repositories for proteomic experimental datasets have been instituted (Prince et al., 2004). PeptideAtlas, supported by the Institute for Systems Biology (ISB), is an example of one such repository (Desiere et al., 2006; Deutsch et al., 2005). As of this writing, PeptideAtlas holds the raw and processed mass spectra for 76 experiments. The Proteomics Identifications Database (PRIDE), supported by EBI, currently holds data from over 1655 experiments (Jones et al., 2006).

Though GEO, ArrayExpress, SMD, and PeptideAtlas are already incredible resources for gene expression measurements, these measurements are poorly indexed. Even for microarray data stored in standardized formats like MIAME and MAGE-ML (Brazma et al., 2001; Spellman et al., 2002), the contextual annotations needed to determine relevance of a dataset are unfortunately still represented by unstructured narrative text. Determining the phenotypes, diseases, and environmental contexts studied by these experiments is no longer a tractable manual process. Most repository websites provide only a free-text based search facility for experimental annotations.

A case example demonstrates the significance of the problem. At the time of this writing, searching on the NCBI GEO website yields 27 datasets annotated with the term "breast cancer." However, a search for "breast tumor" yields only 8 datasets. A search for the plural, "breast tumors," yields 9 datasets, while a search for "breast carcinoma" yields only 6 datasets. Breast cancer is one of the most important cancers affecting women, but searching for existing microarray datasets relevant to this important disease yields arbitrary results and is fraught with false negatives. The basic problem is that data in these repositories are not consistently annotated with terms from a standardized vocabulary. While some initial work has been done in re-annotating these datasets using a standardized vocabulary (Butte and Kohane, 2006), investigators interested in finding relevant data will have to try multiple terms to comprehensively find data.

Another caveat is that much of the most relevant clinical data associated with high-bandwidth measurements may not be stored in international repositories. For example, a dataset of microarray measurements on patients with varying lengths of survival after cancer diagnosis may be available in GEO, but the survival times may not be. Such clinical data may be available on an investigator's laboratory website or in an associated publication, but is often not publicly available at all.

BIOINFORMATICS VOCABULARIES AND ONTOLOGIES

A cell biologist interested in finding clinical samples relevant to a disease being studied in the lab will not find those samples unless she can "speak the same language" as the pathologists storing and indexing the samples. A physician interested in learning more about genes that are significantly different in his experiments will not be able to learn more about the pathways implicated by those genes, unless he can "speak the same language" as those geneticists providing annotative information for those genes. In the application of bioinformatics to genomic and personalized medicine, it is often the case that shifting focus like this crucially depends on the use of a standardized vocabulary. The formal modeling and representation of concepts of knowledge, the terms and other attributes used to describe these concepts, and the relationships of these concepts, is called a *ontology*. The importance of using proper ontologies in biomedical research has been described in recent publications (Blake, 2004; Soldatova and King, 2005). Here, we will describe some of the more commonly used ontologies used in the research of genomic medicine (Table 10.2).

Clinical vocabularies

The largest disease ontology (or nosology) in use today is the Systemized Nomenclature of Medicine-Clinical Term (SNOMED-CT) ontology. SNOMED-CT has a lineage spanning 75 years and is used by pathologists world-wide (Chute, 2000). SNOMED-CT has over 340,000 biomedical concepts in 18 hierarchies with 1.3 million relationships that classify diseases based on syndromic and pathophysiological mechanisms (Cho et al., 1998). For example, relations in SNOMED-CT assert that *clear cell carcinoma of the kidney* is a *malignant tumor of the kidney* (i.e., a type of tumor of a particular organ) and separately is also

TABLE 10.2 Some of the ontologies and vocabularies in use in bioinformatics

Ontology and vocabulary	Use
Systemized Nomenclature of Medicine-Clinical Term (SNOMED-CT)	Pathological based classification of disease
International Classification of Diseases (ICD) with Clinical Modification	Diagnosis codes for billing, public health, and epidemiology
Gene Ontology (GO)	Molecular functions, biological processes, and cellular components of proteins
Open Biomedical Ontologies (OBO)	Includes Gene Ontology and many other ontologies describing cell types and anatomy
InterDom	Protein–protein interactions
InterPro	Protein domains
Unified Medical Language System (UMLS)	Unified compendium of 140 biomedical vocabularies

a *malignant neoplasm of the retroperitoneum* (i.e., a type of tumor in a particular anatomical location).

Another disease classification more familiar to physicians is the International Classification of Diseases (ICD) with Clinical Modification, a nosology with origins in the 1850s. ICD-9-CM has nearly 7000 codes, while ICD-10-CM has over 14,000 codes to describe causes of morbidity and mortality (World Health Organization, 2005). Both ICD-9-CM and ICD-10-CM are in use by various health care providers, the US Federal Government, and the World Health Organization. Though most health care institutions in the United States use ICD-9, ICD-10 was recently approved for use by the US Congress. To continue the example above, relations in ICD-9-CM assert that a *malignant neoplasm of the kidney* is a *malignant neoplasm of the genitourinary organs*. Importantly, however, the limited resolution in ICD-9-CM does not permit the specific diagnosis of *clear cell carcinoma of the kidney*.

There are several important differences between the SNOMED-CT and the ICD nosologies. As shown above, SNOMED-CT has greater expressiveness than ICD to represent diseases. SNOMED provides a rich set of relationships between medical concepts. ICD provides only a hierarchical relationship between concepts.

Despite their differences, it is important to note that both SNOMED-CT and ICD are viewed as administrative nosologies, typically used for billing, reporting, electronic medical records, and decision support systems. However, important sources of stored tissues in medical institutions, such as frozen tissue banks, may only be accessible through SNOMED or ICD codes, and these coding systems commonly serve as an index into these repositories.

Genomic Ontologies

The Gene Ontology (GO) is a taxonomy that is used to describe the normal *molecular* function of proteins, the *cellular components* in which proteins operate, and the larger *biological processes* in which they participate (Ashburner et al., 2000). At the time of this writing, GO currently holds 7470 molecular functions, 1823 cellular components, and 12,250 biological processes. Over 186,000 genes across a variety of species have been assigned to one or more GO categories, but this essentially translates to one-third of human genes having some coded function.

The easiest way to find the GO classification for a gene is to execute a query using Entrez Gene. For example, a query for the gene coding for the insulin receptor (INSR) yields 13 molecular functions (e.g., *phosphoinositide 3-kinase binding*), 7 biological processes (e.g., *carbohydrate metabolism*), and 2 cellular components. It is important to note that an actual publication detailing the evidence for each annotation is available for many GO annotations, and these publications can be easily referenced. It is also important to note that the GO annotations for genes can change quickly, as both the annotations for genes and GO itself continue to be "works in progress."

The GO is actually one of 61 ontologies in the set known as Open Biomedical Ontologies (OBO) (Mungall, 2004). These ontologies are open, in the sense that they can be freely downloaded and used without constraints. Other ontologies in OBO include ontologies to describe the gross anatomy of flies, worms, and mice, cell types, and functional portions of proteins, as well as ontologies to describe cancers and other concepts related to humans.

Other vocabularies used for gene and protein annotation include InterDom (Ng et al., 2003), InterPro (Mulder et al., 2003), and PRINTS (Attwood et al., 2003) for protein domains, and other gene identifiers include the stable NCBI Gene identifier (Wheeler et al., 2006) and the unstable NCBI UniGene identifier (Wheeler et al., 2000).

Unified Medical Language System

The Unified Medical Language System (UMLS) is the largest available compendium of biomedical vocabularies, containing 140 biomedical vocabularies with over 1.2 million concepts and 41 million relations between concepts (Bodenreider, 2004; Butte and Kohane, 2006). UMLS already unifies vocabularies used extensively in molecular biology and genomics, such as the Medical Subject Headings (MeSH), NCBI Taxonomy, and the GO, with medical vocabularies including the International Classification of Diseases and SNOMED Clinical Terms (Ashburner et al., 2000; ICD, 2003; Wheeler et al., 2004). The UMLS is a unified vocabulary, which means that concepts listed in multiple vocabularies are brought together. For example, the concept of *fever*, which holds the UMLS concept unique identifier of *C0015967*, is represented in 78 component vocabularies, such as *D005334* in MeSH, *386661006* and *50177009* in SNOMED-CT, *780.6* in ICD-9-CM, *10016558* in the Medical Dictionary for Regulatory Activities Terminology (MedDRA), *X25* in Perioperative Nursing DataSet, *GO:0001660* in GO, and *U001776* in the Library of Congress Subject Headings. Each of these individual codes represents the same concept, and thus have been unified to a single biomedical concept in UMLS.

Because of its ability to span nearly every other relevant vocabulary, UMLS best serves as a bridging vocabulary, providing terms commonly used by both physicians as well as molecular biologists. Given the support for UMLS and its component vocabularies shown by the National Library of Medicine and the US Department of Health and Human Services, it is likely that terms and concepts from UMLS will constitute important portions of the future electronic health records of patients in the United States (see also Chapter 12). As samples from patients continue to be valuable for study using high-bandwidth measurements, tapping into the medical records associated with these samples will only grow in importance and UMLS can help enable these secondary research uses of medical records.

FREELY AVAILABLE BIOINFORMATICS TOOLS

There is an incredible preponderance of software tools freely available for research in bioinformatics. As an example, as of this writing, over 300 articles are published each year specifically describing bioinformatics software tools in the journals *BMC Bioinformatics*, *Bioinformatics*, *Nucleic Acids Research*, and *Genome Biology*. The vast majority of these are freely available for

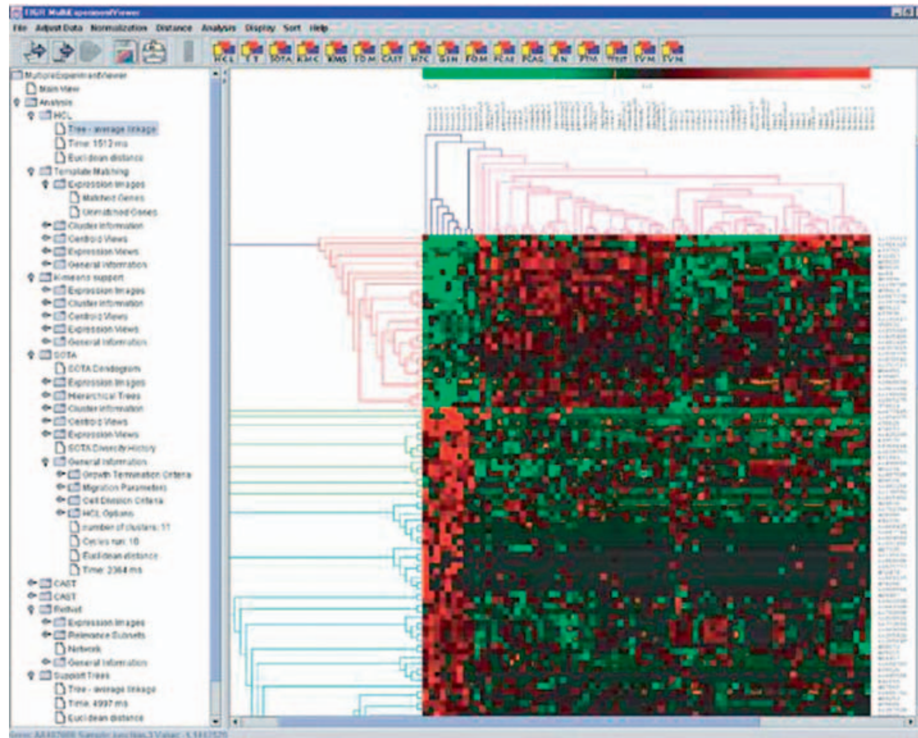


Figure 10.8 Screen shot of the MeV (Saeed et al., 2006). Reprinted from Methods in Enzymology, Volume 411, Saeed, A.I., Bhagabati, N.K., Braisted, J.C., Liang, W., Sharov, V., Howe, E.A., Li, J., "TM4 microarray software suite", 134–93, copyright 2006, with permission from Elsevier.

researchers. Describing all of the available free tools is clearly intractable, though a recent book and paper serve as nice collections of most of the commonly used tools (Dudoit et al., 2003a; Parmigiani, 2003). Here instead, we will arbitrarily focus on a few important tools useful in the study of genomic and personalized medicine, dividing our set into two groups: those useful in the analysis of high-bandwidth data and those useful in the interpretation of results.

Analytic Tools

The Multiexperiment Viewer (MeV) is an easy-to-use analytic tool initially developed at the Institute for Genomic Research and currently maintained at the Dana Farber Cancer Institute (Saeed et al., 2006). It specifically supports 17 tools for the analysis of gene expression microarrays, and presents these tools using easy-to-use on screen buttons (Figure 10.8). It runs under the Macintosh, Windows, and Linux operating systems, and the programming source code is available and can be modified by bioinformatics programmers. At the time of this writing, MeV has been cited by nearly 200 publications and has been extensively documented.

GenePattern is a sophisticated analytic tool developed and maintained by the Broad Institute (Reich et al., 2006). GenePattern supports the analysis of multiple types of high-bandwidth measurements, including gene expression microarrays, proteomics, and single nucleotide polymorphisms (SNP). Each analytic and visualization tool is contained within a module, which can be chained together to represent a reproducible pipeline for analysis (Figure 10.9). GenePattern runs under the Macintosh, Windows, and Linux operating systems. GenePattern is well-documented and supported, and as of this writing, approximately 30 papers have cited its use.

GenePattern clearly excels in its advanced features. One particularly useful feature of GenePattern is that it offers a modular architecture allowing users to plug in new analytic tools or visualization methods as they become available. These tools can be created to interface with other bioinformatics analytic tools, including Matlab and R and Bioconductor (described below). GenePattern can distribute the work required for an analysis onto a parallel computer cluster, which, if available, can speed the time required to complete an analysis. In addition, GenePattern supports the output of measurements into the standardized formats necessary for submission to international repositories, a step that is increasingly required before a top-tier journal accepts a manuscript.

At the opposite extreme of user-friendliness lies the statistical computing and graphics environment known as R (R Development Core Team, 2004). Based on the S statistical computing environment developed by Bell Labs in the late 1970s, R is the most flexible of the tools listed here. R is almost essentially "command line" driven, in that commands to execute statistical and visualization tools are typed and interpreted serially, but most users write scripts and sophisticated programs that can be run to generate the appropriate output and figures. There is essentially no user-interface (i.e., menus and windows) that can be used to perform analyzes; getting up-to-speed with R involves climbing a steep learning curve.

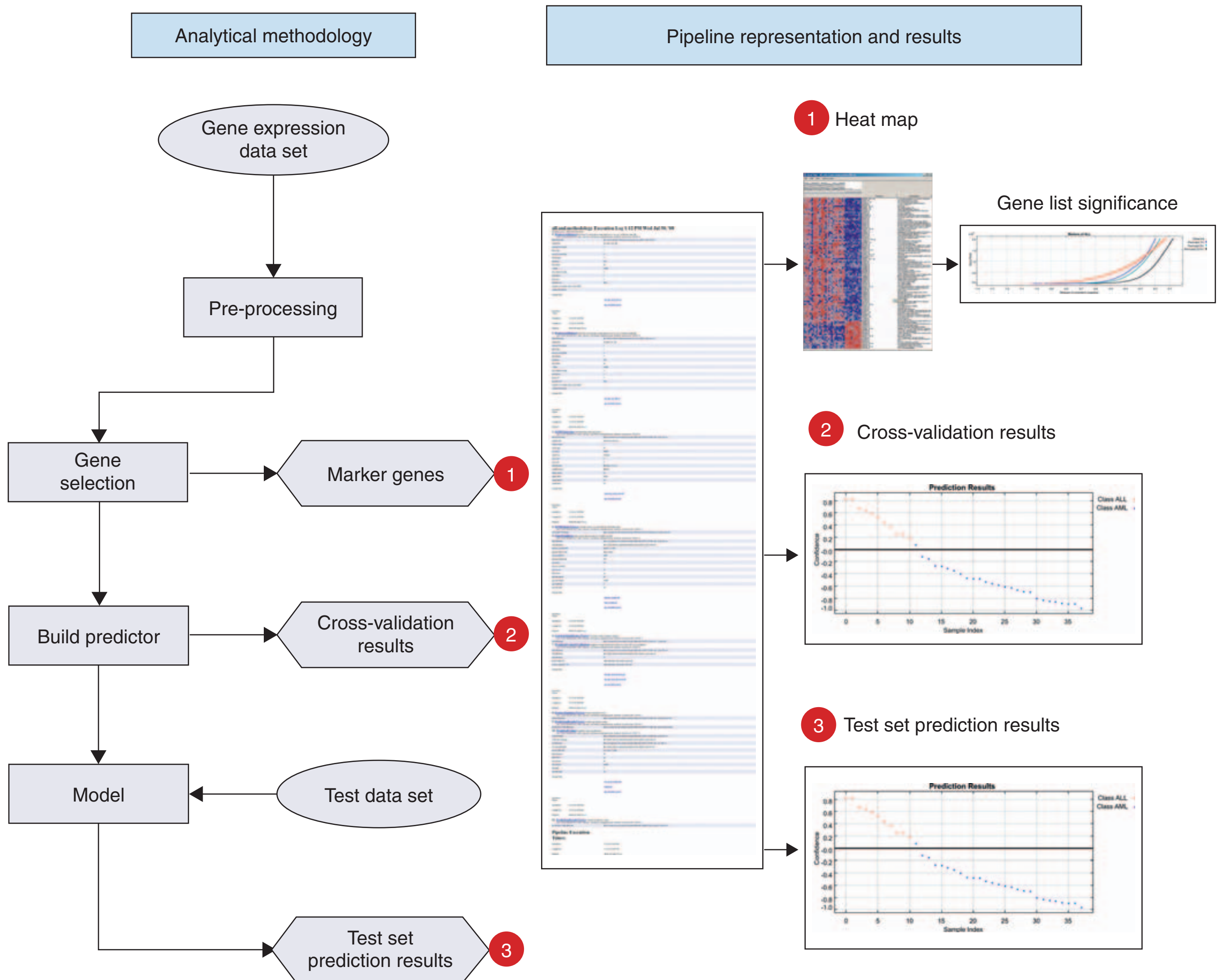


Figure 10.9 The conceptual analytic "pipeline" needed to analyze high-bandwidth molecular data is represented on the left, while its computational equivalent is represented on the right. Using GenePattern, these steps can be coded and documented, yielding reproducible research results (Reich et al., 2006). Reprinted by permission from Macmillan Publishers Ltd, *Nature Genetics* 38, 500, copyright 2006.

Despite the severe user-interface handicaps, the most important advantage of R is its community of users. Because of its designed similarity to S, users of R can benefit from decades of developed analytic tools (some statistical tool plug-ins for R and S available on the Internet date back to the late 1980s). The most important set of plug-in tools for research in genomic medicine are those contained in the Bioconductor package, which offers tools for reading microarray data, processing raw data files, common supervised and unsupervised machine learning methods, and visualization methods (Gentleman et al., 2004).

More than 50 books have been written that explain both commonly-used and cutting-edge statistical techniques using R (and S), including a few specifically for bioinformatics and Bioconductor (Gentleman et al., 2004; Parmigiani, 2003). R runs under the Macintosh, Windows, and Linux operating systems and its source code is available.

Intepretation

After several analyses, it quickly becomes obvious that the rate-limiting step in experiments involving high-bandwidth measurements is neither the handling of the biological samples nor the actual statistical or numerical work, but instead the post-analytic work in determining what the results actually mean. First, detailed names and knowledge might not yet exist for genes and proteins that have been found to be significantly involved in an experiment, even though these elements may have already been measured for many years. This complicates the interpretation of results. The official gene name, predicted

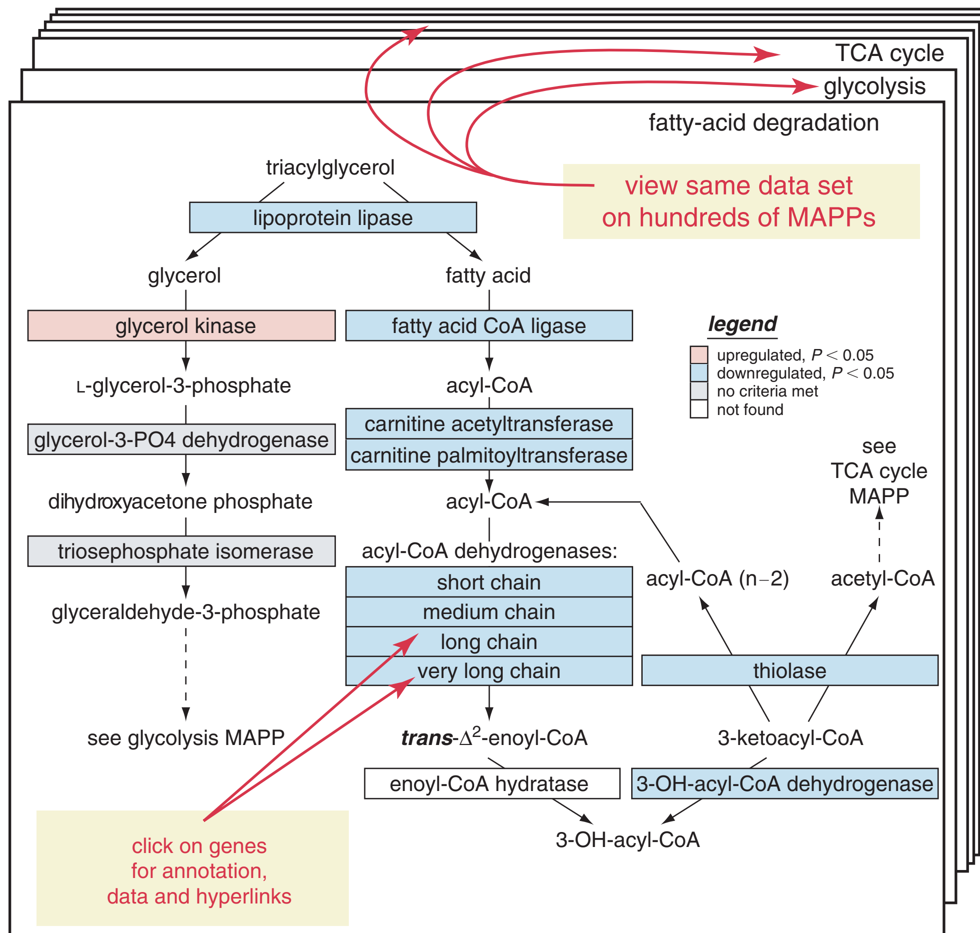


Figure 10.10 GenMAPP has the ability to color-code proteins participating in biological pathways based on measurements (Dahlquist et al., 2002). Reprinted by permission from Macmillan Publishers Ltd, *Nature Genetics* 31, 19, copyright 2002.

protein domains, or GO classification may become available for a gene or protein as early as tomorrow, or as late as years from now. Operationally, this means that one is never done analyzing a set of microarray data. One has to develop the infrastructure to constantly reinvestigate genes and proteins from analyzes performed in the past. It may be next week, for example, that new knowledge about a protein that was significantly implicated in an analysis performed three months ago gets published, finally leading to a novel and important finding.

The challenges in determining the proper analytic methods to use are usually only a short-term difficulty, and typically after the "functional genomics pipeline" has been established, the rate-limiting step shifts to the post-analytic challenges (Kohane et al., 2002). In the future, truly demonstrating a "return on investment" from high-bandwidth molecular measurements depends on taking findings past the measurement stage and integrating them with the rest of the research pipeline, including interpretation and validation. The list of elements resulting from an analysis should not be viewed as an end in itself; its real value only increases as that list moves through biological validation, ranging from the numerical verification of measurement levels with alternative techniques, to the ascertainment of the meaning of the results, such as finding common mechanisms or biological pathways involving the genes or proteins. While tools that link measurement elements back to known biological pathways are still in their infancy, some have been shown to be quite useful.

GenMAPP is a software tool that can visualize of high-bandwidth molecular measurements using pre-drawn diagrams illustrating biological pathways (Dahlquist et al., 2002). Though GenMAPP comes with nearly a thousand useful biological pathways across numerous species, GenMAPP also comes with an editor enabling a biologist or bioinformatician to draw any new pathway (Figure 10.10). Because of this, the GenMAPP user base has developed into a large community of users who share pathways. GenMAPP runs only under the Windows operating system. At the time of this writing, GenMAPP has been cited by nearly 600 publications and has been extensively documented.

The Kyoto Encyclopedia of Genes and Genomes (KEGG) is an extensive web-based database of biological pathways and functions under development for more than a decade (Kanehisa, 1997). Similar to GenMAPP, KEGG offers tools that can color-code diagrams of biological pathways based on quantitative measurements, instead using a web-based interface (Figure 10.11). KEGG offers more than 42,000 biological pathways for nearly 500 species.

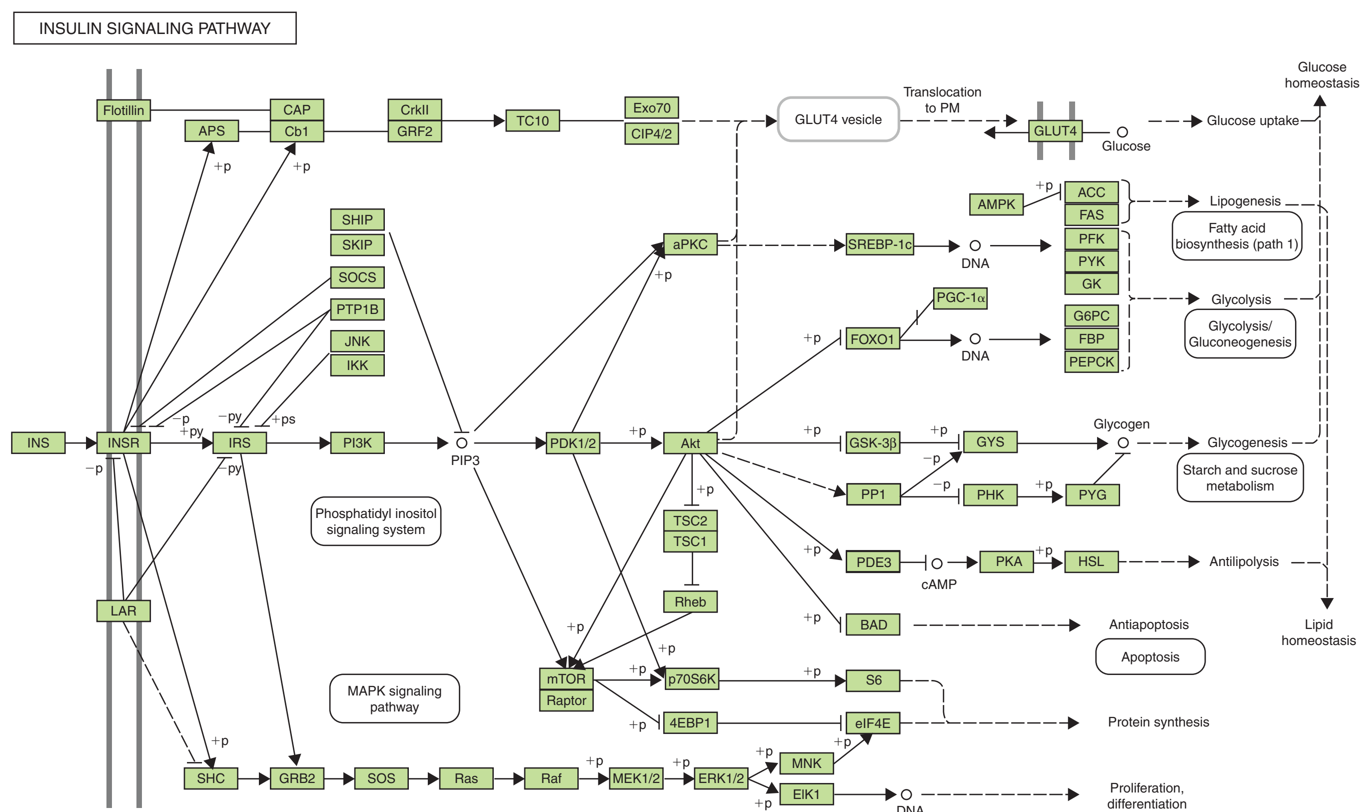


Figure 10.11 The insulin signaling pathway, as represented in KEGG as of this writing. A click on any protein yields additional data about that protein.

Beyond biological pathways, KEGG also provides knowledge on ligands, metabolites, small-molecule drugs, and diseases. At the time of this writing, KEGG has been cited by nearly 1000 publications.

Other free web-based resources available to study genes by known function and protein domains include Entrez Gene and GeneCards (Safran et al., 2002; Wheeler et al., 2006).

2009 UPDATE

There has been astounding growth in biomedical databases over the past two years. The January issue of *Nucleic Acids Research*, in which significant biomedical databases are listed annually, reached 1170 databases in 2009, demonstrating a 16% increase from the previous year (Galperin and Cochrane, 2009). New databases this year include those containing microRNA and their binding sites as well as data and genomes of pathogens causing infectious diseases in humans, including Giardia and Plasmodium.

Individual databases have also demonstrated growth. The NCBI Gene Expression Omnibus, an international repository for gene expression microarray data, has (as of this writing) over 278,000 microarrays from 11,000 experiments, demonstrating 1.4× growth over the past year (Sayers et al., 2009). The EBI ArrayExpress database holds another 100,000 microarrays from 3400 experiments (Parkinson et al., 2009). Even greater growth has been seen in proteomics databases; the EBI PRIDE holds nearly 9000 samples yielding over 17 million mass spectra, showing 6.7-fold growth in the past year. In most cases, the data from these sources are easily and quickly downloaded.

The most significant new source of data is from genome-wide association studies (GWAS) (McCarthy et al., 2008). As of this writing, the NHGRI-curated database at http://genome.gov/gwastudies lists 245 GWAS publications. Typically the list of SNPs with significant association is publicly available for many of these, while the raw genotyping measurements are available for fewer. Genotype and phenotype data on GWAS funded by the United States. NIH are held at the NCBI Database of Genotypes and Phenotypes (dbGAP). As of this writing, dbGAP has grown to include 35 studies (up from 14 one year prior), with nearly 90,000 human samples with genetics and phenotyping. The Framingham Heart Study itself contributes 102 separate substudies. Compared to gene

expression microarray and proteomics data, data on genetics are not as easily downloaded; access to genetics data sometimes requires pre-approval from both local and repository Institutional Review Boards (IRB). Most available genetics data are in the form of measurements from SNP assays, where between 50,000 and 2 million SNPs are measured across individuals. Expectations are high that over the coming years, more genetics data will instead come from high-throughput sequencing, as the capabilities of pyrosequencing, single molecule sequencing, and other technologies develop (Schuster, 2008). However, worries about the re-identifiability of individual study participants from publicly-available GWAS data have led to increased vigilance and regulation to the accessibility of this data (Homer et al., 2008).

Clinical data, including those from electronic health records (see Chapter 12), are just starting to be used in large-scale genetics experiments. A recent GWAS for genetic variants influencing the maintenance dose of warfarin, a commonly-used anti-coagulant, used electronic medical records on bleeding events as part of the phenotyping (Rieder et al., 2005). This type of approach is being built into the research infrastructure at particular medical institutions, where DNA biobanks are linked to phenotypic information from electronic medical records (Roden et al., 2008). As more academic research centers are enabled to build such infrastructures through the NIH Roadmap Institutional Clinical and Translational Science Awards (CTSA), we can expect to see more studies bridging the methodologies clinical informatics and bioinformatics (National Institutes of Health, 2005).

2009 UPDATE REFERENCES

Galperin, M.Y. and Cochrane, G.R. (2009). Nucleic acids research annual database issue and the NAR online molecular biology database collection in 2009. *Nucleic Acids Res* 37(Database issue), D1–D4.

Homer, N., Szelinger, S., Redman, M., Duggan, D., Tembe, W., Muehling, J., Pearson, J.V., Stephan, D.A., Nelson, S.F. and Craig, D.W. (2008). Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genet* 4(8), e1000167.

McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P. and Hirschhorn, J.N. (2008). Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nat Rev Genet* 9(5), 356–369.

National Institutes of Health. (2005). Institutional Clinical and Translational Science Award 2007. Bethesda, MD.

Parkinson, H., Kapushesky, M., Kolesnikov, N., Rustici, G., Shojatalab, M., Abeygunawardena, N., Berube, H., Dylag, M., Emam, I., Farne, A. et al. (2009). ArrayExpress update – From an archive of functional genomics experiments to the atlas of gene expression. *Nucleic Acids Res* 37(Database issue), D868–D72.

Rieder, M.J., Reiner, A.P., Gage, B.F., Nickerson, D.A., Eby, C.S., McLeod, H.L., Blough, D.K., Thummel, K.E., Veenstra, D.L. and Rettie, A.E. (2005). Effect of VKORC1 haplotypes on transcriptional regulation and warfarin dose. *N Engl J Med* 352(22), 2285–2293.

Roden, D.M., Pulley, J.M., Basford, M.A., Bernard, G.R., Clayton, E.W., Balser, J.R. and Masys, D.R. (2008). Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin Pharmacol Ther* 84(3), 362–369.

Sayers, E.W., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Edgar, R., Federhen, S. et al. (2009). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 37(Database issue), D5–D15.

Schuster, S.C. (2008). Next-generation sequencing transforms today's biology. *Nat Methods* 5(1), 16–18.

ACKNOWLEDGEMENTS

The work was supported by grants from the Lucile Packard Foundation for Children's Health, National Library of Medicine (K22 LM008261), National Institute of General Medical Sciences (R01 GM079719), Howard Hughes Medical Institute, and the Pharmaceutical Research and Manufacturers of America Foundation. The author is a scientific advisor to Genstruct, Inc., a computational systems biology company.

REFERENCES

Adida, B. and Kohane, I.S. (2006). GenePING: secure, scalable management of personal genomic data. *BMC Genomics* 7, 93.

Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X. et al. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403, 503–511.

Alter, O., Brown, P.O. and Botstein, D. (2000). Singular value decomposition for genome-wide expression data processing and modeling. *Proc Natl Acad Sci USA* 97, 10101–10106.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T. et al. (2000). Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25, 25–29.

Attwood, T.K., Bradley, P., Flower, D.R., Gaulton, A., Maudling, N., Mitchell, A.L., Moulton, G., Nordle, A., Paine, K., Taylor, P. et al. (2003). PRINTS and its automatic supplement, prePRINTS. *Nucleic Acids Res* 31, 400–402.

Bair, E. and Tibshirani, R. (2004). Semi-supervised methods to predict patient survival from gene expression data. *PLoS Biol* 2, E108.

Baldauf, S.L., Roger, A.J., Wenk-Siefert, I. and Doolittle, W.F. (2000). A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* 290, 972–977.

Baldi, P. and Long, A.D. (2001). A Bayesian framework for the analysis of microarray expression data: regularized *t*-test and statistical inferences of gene changes. *Bioinformatics* 17, 509–519.

Ball, C.A., Brazma, A., Causton, H., Chervitz, S., Edgar, R., Hingamp, P., Matese, J.C., Parkinson, H., Quackenbush, J., Ringwald, M. et al. (2004). Submission of microarray data to public repositories. *PLoS Biol* 2, E317.

Ball, C.A., Awad, I.A., Demeter, J., Gollub, J., Hebert, J.M., Hernandez-Boussard, T., Jin, H., Matese, J.C., Nitzberg, M., Wymore, F. et al. (2005). The Stanford Microarray Database accommodates additional microarray platforms and data formats. *Nucleic Acids Res* 33, D580–D582.

Ben-Dor, A., Bruhn, L., Friedman, N., Nachman, I., Schummer, M. and Yakhini, Z. (2000). Tissue classification with gene expression profiles. *J Comput Biol* 7, 559–583.

Bhattacharjee, A., Richards, W.G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M. et al. (2001). Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci USA* 98, 13790–13795.

Bild, A.H., Yao, G., Chang, J.T., Wang, Q., Potti, A., Chasse, D., Joshi, M.B., Harpole, D., Lancaster, J.M., Berchuck, A. et al. (2006). Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 439, 353–357.

Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A. et al. (2000). Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* 406, 536–540.

Blake, J. (2004). Bio-ontologies-fast and furious. *Nat Biotechnol* 22, 773–774.

Bodenreider, O. (2004). The Unified Medical Language System (UMLS): Integrating biomedical terminology. *Nucleic Acids Res* 32(Database issue), D267–D270.

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C. et al. (2001). Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 29, 365–371.

Brown, M.P., Grundy, W.N., Lin, D., Cristianini, N., Sugnet, C.W., Furey, T.S., Ares, M, Jr. and Haussler, D. (2000). Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc Natl Acad Sci USA* 97, 262–267.

Brunner, H.G. and van Driel, M.A. (2004). From syndrome families to functional genomics. *Nat Rev Genet* 5, 545–551.

Butte, A. (2002). The use and analysis of microarray data. *Nat Rev Drug Discov* 1, 951–960.

Butte, A. and Kohane, I. (1999). Unsupervised knowledge discovery in medical databases using relevance networks. In *Fall Symposium, American Medical Informatics Association* (N. Lorenzi, ed.), Hanley and Belfus, Washington, DC, pp. 711–715.

Butte, A.J. and Kohane, I.S. (2000). Mutual information relevance networks: Functional genomic clustering using pairwise entropy measurements. *Pac Symp Biocomput*, 418–429.

Butte, A.J. and Kohane, I.S. (2006). Creation and implications of a phenome-genome network. *Nat Biotechnol* 24, 55–62.

Butte, A.J., Tamayo, P., Slonim, D., Golub, T.R. and Kohane, I.S. (2000). Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proc Natl Acad Sci USA* 97, 12182–12186.

Butte, A.J., Ye, J., Niederfellner, G., Rett, K., Häring, H., White, M.F. and Kohane, I.S. (2001). Determining significant fold differences in gene expression analysis. *Pac Symp Biocomput*, 6–17.

Butte, A.J., Bao, L., Reis, B.Y., Watkins, T.W. and Kohane, I.S. (2002). Comparing the similarity of time-series gene expression using signal processing metrics. *J Biomed Inform* 34, 396–405.

Centers for Disease Control and Prevention (2004). Genetic testing for breast and ovarian cancer susceptibility: Evaluating direct-to-consumer marketing – Atlanta, Denver, Raleigh-Durham, and Seattle, 2003. *MMWR Morb Mortal Wkly Rep* 53, 603–606.

Chang, H.Y., Chi, J.T., Dudoit, S., Bondre, C., van de Rijn, M., Botstein, D. and Brown, P.O. (2002). Diversity, topographic differentiation, and positional memory in human fibroblasts. *Proc Natl Acad Sci USA* 99, 12877–12882.

Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X.C., Stern, D., Winkler, J., Lockhart, D.J., Morris, M.S. and Fodor, S.P. (1996). Accessing genetic information with high-density DNA arrays. *Science* 274, 610–614.

Cho, R.J., Fromont-Racine, M., Wodicka, L., Feierbach, B., Stearns, T., Legrain, P., Lockhart, D.J. and Davis, R.W. (1998). Parallel analysis of genetic selections using whole genome oligonucleotide arrays. *Proc Natl Acad Sci USA* 95, 3752–3757.

Chow, M.L., Moler, E.J. and Mian, I.S. (2001). Identifying marker genes in transcription profiling data using a mixture of feature relevance experts. *Physiol Genomics* 5, 99–111.

Chute, C.G. (2000). Clinical classification and terminology: Some history and current observations. *J Am Med Inform Assoc* 7, 298–303.

Clayton, T.A., Lindon, J.C., Cloarec, O., Antti, H., Charuel, C., Hanton, G., Provost, J.P., Le Net, J.L., Baker, D., Walley, R.J. et al. (2006). Pharmaco-metabonomic phenotyping and personalized drug treatment. *Nature* 440, 1073–1077.

Crescenzi, M. and Giuliani, A. (2001). The main biological determinants of tumor line taxonomy elucidated by a principal component analysis of microarray data. *FEBS Lett* 507, 114–118.

Dahlquist, K.D., Salomonis, N., Vranizan, K., Lawlor, S.C. and Conklin, B.R. (2002). GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nat Genet* 31, 19–20.

Debouck, C. and Goodfellow, P.N. (1999). DNA microarrays in drug discovery and development. *Nat Genet* 21, 48–50.

DeRisi, J., Penland, L., Brown, P.O., Bittner, M.L., Meltzer, P.S., Ray, M., Chen, Y., Su, Y.A. and Trent, J.M. (1996). Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat Genet* 14, 457–460.

Desiere, F., Deutsch, E.W., King, N.L., Nesvizhskii, A.I., Mallick, P., Eng, J., Chen, S., Eddes, J., Loevenich, S.N. and Aebersold, R. (2006). The PeptideAtlas project. *Nucleic Acids Res* 34, D655–D658.

Deutsch, E.W., Eng, J.K., Zhang, H., King, N.L., Nesvizhskii, A.I., Lin, B., Lee, H., Yi, E.C., Ossola, R. and Aebersold, R. (2005). Human Plasma PeptideAtlas. *Proteomics* 5, 3497–3500.

Dudoit, S., Yang, Y.H., Callow, M.J. and Speed, T.P. (2002). Statistical methods for identifying differentially expressed genes in replicated cDNA Microarray experiments. *Statistica Sinica* 12, 111–139.

Dudoit, S., Gentleman, R.C. and Quackenbush, J. (2003a). Open source software for the analysis of microarray data. *Biotechniques*(Suppl), 45–51.

Dudoit, S., Shaffer, J.P. and Boldrick, J.C. (2003b). Multiple hypothesis testing in microarray experiments. *Statis. Sci.* 18, 71–103.

Edwards, J.S., Ibarra, R.U. and Palsson, B.O. (2001). *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat Biotechnol* 19, 125–130.

Eisen, M.B., Spellman, P.T., Brown, P.O. and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95, 14863–14868.

Espina, V., Mehta, A.I., Winters, M.E., Calvert, V., Wulfkuhle, J., Petricoin, E.F., 3rd, and Liotta, L.A. (2003). Protein microarrays: Molecular profiling technologies for clinical specimens. *Proteomics* 3, 2091–2100.

Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R.N. and Willmitzer, L. (2000). Metabolite profiling for plant functional genomics. *Nat Biotechnol* 18, 1157–1161.

Fliri, A.F., Loging, W.T., Thadeio, P.F. and Volkmann, R.A. (2005). Analysis of drug-induced effect patterns to link structure and side effects of medicines. *Nat Chem Biol* 1, 389–397.

Freimer, N. and Sabatti, C. (2003). The human phenome project. *Nat Genet* 34, 15–21.

Friedman, N., Linial, M., Nachman, I. and Pe'er, D. (2000). Using Bayesian networks to analyze expression data. *J Comput Biol* 7, 601–620.

Furey, T.S., Cristianini, N., Duffy, N., Bednarski, D.W., Schummer, M. and Haussler, D. (2000). Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* 16, 906–914.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J. et al. (2004). Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol* 5, R80.

Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K. and Weissman, J.S. (2003). Global analysis of protein expression in yeast. *Nature* 425, 737–741.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A. et al. (1999). Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* 286, 531–537.

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H. and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 17, 994–999.

Henneman, L., Timmermans, D.R. and Van Der Wal, G. (2006). Public attitudes toward genetic testing: Perceived benefits and objections. *Genet Test* 10, 139–145.

Hestvik, A.L., Hmama, Z. and Av-Gay, Y. (2003). Kinome analysis of host response to mycobacterial infection: A novel technique in proteomics. *Infect Immun* 71, 5514–5522.

Hilsenbeck, S.G., Friedrichs, W.E., Schiff, R., O'Connell, P., Hansen, R.K., Osborne, C.K. and Fuqua, S.A. (1999). Statistical analysis of array expression data as applied to the problem of tamoxifen resistance. *J Natl Cancer Inst* 91, 453–459.

ICD (2003). *International Classification of Diseases: 9th revision, Clinical Modification (ICD-9-CM), 6th edition*. Centers for Medicare & Medicaid Services, Washington DC.

Jessani, N., Liu, Y., Humphrey, M. and Cravatt, B.F. (2002). Enzyme activity profiles of the secreted and membrane proteome that depict cancer cell invasiveness. *Proc Natl Acad Sci USA* 99, 10335–10340.

Johnson, G.C., Esposito, L., Barratt, B.J., Smith, A.N., Heward, J., Di Genova, G., Ueda, H., Cordell, H.J., Eaves, I.A., Dudbridge, F. et al. (2001). Haplotype tagging for the identification of common disease genes. *Nat Genet* 29, 233–237.

Jones, P., Cote, R.G., Martens, L., Quinn, A.F., Taylor, C.F., Derache, W., Hermjakob, H. and Apweiler, R. (2006). PRIDE: A public repository of protein and peptide identifications for the proteomics community. *Nucleic Acids Res* 34, D659–D663.

Kanehisa, M. (1997). A database for post-genome analysis. *Trends Genet* 13, 375–376.

Khoury, M.J., McCabe, L.L. and McCabe, E.R. (2003). Population screening in the age of genomic medicine. *N Engl J Med* 348, 50–58.

Kohane, I.S., Kho, A.T. and Butte, A.J. (2002). *Microarrays for an integrative genomics*. MIT Press, Cambridge, Massachusetts.

Kohane, I.S., Masys, D.R. and Altman, R.B. (2006). The incidentalome: A threat to genomic medicine. *JAMA* 296, 212–215.

Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.P., Subramanian, A., Ross, K.N. et al. (2006). The Connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313, 1929–1935.

Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I. et al. (2002). Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298, 799–804.

Liang, S., Fuhrman, S. and Somogyi, R. (1998). Reveal, a general reverse engineering algorithm for inference of genetic network architectures. *Pac Symp Biocomput*, 18–29.

Linnâe, C.V. and Schrèoder, J. (1763). Genera morborum, apud C. E. Steinert, Upsaliµ.

Liotta, L.A., Espina, V., Mehta, A.I., Calvert, V., Rosenblatt, K., Geho, D., Munson, P.J., Young, L., Wulfkuhle, J. and Petricoin, E.F., 3rd (2003). Protein microarrays: Meeting analytical challenges for clinical applications. *Cancer Cell* 3, 317–325.

Mulder, N.J., Apweiler, R., Attwood, T.K., Bairoch, A., Barrell, D., Bateman, A., Binns, D., Biswas, M., Bradley, P., Bork, P. et al. (2003). The InterPro Database, 2003 brings increased coverage and new features. *Nucleic Acids Res* 31, 315–318.

Mungall, C.J. (2004). Obol: Integrating language and Meaning in Bio-Ontologies. *Comp Funct Genom* 5, 509–520.

Nature (2002). Microarray standards at last. *Nature* 419, 323.

Ng, S.K., Zhang, Z., Tan, S.H. and Lin, K. (2003). InterDom: A database of putative interacting protein domains for validating predicted protein interactions and complexes. *Nucleic Acids Res* 31, 251–254.

Nguyen, D.V. and Rocke, D.M. (2002). Partial least squares proportional hazard regression for application to DNA microarray survival data. *Bioinformatics* 18, 1625–1632.

Odom, D.T., Zizlsperger, N., Gordon, D.B., Bell, G.W., Rinaldi, N.J., Murray, H.L., Volkert, T.L., Schreiber, J., Rolfe, P.A., Gifford, D.K. et al. (2004). Control of pancreas and liver gene expression by HNF transcription factors. *Science* 303, 1378–1381.

Park, P.J., Tian, L. and Kohane, I.S. (2002). Linking gene expression data with patient survival times using partial least squares. *Bioinformatics* 18(Suppl 1), S120–S127.

Parkinson, H., Sarkans, U., Shojatalab, M., Abeygunawardena, N., Contrino, S., Coulson, R., Farne, A., Lara, G.G., Holloway, E., Kapushesky, M. et al. (2005). ArrayExpress – a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res* 33, D553–D555.

Parmigiani, G. (2003). *The analysis of gene expression data: Methods and software*. Springer, New York.

Pavlidis, P. and Noble, W.S. (2001). Analysis of strain and regional variation in gene expression in mouse brain. *Genome Biol* 2.

Perou, C.M. (2001). Show me the data! *Nat Genet* 29, 373.

Petricoin, E.F., Ardekani, A.M., Hitt, B.A., Levine, P.J., Fusaro, V.A., Steinberg, S.M., Mills, G.B., Simone, C., Fishman, D.A., Kohn, E.C. et al. (2002). Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 359, 572–577.

Potti, A., Mukherjee, S., Petersen, R., Dressman, H.K., Bild, A., Koontz, J., Kratzke, R., Watson, M.A., Kelley, M., Ginsburg, G.S. et al. (2006). A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. *N Engl J Med* 355, 570–580.

Prince, J.T., Carlson, M.W., Wang, R., Lu, P. and Marcotte, E.M. (2004). The need for a public proteomics repository. *Nat Biotechnol* 22, 471–472.

Quackenbush, J. (2004). Data standards for "omic" science. *Nat Biotechnol* 22, 613–614.

Quackenbush, J. (2006). Microarray analysis and tumor classification. *N Engl J Med* 354, 2463–2472.

Quinlan, J. (1992). *C4.5: Programs for machine learning*. Morgan Kaufmann, San Mateo, California.

R Development Core Team (2004). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramaswamy, S., Tamayo, P., Rifkin, R., Mukherjee, S., Yeang, C.H., Angelo, M., Ladd, C., Reich, M., Latulippe, E., Mesirov, J.P. et al. (2001). Multiclass cancer diagnosis using tumor gene expression signatures. *Proc Natl Acad Sci USA* 98, 15149–15154.

Ramoni, M.F., Sebastiani, P. and Kohane, I.S. (2002). Cluster analysis of gene expression dynamics. *Proc Natl Acad Sci USA* 99, 9121–9126.

Raychaudhuri, S., Stuart, J.M. and Altman, R.B. (2000). Principal components analysis to summarize microarray experiments: Application to sporulation time series. *Pac Symp Biocomput*, 455–466.

Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P. and Mesirov, J.P. (2006). GenePattern 2.0. *Nat Genet* 38, 500–501.

Reis, B.Y., Butte, A.J. and Kohane, I.S. (2001). Extracting knowledge from dynamics in gene expression. *J Biomed Inform* 34, 15–27.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E. et al. (2000). Genome-wide location and function of DNA binding proteins. *Science* 290, 2306–2309.

Rhodes, D.R., Yu, J., Shanker, K., Deshpande, N., Varambally, R., Ghosh, D., Barrette, T., Pandey, A. and Chinnaiyan, A.M. (2004). Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proc Natl Acad Sci USA* 101, 9309–9314.

Robinson, W.H., DiGennaro, C., Hueber, W., Haab, B.B., Kamachi, M., Dean, E.J., Fournel, S., Fong, D., Genovese, M.C., de Vegvar, H.E. et al. (2002). Autoantigen microarrays for multiplex characterization of autoantibody responses. *Nat Med* 8, 295–301.

Roche, P.A. and Annas, G.J. (2001). Protecting genetic privacy. *Nat Rev Genet* 2, 392–396.

Roche, P.A. and Annas, G.J. (2006). DNA testing, banking, and genetic privacy. *N Engl J Med* 355, 545–546.

Ross, D.T., Scherf, U., Eisen, M.B., Perou, C.M., Rees, C., Spellman, P., Iyer, V., Jeffrey, S.S., Van de Rijn, M., Waltham, M. et al. (2000). Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet* 24, 227–235.

Rumelhart, D. and McClelland, J. University of California SDPRG (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. MIT Press, Cambridge, Mass.

Saeed, A.I., Bhagabati, N.K., Braisted, J.C., Liang, W., Sharov, V., Howe, E.A., Li, J., Thiagarajan, M., White, J.A. and Quackenbush, J. (2006). TM4 microarray software suite. *Methods Enzymol* 411, 134–193.

Safran, M., Solomon, I., Shmueli, O., Lapidot, M., Shen-Orr, S., Adato, A., Ben-Dor, U., Esterman, N., Rosen, N., Peter, I. et al. (2002). GeneCards(TM) 2002: Towards a complete, object-oriented, human gene compendium. *Bioinformatics* 18, 1542–1543.

Sarwal, M., Chua, M.S., Kambham, N., Hsieh, S.C., Satterwhite, T., Masek, M. and Salvatierra, O, Jr. (2003). Molecular heterogeneity in acute renal allograft rejection identified by DNA microarray profiling. *N Engl J Med* 349, 125–138.

Simon, I., Siegfried, Z., Ernst, J. and Bar-Joseph, Z. (2005). Combined static and dynamic analysis for determining the quality of time-series expression profiles. *Nat Biotechnol* 23, 1503–1508.

Singh, D., Febbo, P.G., Ross, K., Jackson, D.G., Manola, J., Ladd, C., Tamayo, P., Renshaw, A.A., D'Amico, A.V., Richie, J.P. et al. (2002). Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* 1, 203–209.

Soldatova, L.N. and King, R.D. (2005). Are the current ontologies in biology good ontologies? *Nat Biotechnol* 23, 1095–1098.

Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S. et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 98, 10869–10874.

Spellman, P.T., Miller, M., Stewart, J., Troup, C., Sarkans, U., Chervitz, S., Bernhart, D., Sherlock, G., Ball, C., Lepage, M. et al. (2002). Design and implementation of microarray gene expression markup language (MAGE-ML). *Genome Biol* 3, RESEARCH0046.

Storey, J.D. and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 100, 9440–9445.

Su, A.I., Welsh, J.B., Sapinoso, L.M., Kern, S.G., Dimitrov, P., Lapp, H., Schultz, P.G., Powell, S.M., Moskaluk, C.A., Frierson, H.F, Jr. et al. (2001). Molecular classification of human carcinomas by use of gene expression signatures. *Cancer Res* 61, 7388–7393.

Szallasi, Z. and Liang, S. (1998). Modeling the normal and neoplastic cell cycle with "realistic Boolean genetic networks": Their application for understanding carcinogenesis and assessing therapeutic strategies. *Pac Symp Biocomput*, 66–76.

Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E.S. and Golub, T.R. (1999). Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. *Proc Natl Acad Sci USA* 96, 2907–2912.

Toronen, P., Kolehmainen, M., Wong, G. and Castren, E. (1999). Analysis of gene expression data using self-organizing maps. *FEBS Lett* 451, 142–146.

Tseng, Y.H., Butte, A.J., Kokkotou, E., Yechoor, V.K., Taniguchi, C.M., Kriauciunas, K.M., Cypess, A.M., Niinobe, M., Yoshikawa, K., Patti, M.E. et al. (2005). Prediction of preadipocyte differentiation by gene expression reveals role of insulin receptor substrates and necdin. *Nat Cell Biol* 7, 601–611.

Tusher, V.G., Tibshirani, R. and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 98, 5116–5121.

van de Rijn, M. and Gilks, C.B. (2004). Applications of microarrays to histopathology. *Histopathology* 44, 97–108.

Wang, L., Wu, Q., Qiu, P., Mirza, A., McGuirk, M., Kirschmeier, P., Greene, J.R., Wang, Y., Pickett, C.B. and Liu, S. (2001). Analyses of p53 target genes in the human genome by bioinformatic and microarray approaches. *J Biol Chem* 276, 43604–43610.

Weinmann, A.S., Yan, P.S., Oberley, M.J., Huang, T.H. and Farnham, P.J. (2002). Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. *Genes Dev* 16, 235–244.

Weinstein, J.N., Myers, T.G., O'Connor, P.M., Friend, S.H., Fornace, A.J, Jr., Kohn, K.W., Fojo, T., Bates, S.E., Rubinstein, L.V., Anderson, N.L. et al. (1997). An information-intensive approach to the molecular pharmacology of cancer. *Science* 275, 343–349.

Welsh, J.B., Sapinoso, L.M., Kern, S.G., Brown, D.A., Liu, T., Bauskin, A.R., Ward, R.L., Hawkins, N.J., Quinn, D.I., Russell, P.J. et al. (2003). Large-scale delineation of secreted protein biomarkers overexpressed in cancer tissue and serum. *Proc Natl Acad Sci USA* 100, 3410–3415.

Wen, X., Fuhrman, S., Michaels, G.S., Carr, D.B., Smith, S., Barker, J.L. and Somogyi, R. (1998). Large-scale temporal gene expression mapping of central nervous system development. *Proc Natl Acad Sci USA* 95, 334–339.

Wheeler, D.L., Chappey, C., Lash, A.E., Leipe, D.D., Madden, T.L., Schuler, G.D., Tatusova, T.A. and Rapp, B.A. (2000). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 28, 10–14.

Wheeler, D.L., Church, D.M., Edgar, R., Federhen, S., Helmberg, W., Madden, T.L., Pontius, J.U., Schuler, G.D., Schriml, L.M., Sequeira, E. et al. (2004). Database resources of the National Center for Biotechnology Information: update. *Nucleic Acids Res* 32(Database issue), D35–D40.

Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Edgar, R., Federhen, S. et al. (2006). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 34, D173–D180.

World Health Organization (2005). *International Statistical Classification of Diseases and Health Related Problems*. Tenth Revision/Ed, Geneva.

Wuensche, A. (1998). Genomic regulation modeled as a network with basins of attraction. *Pac Symp Biocomput*, 89–102.

RECOMMENDED RESOURCES

Websites

ArrayExpress	http://www.ebi.ac.uk/arrayexpress/
Bioconductor	http://www.bioconductor.org/
Entrez Gene	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
Gene Expression Omnibus	http://www.ncbi.nlm.nih.gov/geo/
GenePattern	http://www.broad.mit.edu/cancer/software/genepattern/
GenMAPP	http://www.genmapp.org/
International Classification of Diseases	http://www.who.int/classifications/icd/en/
Kyoto Encyclopedia of Genes and Genomes	http://www.genome.jp/kegg/
MultiExperiment Viewer	http://www.tm4.org/mev.html
Open Biomedical Ontologies	http://obo.sourceforge.net/
PeptideAtlas	http://www.peptideatlas.org/
Proteomics Identifications Database	http://www.ebi.ac.uk/pride/
R	http://www.r-project.org/
Stanford Microarray Database	http://genome-www5.stanford.edu/
Unified Medical Language System	http://umlsks.nlm.nih.gov/

Books

Brown, T.A. (2006). *Genomes 3, 3rd edition*. Garland Science Publisher, New York.

Review articles

Butte, A. (2002). The use and analysis of microarray data. *Nat Rev Drug Discov* 1, 951–960.

Quackenbush, J. (2006). Microarray analysis and tumor classification. *N Engl J Med* 354, 2463–2472.

Dudoit, S., Gentleman, R.C. and Quackenbush, J. (2003). Open source software for the analysis of microarray data. *Biotechniques Suppl*, 45–51.

CHAPTER

Systems Biology and Systems Medicine

Nathan D. Price, Lucas B. Edelman, Inyoul Lee, Hyuntae Yoo, Daehee Hwang, George Carlson, David J. Galas, James R. Heath and Leroy Hood

INTRODUCTION

A new approach to biology, which we call *systems biology*, has emerged over the past 10 years or so – an approach that looks at biology as an informational science, studies biological systems as a whole, and recognizes that biological information in living systems is captured, transmitted, modulated, and integrated by biological networks that pass this information to molecular machines for execution. This approach differs from early "systems approaches to biology" in that it attempts both a bottom-up approach (from large molecular datasets) and a top-down approach (from computational modeling and simulations) where there is an attempt to trace complex observations of phenotype back to the digital core encoded in the genome.

New measurement and visualization technologies, together with powerful computational and modeling tools, have transformed systems biology by making it possible for the first time to execute the five uniquely defining features of contemporary systems biology:

1. measurements of the various types of biological information that are global or comprehensive to the greatest extent possible (e.g., measure the digital code of all genes or, for example, the concentrations of all mRNAs, all proteins, all metabolites, etc.);
2. the different levels of information (DNA, RNA, protein, etc.) must be integrated to understand or capture how the environment has modified the basic digital information of the genome at each level of the biological information hierarchy (DNA, RNA, proteins, interactions, biological networks, cells, individuals and ecologies) and thereby to induce biological responses;
3. all biological systems (e.g., networks) must be studied dynamically as they capture, transmit, integrate and utilize the biological information necessary for the execution the two most fundamental responses of living organisms – development or physiological responses to the organism's environment;
4. all measurements must be quantitatively determined to the greatest extent possible; and
5. the global and dynamic data from the variety of information hierarchies must be integrated and modeled.

These models essentially create hypotheses about biological functions and mechanisms of disease that can then be tested experimentally by systems perturbations – the enumeration of new datasets, their integration, modifications of models, in an iterative manner, until the working models reflect the reality of the experimental data. Once the models accurately fit the biological systems measurements, they can then be used to predict the results of perturbing the system in new ways (e.g., for designing treatment strategies or for identifying likely underlying causes for disease).

The information revolution in biology will enable systems medicine to emerge in the next 10 years. Key to this transformation will be harnessing computationally the vast amount of biological information becoming available through rapid

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

advancement in measurement, visualization and computational technologies. Two fundamental types of biological information, the digital information of the genome and interacting information from the environment, are integrated together to specify the five fundamental mechanisms of life – evolution, development, physiological responses, aging and the onset and progression of disease (Hood et al., 2004). Advances in high-throughput DNA sequencing will enable each person's genome to be sequenced rapidly and at a reasonable cost, providing digital information for each person and making possible the prediction of increasingly accurate probabilistic health futures. Also, the advance of high-throughput measurement technologies will enable the assessment of dynamic environmental information emerging from the integration of genome and environmental information from each individual, as reflected, for example, by the changing levels of proteins in the blood – thus providing a real time (current) health assessment of the individual. These technologies will generate tremendously large, dynamical datasets about health states and hence about the states of relevant biological networks of the individual patients. This detailed information will arise through the use of perturbation experiments coupled with global experimental technologies, and progress in systems biology research is increasingly elucidating the functions and structures of these networks.

A systems approach to medicine argues that disease arises from disease-perturbed biological networks and that the dynamically changing, altered patterns of gene expression that are controlled by these perturbed networks give rise to the disease manifestations. Here, we present a systems view of biology and disease together with a discussion of some recent advances in state-of-the-art *in vitro* and *in vivo* diagnostics technologies, and we suggest how, as these technologies mature, they will move us towards a future of predictive, personalized, preventive, and participatory medicine (*P4 medicine*) (Hood et al., 2004).

SYSTEMS SCIENCE IN BIOLOGY AND MEDICINE

Systems Biology

How do systems approaches to biology and medicine lead to a revolutionary view of medicine? We present this systems view in some detail, because an understanding of P4 medicine and its implications for society are predicated upon understanding these systems principles. Two primary domains of biological information lend themselves readily to such systems-level analysis: the static, digital information of the genome, and the dynamic information arising from environmental interactions with the subcellular, cellular, and tissue levels of organization (Hood et al., 2004). Digital genome information encodes two types of biological networks – *protein interactions* and *gene regulatory networks*. Protein networks transmit and use biological information for development, physiology and metabolism. Gene regulatory networks – transcription factors, their regulatory binding sites and the small RNAs that regulate networks of other transcription factors and other RNAs interacting with one another – receive information from signal-transduction networks, integrate and modulate it, and convey the processed information to networks of genes or molecular machines that execute developmental and physiological functions. In biological systems, these two types of networks are closely integrated. The organization of these networks can be inferred from various types of measurements including, for example, global measurements of dynamically changing levels of mRNAs and proteins during developmental and physiological responses, as well as large-scale measurements of protein–protein and protein–DNA interactions. There are multiple hierarchical levels of organization and information (for example, DNA, RNA and protein networks, cellular and metabolic networks, and organization and responses of organ systems). To understand biological systems, information must be gathered from as many information levels as possible and integrated into models that generate hypotheses about how biology works. Let us now consider the logic of systems approaches to medicine.

Systems Medicine

The central premise of systems medicine is that disease-perturbed networks alter patterns of expression in genes and proteins – and that these altered patterns encode the dynamic pathophysiology of the disease and necessarily results in altered molecular fingerprints that can be detected clinically. Through advances in measurement technologies and computational analysis tools for tissue and blood analyses, we will be able to read these signals to make a multitude of diagnoses to distinguish health from disease and to determine the nature of any pathology. Multi-parameter analyses of biological information will be key to tracking these altered patterns in disease-perturbed networks. Diagnostic methods of the past have been pauci-parameter in nature – usually measuring just a single parameter relevant to a specific disease state (e.g., PSA levels to assess prostate health), and so our ability to track health and disease has been limited. Current and emerging technologies are creating a transition into a new era of predictive, preventative, and personalized medicine. A causal disease perturbation could be the result of specific, disease-causing DNA mutations, pathogenic organisms, or other pathological environmental factors such as toxins. Molecular fingerprints of pathological processes can take on many molecular forms, including the analyses of proteins (Wang et al., 2005a, b), DNA (Papadopoulou et al., 2006), RNA (Scherzer et al., 2007), and metabolites (Solanky et al., 2003), as well as informative, post-translational modifications to these molecules such as protein phosphorylation.

Signals related to health and disease can be found in multiple sites. For instance, many bodily fluids such as the blood, urine, saliva, cerebral spinal fluid and so forth can be sampled to identify evidence of perturbed molecular fingerprints reflecting the altered expression patterns of genes and proteins in disease-perturbed biological networks. Of these, the blood is likely the most information rich organ (or fluid) in that it bathes all of the tissues in the body, and it is easily accessible for diagnostic procedures. In addition to the biomolecules secreted into the blood

from cells and tissues throughout the body, the transcriptomes and proteomes of cells circulating in the blood (e.g., white blood cells) are also potentially an abundant source of biomedically important information (Buttner et al., 2007). Thus, the amount of information available in the blood about health and disease is enormous if we learn to read and interpret the molecular signals.

These dynamically changing, disease-perturbed networks lead to two important conclusions. First, some significant network nodal points change before the related clinical or histological changes are evident. Therefore, labeled biomarkers that are specific for the changing nodes or the biological processes they regulate could be used for *in vivo* imaging diagnostics even before symptoms arise, as has already been shown in patients (Golub et al., 1999; Quackenbush, 2006; Ramaswamy et al., 2001; Wang et al., 2005a, b). Alternatively, if some of these altered nodes encode secreted proteins, they could provide readily accessible *in vitro* diagnostic blood markers for early disease detection. Second, many of the sub-networks of proteins that change during the onset of disease affect changes in phenotypic traits that are consistent with the pathology of the disease.

About 900 perturbed mRNAs appear to encode the core prion-disease response, and, during the progression of the prion-disease process, several hundred undergo significant gene expression changes well before the clinical signs of prion disease. Many of these potential early disease "sentinels" are predicted to be secreted into the blood and therefore represent potential protein biomarkers for early disease diagnosis through blood protein analysis. One interesting new diagnostic method evaluates relative expression reversals in protein concentrations or gene expression levels. This procedure eliminates the need for data normalization or the establishment of population-wide thresholds. This approach has been successfully used to identify robust and accurate classifiers for prostate cancer (Xu et al., 2005), sarcomas (Price et al., 2007), and a variety of other cancers (Tan et al., 2005), as well as to predict treatment outcomes in breast cancer (Ma et al., 2004), demonstrating the efficacy of even relatively low-order systems analysis in medicine.

Another key issue that needs to be addressed is how to best decouple two primary dimensions of disease information: namely, disease stratification (e.g., which type of prostate cancer is present) and disease progression (e.g., stage of development of a particular prostate cancer stratification in time). Molecular signature changes can indicate the presence of different diseases (stratification) as well as their stage (progression). Thus, a key challenge in future will be to develop analysis tools that will enable us to differentiate the location of the physiologic state in regards to both of these critical clinical dimensions.

MULTI-PARAMETER BLOOD-BOURNE BIOMARKERS

In principle, it is clear how blood samples containing secreted molecules from all of the tissues in the body can be used as a window into health or disease states (Anderson and Anderson, 2002; Fujii et al., 2004; Hood et al., 2004; Lathrop et al., 2003; Lee et al., 2006). In practice, however, the task of identifying markers of disease states in the vast array of secreted proteins can seem daunting. There are millions of different proteins present in the blood and they are expressed at levels that probably differ by 12 orders of magnitude (10^{12}). Thus, there are indeed very significant challenges associated with defining relevant biomarkers in the plasma proteome, as is evidenced by the paucity of blood protein markers found thus far, despite significant efforts (Wilson, 2006). This fact highlights the need to develop well-founded systems approaches to diagnostics in order to hasten the identification of such markers and to harness the tremendous information potential of the blood. While the potential information available to diagnose heath and disease is enormous, it is also true that the challenges of separating signal from background noise (biological or measurement noise) are also very significant. Sources of noise include error in measurements, polymorphisms in the population, environmental variations, stochastic variations, as well as signal dilution through mixing and other processes of molecule transport. Thus, enhancing signal while reducing noise will be a primary theme of research in predictive medicine for the foreseeable future.

An important strategy for dealing with noise arising from genetic polymorphisms and health histories in the population are through dynamic, subtractive analyses carried out in the individual patient. In other words, measurements taken from individual patients at different time points (longitudinal data) can be used to perform subtractive analyses where only the differences observed are considered. Thus, each patient becomes their own control, which eliminates many of the sources of noise. Eventually, it will be important for each patient to have biannual blood analyses taken so that changes can be observed relative to the background of what is normal *in each individual* rather than relative to the population at large. Such databases of measurements will be essential to enabling personalized and predictive medicine. Using each patient as their own control is one of the essential features of the emerging personalized medicine.

Organ-Specific Blood Protein Fingerprints

We discuss the organ-specific blood protein fingerprint approach to diagnostics because it lies at the very heart of the predictive medicine that will emerge over the next 10 years. Our feeling is that this will be one of the central data gathering strategies for predictive and personalized medicine and, accordingly, we will discuss it in detail. The idea is that disease arises from dynamically changing disease-perturbed networks, that the diseased organs will secrete proteins into the blood, and that if these proteins are encoded by disease-perturbed networks their levels of expression in the blood will be altered in a manner that reflects the specific nature of the disease. Indeed, this idea is the basis for the very broad range of blood biomarker studies that are being carried out by many scientific centers (Anderson and Anderson, 2002; Fujii et al., 2004; Hood et al., 2004; Lathrop et al., 2003; Lee et al., 2006). The major difficulty with this simple view is that if you identify multiple blood proteins whose changes are specific for a

particular disease state (as compared against normal controls) and then examine the same blood markers for that disease in bloods drawn from individuals, with say 10 other pathologies, these marker proteins can change in unpredictable ways since multiple organs control the expression of most blood proteins and these organs respond differently to various environmental signals. The important point is that if a marker protein synthesized in five organs changes in the blood, we cannot be certain which organ(s) is the origin of the change. While that marker may sample a biological network relevant to disease diagnosis, since its origin is not clear its use as a disease biomarker may raise more questions than it answers. The solution to this dilemma is clear – employ organ-specific blood protein biomarkers whose changes must therefore reflect changes only in the organ itself. If enough of these organ-specific blood proteins are sampled, they will represent a survey of many different biological networks in the organ of interest and will provide sufficient diagnostic information for any disease. Eventually, correlation of these organ-specific biomarkers with more general biomarkers may prove to be the best long-term strategy for developing diagnostic fingerprints.

We have several lines of preliminary evidence that suggest this organ-specific blood protein approach will be effective. In prostate cancer, for example, there are disease-mediated altered patterns of mRNA and protein expression in the prostate. Some of these genes are expressed primarily in the prostate (organ-specific products) and some of these organ-specific proteins are secreted into the blood, where they collectively constitute a protein molecular fingerprint comprised of say 100 or more proteins whose relative concentration levels probably report the status of the biological networks in the prostate gland. We have demonstrated that changes in the blood concentrations of several of these prostate-specific blood proteins reflect the various stages of prostate cancer and, as discussed above, various brain-specific blood proteins also reflect the progression of prion disease in mice. We propose that the distinct expression levels of the individual proteins in each fingerprint represent a multi-parameter (and therefore potentially information-rich) diagnostic indicator reflecting the dynamic behavior of, for example, the disease-perturbed networks from which they arise. The analysis of 50 or so organ-specific proteins should allow us to both stratify diseases in the organ as well as determine their stage of progression. We have identified tens to hundreds of organ-specific transcripts in each of the 40 or so organs that we have examined in mouse and human.

We can envision a time over the next 5–10 years when 50 or so organ-specific blood proteins will be identified for each of the 50 or so major organs and tissues in humans – so that computational analyses of the relative concentrations of the protein components in these organ-specific fingerprints will enable blood to become the primary window into health and disease. When we analyze data from these blood indicators, we also may be in a position to identify the dynamically changing disease-perturbed network(s). The analysis of these dynamic networks will allow us to study in detail the pathophysiology of the disease response and hence be in a position to think of new approaches to therapy and prevention. To generate the ability to assess simultaneously all 50 organ blood protein fingerprints in patients, we ultimately need to develop the microfluidic or nanotechnology tools for making perhaps 2500 rapid protein measurements (e.g., 50 proteins from each of 50 human organs) from a droplet of blood. Nanotechnology is necessary because only this severe miniaturization can allow the necessary thousands of measurements from a single drop of blood. In order to reach this stage we will also have to create the appropriate computational tools to capture, store, analyze, integrate, model, and visualize the information arising from this approach.

EMERGING *IN VIVO* AND *IN VITRO* TECHNOLOGIES

In vitro Measurement Technologies

P4 medicine will require that *in vitro* measurements of thousands of blood proteins be executed rapidly, automatically and inexpensively on small blood samples. This will require miniaturization, parallelization, integration and automation of tissue and blood purification and measurement technologies. In short, it will require making *in vitro* measurements inexpensively – perhaps for a penny or less per protein measured. This need for inexpensive *in vitro* measurements is driving the development of integrated microfluidics and nanotechnologies for *in vitro* diagnostics. We further argue that the complex changes in these protein levels must be analyzed computationally to search for the patterns that correlate with particular diseases – and indeed to stratify each disease as well as determine its stage of progression.

To meet the expanding requirements of systems medicine, *in vitro* measurements of cells, proteins, mRNAs, etc., whether for fundamental biological studies or for pathological analysis, must not just be inexpensive, but must also be sensitive, quantitative, rapid, and executed on very small quantities of tissues, cells, serum, etc. We are working towards developing chip platforms that can take a few microliters (a finger prick) of blood, separate the plasma from the serum, and then measure on the order of 100 or more proteins and/or mRNAs quantitatively, with high sensitivity and specificity, and in a few minutes time. We can envision a time 5–10 years hence when small hand-held devices will be able to make these 2500 measurements from a fraction of a droplet of blood, send them *via* wireless to a server for analysis and consequently inform the patient and physician as to the status of the patient. While many fundamental scientific challenges remain to be solved before this goal is achieved, none of those challenges appears insurmountable at this point.

One will also be able to use the blood cells as powerful diagnostic markers – either of infectious diseases or of genetic diseases. Microfluidic cell-sorting technologies for being able to sort blood cells into their 10 or so individual types for analysis are now available. Even more important is the emergence of

single-cell analytic tools where DNA, mRNAs or small RNAs, proteins and even metabolites can be analyzed rapidly from individual cells. The cells can also be perturbed with appropriate environmental stimuli to identify defects. It is possible that an appropriate analysis of immune cells (both innate and adaptive) from the blood will reveal important information about past antigenic history of the individual as well as current state of immunological responsiveness. Similarly, analysis of rare blood cells such as circulating cancer cells may also be utilized to guide therapies.

Another type of *in vitro* measurement will be a determination of the complete genome sequence of individuals with a nanotechnology approach to sequencing single strands of DNA on a massively parallel basis (billions of DNA strands simultaneously). A device like this will emerge over the next 5–10 years and will allow millions, if not billions, of individual human genome sequences to be determined rapidly, inexpensively and in a massively parallel manner. The 2500 organ-specific blood marker measurements and the complete individual human genome sequences will be the heart of the predictive and personalized medicine that will emerge over the next 10 years or so.

While transcriptomic approaches have proven useful for identifying informative molecular signatures for a number of diseases (Quackenbush, 2006), high-throughput proteomic characterizations have lagged behind. The reason for this difference is clearly the more difficult challenge of measuring proteins compared to mRNA transcripts. Emerging proteomics technologies hold the promise of greatly improving our ability to make detailed disease assessments using protein-based molecular signatures. One key advantage of protein signatures relative to gene expression is that the proteins found in the blood and other accessible bodily fluids exhibit slower degradation rates than their mRNA counterparts – and of course proteins are the direct agents mediating the disease process itself. Thus, protein concentration signatures represent an important class of molecular signature for disease diagnosis.

The Imaging of Proteomics and Blood Protein Biomarkers

The blood is an ideal organ for identifying biomarkers because it interacts with virtually every organ and major tissue type in the body, and each secretes proteins into the blood. Moreover, the blood is easily accessible for diagnostic studies – and presumably the analysis of changes in protein levels or protein modifications that could serve as surrogates or reflections of health and disease. As noted earlier, the blood is an enormously complex organ; it contains hundreds of thousands to millions of proteins whose concentrations range over 12 orders of magnitude. Since there is for proteins no equivalent to the polymerase chain reaction (PCR) amplification procedure, only the more abundant proteins can be seen by conventional protein analysis studies. It should also be pointed out that many different features of proteins ultimately must be characterized or quantified: identification, quantification, chemical modifications, alternative mRNA splicing products, cellular localizations, three-dimensional structures (and their dynamics), as well as their functions. We here are only concerned with the quantification and identification of proteins.

There are two general approaches to the analysis of protein mixtures. In one case protein-capture agents such as antibodies are used – and with appropriate controls these approaches can quantify the levels of proteins. These techniques include Western blot analyses, ELISA assays, surface plasmon resonance and protein chips. A second approach is to use mass spectrometry. Since proteins have large masses that are difficult to analyze accurately by mass spectrometry, most proteins are analyzed after converting them into tryptic peptides (or other proteolytic fragments) whose masses are far smaller and hence more easily analyzed. Because blood is such a complex mixture, proteins of interest are often enriched by some type of fractional procedure (charge, hydrophobility, size and/or the presence of modifications such as glycosylation or phosphorylation), before (or after) their conversion into peptides and analysis by mass spectrometry. Isotopically labeled and synthesized peptides may be used to identify and quantify (relative or absolute) the same peptides from unknown samples (and hence one obtains a proxy for the quantification of their corresponding proteins) very effectively by mass spectrometry.

We discuss below examples of some of these techniques.

Antibody Microarrays Using Surface Plasmon Resonance Imaging (SPRI)

Protein chip methods hold great potential for broad quantitative screens of proteins, and a variety of techniques have been developed based on antibody binding (Haab et al., 2001; Olle et al., 2005; Song et al., 2007). Various types of antibody arrays have been used for biomarker discovery and protein profiling of serum from patients with prostate, lung, pancreas, and bladder cancer (Gao et al., 2005; Miller et al., 2003; Orchekowski et al., 2005; Sanchez-Carbayo et al., 2006). One emerging approach with tremendous promise is SPRI (Hu et al., 2007; Huber and Mueller, 2006; Koga et al., 2006), which enables real-time, label-free measurement of protein expression by large numbers of different antibodies. SPR-based chips have a detection sensitivity of 10–100 times less than ELISA (Hu et al., 2007), but have a spatial resolution down to approximately 4 μm (Lyon et al., 1998). It is thus possible to print up to 800 unique antibodies on Lumera Nanocapture Gold™ microarray slides and monitor the abundance of the target proteins in real time (Hu et al., 2007), even in complex samples such as blood, because the same slide can be regenerated for reuse many times in 10 minutes or less (Z. Hu, C. Lausted, unpublished observations). We are in the process of automating the sample introduction procedures using microfluidic chips, which means that SPRI has the capacity necessary to screen rapidly through tens to hundreds of patient samples. Thus, this approach holds tremendous promise to be able to screen through large numbers of proteins, including secreted proteins and cell-surface markers and not only measure their presence,

but also abundance and the dynamics of their binding. The limitation of this technique is its dependence on the affinity and specificity of the antibodies it employs for detection – cross-reactivities in complex protein mixtures (such as blood) can pose significant problems.

DNA-Encoded Antibody Libraries (DEAL)

One recently developed technique that offers great potential for detailed analyses is DEAL. The primary advantage of DEAL is that it uses a single, robust chemistry – that of DNA hybridization – to spatially localize and detect proteins, mRNAs, and cells, all in a multiplexed fashion. Antibodies are typically too fragile to survive the fabrication procedures associated with assembling robust microfluidics chips, but DNA oligomers are significantly more robust. DEAL thus enables the detection of panels of protein biomarkers within a microfluidics environment and from very small quantities of biological material (100 nanoliters or so) (Yang et al., 2006). This amount of plasma can be readily separated from whole blood on-chip, thus allowing for the measurement of serum biomarkers from a finger-prick of whole blood. In addition, within the environment of flowing microfluidics, the rate-limiting step in performing a surface immunoassay is the kinetics of the binding of the analyte to the surface-bound capture agent (Zimmermann et al., 2005). Thus, DEAL-based immunoassays can be executed very rapidly.

The versatility of DEAL also enables multiplexed cell sorting and localization, followed by few- or single-cell measurements of protein, RNA, and other biomolecules (Bailey et al., 2007). DEAL can be engineered into a highly sensitive and very rapid measurement technique, with a reported detection limit of 10 fM for the protein IL-2, 150 times more sensitive than the corresponding commercial ELISA assay. This sensitivity can be applied to the isolation of rare cells based on combinations of cell-surface markers, enabling the isolation and addressing of individual cancer stem cells. DEAL can also be used to make single-cell measurements of secreted proteins from each of these isolated single cells. Thus, DEAL offers superb sensitivity and the ability to perform spatially multiplexed detection for characterization of rare cell types, such as circulating cancer cells or cancer stem cells. These advantages still face the inherent limitations of antibodies, so the development of new approaches to generating protein-capture agents is a critical part of future development of comprehensive blood diagnostics.

Mass Spectrometry-based Techniques

Isotopic Tagging for Relative and Absolute Quantification (iTRAQ) of Proteins Analyzed by Mass Spectrometry

Stable isotope labeling enables the quantitative analysis of protein concentrations through mass spectrometry (MS). One state-of-the-art technique for quantitative MS is iTRAQ (Ross et al., 2004), which uses stable isotope labeling of proteolytic peptides. This technique modifies primary amino acid groups of peptides by linking a mass balance group, and a reporter group by forming an amide bond. When MS/MS is used for analysis with iTRAQ-tagged peptides, the mass balancing carbonyl moiety is released as a neutral fragment generating 4–8 distinct sets of peptides whose relative abundances can be determined quantitatively. Because eight different iTRAQ reagents are currently available, comparative analysis of a set of two to eight samples is feasible within a single MS/MS run (Hu et al., 2007). The, iTRAQ technology represents the state-of-the-art in quantitative proteomics and represents a promising technology for using proteomics to differentiate key differences in protein networks.

Glycopeptide Capture – Front-End Enrichment of Blood Proteins Containing Sugar Residues

Mass spectrometry-based methods will allow for the identification of proteins spanning approximately three orders of magnitude in concentration from a given sample. Therefore, methods that can select specified fractions of the proteome are important for simplifying the sample sufficiently to identify the proteins of interest. One recently developed approach is the shotgun glycopeptides capture approach (Sun et al., 2007). This approach selects for N-linked or O-linked glycosylated peptides by selectively coupling these peptides to beads – allowing the uncoupled peptides to be washed away – then the glycosylated peptides are be released and analyzed by mass spectrometry. Both secreted and cell-surface proteins are enriched for glycopeptides compared with their nuclear and cytoplasmic counterparts. Thus, this approach can be used to, for example, identify candidates for unique cell-surface markers to make identifications of clinically relevant cellular subpopulations.

In Vivo Molecular Diagnostics

As it relates to personalized and predictive medicine, *in vivo* molecular diagnostics will also require the development of a diverse library of molecular imaging probes. These modular tools can be used to:

1. identify the specific location of disease-perturbed networks in patients;
2. link *in vivo* molecular measurements in diseased tissue in patients to *in vitro* measurements;
3. rapidly assess the efficacy of personalized therapeutics; and
4. validate that a drug is hitting its target and inducing the desired pharmacological outcome.

Although there are many *in vivo* imaging modalities, perhaps the best current method from the point of view of personalized and predictive medicine in patients is positron emission tomography (PET) molecular imaging (Czernin and Phelps, 2002). For PET, trace quantities of radiolabeled molecular probes are injected into the patient. As the probes circulate through the body and its various organ systems, they interact with target proteins to provide imaging assays for, as examples, the rate of metabolic processes, the concentration of receptors in signal transduction, enzyme activity, DNA-replication rates, hormone status, pharmacokinetics, and pharmacodynamics.

COMPUTATIONAL AND MATHEMATICAL CHALLENGES IN SYSTEMS MEDICINE

Molecular Signature Classifiers

Computational methods are needed to reduce the high degree of data dimensionality associated with global datasets to identify molecular signatures that can be used for disease diagnosis. Despite notable and significant challenges that remain (Dupuy and Simon, 2007; Simon, 2005), computational analyses to identify molecular signatures from global gene expression datasets that can be used for diagnosis and treatment selection (Quackenbush, 2006) is an area of research that has shown significant promise. These studies typically involve the collection of samples from two or more classes (e.g., cancer versus normal, or responsive versus non-responsive to treatment) and the use of a set of data on which to train the classifier and another set on which to test. In the absence of a true test set, re-sampling methods (such as cross-validation) are generally used to estimate likely performance of the classifier on future data. The ability to generate an accurate classifier is a function of factors such as: (1) the size of the training set relative to the number of features, (2) the computational method used, and (3) the inherent distinctness of the selected phenotypes. Typically, the number of samples is far less than the number of transcripts, and consequently over-fitting is a significant problem. This leads to the need for computational methods that avoid over-fitting when selecting a classifier.

A variety of methods have been applied to disease diagnoses, including approaches based on support vector machines (Furey et al., 2000; Ramaswamy et al., 2001) and relative expression reversals (Geman et al., 2004; Price et al., 2007; Tan et al., 2005; Xu et al., 2005), among many others. Application of these methods has perhaps been applied most extensively to the study of cancer, leading to the discovery of molecular classifiers of varying degrees of accuracy to identify prognostic signatures for breast cancer (Adler and Chang, 2006; Buyse et al., 2006; Dai et al., 2005; Foekens et al., 2006; Glinsky et al., 2004a, b; Goncalves et al., 2006; Ivshina et al., 2006; Liu et al., 2007; Ma et al., 2004; Nuyten et al., 2006; Park et al., 2007; Pawitan et al., 2005; Thomassen et al., 2007; van 't Veer et al., 2002; van de Vijver et al., 2002; Wang et al., 2005a, b; Weigelt et al., 2005), ovarian cancer (De Cecco et al., 2004; Smith, 2002; Spentzos et al., 2004), colon cancer (Barrier et al., 2006; Giacomini et al., 2005), prostate cancer (Dhanasekaran et al., 2001; Glinsky et al., 2004a, b; Halvorsen et al., 2005; Lin et al., 2005; Luo et al., 2002; Singh et al., 2002; Xu et al., 2005), and brain cancer (Fuller et al., 2005; Kim et al., 2002), among others.

Biological Networks and the Interactome

Among the most promising opportunities presented by systems biology is the integrated assessment of multiple biological parameters to magnify our understanding of complex developmental or pathological states. The interaction among all individual biomolecules in complex regulatory, signal transduction, and feedback networks present in eukaryotic organisms forms the theoretical framework for this higher-order analysis and is referred to as the *interactome*. These dynamic associations include protein–protein interactions, transcriptional regulation, and post-transcriptional gene silencing by short-interfering RNA and microRNA (miRNA) (Ruvkun, 2001). Multiple high-throughput methods have been developed to characterize the interaction between proteins, including the yeast-two-hybrid system (Ito et al., 2001) and surface plasmon resonance (Usui-Aoki et al., 2005). These methods are to determining interactome interactions what microarray studies are to determining relative mRNA concentrations.

2009 UPDATE

An area of intense interest and progress during 2008 has been multiparameter blood-borne biomarkers. An important advance for making multiple protein measurements with high sensitivity in blood was the reported development of an integrated blood barcode chip (Fan et al., 2008). This microfluidic system has the ability to sensitively sample a large panel of protein biomarkers over a wide range of concentrations, a key challenge for the complex proteome in blood. This device holds the promise of inexpensive, highly sensitive measurements from just a pinprick droplet of blood.

In addition to protein biomarkers, a number of findings were made regarding other important possible sources of diagnostic information from the blood to guide systems medicine. One intriguing finding showed that when glioblastoma tumor cells release microvesicles, they are taken up by normal host cells and also ultimately appear in the blood (Skog et al., 2008). This study also showed that the tumor-specific EGFRvIII mutation was detected in serum microvesicles in a subset of patients. Such mutation-specific findings are critical to guide individualized therapies. In another study, it was found that miRNAs not only hold promise as tissue-based biomarkers, but that they also exist in stable forms in human plasma – as was demonstrated in a study of prostate cancer (Mitchell et al., 2008).

Outside of blood, markers in other accessible bodily fluids as well as tissue markers continued to show promises. For example, analysis of the average concentration over 24 h of certain metabolites found in urine was shown to reliably differentiate a number of different phenotypes, including blood pressure linked to coronary heart disease and stroke (Holmes et al., 2008). At the tissue level, a large study involving over 1000 patients showed that the use of panels of molecular markers

conclusively improved diagnosis substantially beyond diagnoses based on accepted risk factors, even in elderly patients (Zethelius et al., 2008). Gene expression and genetic profiles were also used to stratify breast cancer tumors, including cancer cells with stem-cell like properties (Shipitsin et al., 2007), a highly important cell population based on their demonstrated ability to regenerate tumors.

Progress was also made on deciphering protein networks of importance to systems medicine. For example, the proteome-wide mapping of interactions between hepatitis C virus and human proteins was performed (de Chassey et al., 2008). Close to 500 interactions were reported, which were enriched for highly interconnected proteins and that were related to cellular pathways targeted by the hepatitis C virus. Another important advance for uncovering the human proteome interaction network was the development of a novel analytical process for identifying human protein complexes (Glatter et al., 2009). This method was developed to improve upon current methods that are impeded because of low throughput, sensitivity, and robustness of existing procedures (Glatter et al., 2009). With this new method, the experimental identification of the human proteome interaction network should accelerate.

An example of a model for systems medicine can be found from the study of prion disease in mice (Hwang et al., 2009). Here the dynamic onset of the infectious prion disease in eight different inbred strain/prion strain combinations was studied at the level of mRNA in the brain (the affected organ) and showed that a series of interlocking protein networks that surround the prion protein are significantly perturbed across the 150-day span from disease initiation to death. It was shown that differential networks could be derived from comparisons of mRNA-expression patterns in normal and diseased brains at each of eight to ten time points after infection. The initial network changes occur well before the clinical signs of the disease can be detected and predict later widespread histopathological events. By subtractive analyses using various of the inbred mouse strain/prions strain combination and the application of powerful new statistical methods, a core of 333 changing mRNA transcripts were identified that mapped to four biological networks (prion replication and accumulation, astrocyte and microglial activation, bouton and axon degeneration, and nerve-cell death) that appear to encode the core prion-disease response. About two-thirds of these transcripts encoded and explained most of the known aspects of the pathophysiology of the disease and about one-third encoded aspects of the disease here-to-fore unknown. The dynamics of these four biological networks also suggested new approaches to therapy for prion disease. This study demonstrates the power of the systems approach to medicine in unraveling the complexities of disease mechanisms (Hwang et al., 2009).

Finally, high-throughput sequencing of individual human tumor samples is beginning to yield important information into human cancer that will enhance systems medicine going forward. Importantly, genomic sequencing studies perform unbiased genomic analyses that are uncovering novel mutations that are prevalent in tumors but were previously unknown, as was done for glioblastoma multiform (Parsons et al. 2008) as well as breast and colorectal cancers (Wood et al., 2007). In a study of 24 pancreatic cancers, it was found that they contain an average of 63 genetic alterations each (Jones et al., 2008). These mutations were found to define a core set of 12 signaling pathways whose dysregulation help explain the major features of pancreatic tumorigenesis (Jones et al., 2008).

Taken together, progress on many fronts continues to use systems approaches to drive toward a future of personalized, predictive, preventative, and participatory (P4) medicine.

2009 UPDATE REFERENCES

de Chassey, B., Navratil, V., Tafforeau, L., Hiet, M.S., Aublin-Gex, A., Agaugue, S., Meiffren, G., Pradezynski, F., Faria, B.F., Chantier, T. et al. (2008). Hepatitis C virus infection protein network. *Mol Syst Biol* 4, 230.

Fan, R., Vermesh, O., Srivastava, A., Yen, B.K., Qin, L., Ahmad, H., Kwong, G.A., Liu, C.C., Gould, J., Hood, L. et al. (2008). Integrated barcode chips for rapid, multiplexed analysis of proteins in microliter quantities of blood. *Nat Biotechnol* 26(12), 1373–1378.

Glatter, T., Wepf, A., Aebersold, R. and Gstaiger M. (2009). An integrated workflow for charting the human interaction proteome: insights into the PP2A system. *Mol Syst Biol* 5, 237.

Holmes, E., Loo, R.L., Stamler, J., Bictash, M., Yap, I.K., Chan, Q., Ebbels, T., De Iorio, M., Brown, I.J., Veselkov, K.A. et al. (2008). Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* 453(7193), 396–400.

Hwang, D., Lee, I.Y., Yoo, H., Gehlenborg, N., Cho, J.H., Petritis, B., Baxter, D., Pitstick, R., Young, R., Spicer, D. et al. (2009). A systems approach to prion disease. *Mol Syst Biol* 5, 252.

Jones, S., Zhang, X., Parsons, D.W., Lin, J.C., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Kamiyama, H., Jimeno, A. et al. (2008). Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* 321(5897), 1801–1806.

Mitchell, P.S., Parkin, R.K., Kroh, E.M., Fritz, B.R., Wyman, S.K., Pogosova-Agadjanyan, E.L., Peterson, A., Noteboom, J., O'Briant, K.C., Allen, A. et al. (2008). Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A* 105(30), 10513–10518.

Parsons, D.W., Jones, S., Zhang, X., Lin, J.C., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Siu, I.M., Gallia, G.L. et al. (2008). An integrated genomic analysis of human glioblastoma multiforme. *Science* 321(5897), 1807–1812.

Shipitsin, M., Campbell, L.L., Argani, P., Weremowicz, S., Bloushtain-Qimron, N., Yao, J., Nikolskaya, T., Serebryiskaya, T., Beroukhim, R., Hu, M. et al. (2007). Molecular definition of breast tumor heterogeneity. *Cancer Cell* 11(3), 259–273.

Skog, J., Wurdinger, T., van Rijn, S., Meijer, D.H., Gainche, L., Sena-Esteves, M., Curry, W.T., Jr., Carter, B.S., Krichevsky, A.M. and Breakefield, X.O. (2008). Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. *Nat Cell Biol* 10(12), 1470–1476.

Wood, L.D., Parsons, D.W., Jones, S., Lin, J., Sjoblom, T., Leary, R.J., Shen, D., Boca, S.M., Barber, T., Ptak, J. et al. (2007). The genomic landscapes of human breast and colorectal cancers. *Science* 318(5853), 1108–1113.

Zethelius, B., Berglund, L., Sundstrom, J., Ingelsson, E., Basu, S., Larsson, A., Venge, P. and Arnlov, J. (2008). Use of multiple biomarkers to improve the prediction of death from cardiovascular causes. *N Engl J Med* 358(20), 2107–2116.

REFERENCES

Adler, A.S. and Chang, H.Y. (2006). From description to causality: Mechanisms of gene expression signatures in cancer. *Cell Cycle* 5, 1148–1151.

Anderson, N.L. and Anderson, N.G. (2002). The human plasma proteome: History, character, and diagnostic prospects. *Mol Cell Proteomics* 1, 845–867.

Bailey, R.C., Kwong, G.A., Radu, C.G., Witte, O.N. and Heath, J.R. (2007). DNA-encoded antibody libraries: A unified platform for multiplexed cell sorting and detection of genes and proteins. *J Am Chem Soc* 129, 1959–1967.

Barrier, A., Boelle, P.Y., Roser, F., Gregg, J., Tse, C., Brault, D., Lacaine, F., Houry, S., Huguier, M., Franc, B. et al. (2006). Stage II colon cancer prognosis prediction by tumor gene expression profiling. *J Clin Oncol* 24, 4685–4691.

Buttner, P., Mosig, S., Lechtermann, A., Funke, H. and Mooren, F.C. (2007). Exercise affects the gene expression profiles of human white blood cells. *J Appl Physiol* 102, 26–36.

Buyse, M., Loi, S., van't Veer, L., Viale, G., Delorenzi, M., Glas, A.M., d'Assignies, M.S., Bergh, J., Lidereau, R., Ellis, P. et al. (2006). Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst* 98, 1183–1192.

Czernin, J. and Phelps, M.E. (2002). Positron emission tomography scanning: Current and future applications. *Annu Rev Med* 53, 89–112.

Dai, H., van't Veer, L., Lamb, J., He, Y.D., Mao, M., Fine, B.M., Bernards, R., van de Vijver, M., Deutsch, P., Sachs, A. et al. (2005). A cell proliferation signature is a marker of extremely poor outcome in a subpopulation of breast cancer patients. *Cancer Res* 65, 4059–4066.

De Cecco, L., Marchionni, L., Gariboldi, M., Reid, J.F., Lagonigro, M.S., Caramuta, S., Ferrario, C., Bussani, E., Mezzanzanica, D., Turatti, F. et al. (2004). Gene expression profiling of advanced ovarian cancer: Characterization of a molecular signature involving fibroblast growth factor 2. *Oncogene* 23, 8171–8183.

Dhanasekaran, S.M., Barrette, T.R., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K.J., Rubin, M.A. and Chinnaiyan, A.M. (2001). Delineation of prognostic biomarkers in prostate cancer. *Nature* 412, 822–826.

Duarte, N.C., Becker, S.A., Jamshidi, N., Thiele, I., Mo, M.L., Vo, T.D., Srivas, R. and Palsson, B.O. (2007). Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci USA* 104, 1777–1782.

Dupuy, A. and Simon, R.M. (2007). Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J Natl Cancer Inst* 99, 147–157.

Foekens, J.A., Atkins, D., Zhang, Y., Sweep, F.C., Harbeck, N., Paradiso, A., Cufer, T., Sieuwerts, A.M., Talantov, D., Span, P.N. et al. (2006). Multicenter validation of a gene expression-based prognostic signature in lymph node-negative primary breast cancer. *J Clin Oncol* 24, 1665–1671.

Fujii, K., Nakano, T., Kawamura, T., Usui, F., Bando, Y., Wang, R. and Nishimura, T. (2004). Multidimensional protein profiling technology and its application to human plasma proteome. *J Proteome Res* 3, 712–718.

Fuller, G.N., Mircean, C., Tabus, I., Taylor, E., Sawaya, R., Bruner, J.M., Shmulevich, I. and Zhang, W. (2005). Molecular voting for glioma classification reflecting heterogeneity in the continuum of cancer progression. *Oncol Rep* 14, 651–656.

Furey, T.S., Cristianini, N., Duffy, N., Bednarski, D.W., Schummer, M. and Haussler, D. (2000). Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* 16, 906–914.

Gao, W.M., Kuick, R., Orchekowski, R.P., Misek, D.E., Qiu, J., Greenberg, A.K., Rom, W.N., Brenner, D.E., Omenn, G.S., Haab, B.B. et al. (2005). Distinctive serum protein profiles involving abundant proteins in lung cancer patients based upon antibody microarray analysis. *BMC Cancer* 5, 110.

Geman, D., d'Avignon, C., Naiman, D.Q. and Winslow, R.L. (2004). Classifying gene expression profiles from pairwise mRNA comparisons. *Stat Appl Genet Mol Biol* 3(Article19).

Giacomini, C.P., Leung, S.Y., Chen, X., Yuen, S.T., Kim, Y.H., Bair, E. and Pollack, J.R. (2005). A gene expression signature of genetic instability in colon cancer. *Cancer Res* 65, 9200–9205.

Glinsky, G.V., Glinskii, A.B., Stephenson, A.J., Hoffman, R.M. and Gerald, W.L. (2004). Gene expression profiling predicts clinical outcome of prostate cancer. *J Clin Invest* 113, 913–923.

Glinsky, G.V., Higashiyama, T. and Glinskii, A.B. (2004). Classification of human breast cancer using gene expression profiling as a component of the survival predictor algorithm. *Clin Cancer Res* 10, 2272–2283.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A. et al. (1999). Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* 286, 531–537.

Goncalves, A., Esterni, B., Bertucci, F., Sauvan, R., Chabannon, C., Cubizolles, M., Bardou, V.J., Houvenaegel, G., Jacquemier, J., Granjeaud, S. et al. (2006). Postoperative serum proteomic profiles may predict metastatic relapse in high-risk primary breast cancer patients receiving adjuvant chemotherapy. *Oncogene* 25, 981–989.

Haab, B.B., Dunham, M.J. and Brown, P.O. (2001). Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biol* 2, RESEARCH0004.

Halvorsen, O.J., Oyan, A.M., Bo, T.H., Olsen, S., Rostad, K., Haukaas, S.A., Bakke, A.M., Marzolf, B., Dimitrov, K., Stordrange, L. et al. (2005). Gene expression profiles in prostate cancer: Association with patient subgroups and tumour differentiation. *Int J Oncol* 26, 329–336.

Hood, L., Heath, J.R., Phelps, M.E. and Lin, B. (2004). Systems biology and new technologies enable predictive and preventative medicine. *Science* 306, 640–643.

Hu, Z., Hood, L. and Tian, Q. (2007). Quantitative proteomic approaches for biomarker discovery. *Proteomics*. In review.

Huber, W. and Mueller, F. (2006). Biomolecular interaction analysis in drug discovery using surface plasmon resonance technology. *Curr Pharm Des* 12, 3999–4021.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M. and Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci USA* 98, 4569–4574.

Ivshina, A.V., George, J., Senko, O., Mow, B., Putti, T.C., Smeds, J., Lindahl, T., Pawitan, Y., Hall, P., Nordgren, H. et al. (2006). Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer Res* 66, 10292–10301.

Joshi-Tope, G., Gillespie, M., Vastrik, I., D'Eustachio, P., Schmidt, E., de Bono, B., Jassal, B., Gopinath, G.R., Wu, G.R., Matthews, L. et al. (2005). Reactome: A knowledgebase of biological pathways. *Nucleic Acids Res* 33, D428–D432.

Kim, S., Dougherty, E.R., Shmulevich, I., Hess, K.R., Hamilton, S.R., Trent, J.M., Fuller, G.N. and Zhang, W. (2002). Identification of combination gene sets for glioma classification. *Mol Cancer Ther* 1, 1229–1236.

Koga, H., Kyo, M., Usui-Aoki, K. and Inamori, K. (2006). A chip-based miniaturized format for protein-expression profiling: the exploitation of comprehensively produced antibodies. *Electrophoresis* 27, 3676–3683.

Lathrop, J.T., Anderson, N.L., Anderson, N.G. and Hammond, D.J. (2003). Therapeutic potential of the plasma proteome. *Curr Opin Mol Ther* 5, 250–257.

Lee, H.J., Lee, E.Y., Kwon, M.S. and Paik, Y.K. (2006). Biomarker discovery from the plasma proteome using multidimensional fractionation proteomics. *Curr Opin Chem Biol* 10, 42–49.

Lin, B., White, J.T., Lu, W., Xie, T., Utleg, A.G., Yan, X., Yi, E.C., Shannon, P., Khrebtukova, I., Lange, P.H. et al. (2005). Evidence for the presence of disease-perturbed networks in prostate cancer cells by genomic and proteomic analyses: A systems approach to disease. *Cancer Res* 65, 3081–3091.

Liu, R., Wang, X., Chen, G.Y., Dalerba, P., Gurney, A., Hoey, T., Sherlock, G., Lewicki, J., Shedden, K. and Clarke, M.F. (2007). The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N Engl J Med* 356, 217–226.

Luo, J., Dunn, T., Ewing, C., Sauvageot, J., Chen, Y., Trent, J. and Isaacs, W. (2002). Gene expression signature of benign prostatic hyperplasia revealed by cDNA microarray analysis. *Prostate* 51, 189–200.

Lyon, L.A., Musick, M.D. and Natan, M.J. (1998). Colloidal Au-enhanced surface plasmon resonance immunosensing. *Anal Chem* 70, 5177–5183.

Ma, X.J., Wang, Z., Ryan, P.D., Isakoff, S.J., Barmettler, A., Fuller, A., Muir, B., Mohapatra, G., Salunga, R., Tuggle, J.T. et al. (2004). A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer Cell* 5, 607–616.

Miller, J.C., Zhou, H., Kwekel, J., Cavallo, R., Burke, J., Butler, E.B., Teh, B.S. and Haab, B.B. (2003). Antibody microarray profiling of human prostate cancer sera: Antibody screening and identification of potential biomarkers. *Proteomics* 3, 56–63.

Mulquiney, P.J. and Kuchel, P.W. (2003). *Modelling Metabolism with Mathematica Detailed Examples Including Erythrocyte Metabolism*. CRC Press, Boca Raton, Fla.

Nuyten, D.S., Kreike, B., Hart, A.A., Chi, J.T., Sneddon, J.B., Wessels, L.F., Peterse, H.J., Bartelink, H., Brown, P.O., Chang, H.Y. et al. (2006). Predicting a local recurrence after breast-conserving therapy by gene expression profiling. *Breast Cancer Res* 8, R62.

Olle, E.W., Sreekumar, A., Warner, R.L., McClintock, S.D., Chinnaiyan, A.M., Bleavins, M.R., Anderson, T.D. and Johnson, K.J. (2005). Development of an internally controlled antibody microarray. *Mol Cell Proteomics* 4, 1664–1672.

Orchekowski, R., Hamelinck, D., Li, L., Gliwa, E., vanBrocklin, M., Marrero, J.A., Vande Woude, G.F., Feng, Z., Brand, R. and Haab, B.B. (2005). Antibody microarray profiling reveals individual and combined serum proteins associated with pancreatic cancer. *Cancer Res* 65, 11193–11202.

Palsson, B. (2004). Two-dimensional annotation of genomes. *Nat Biotechnol* 22, 1218–1219.

Papadopoulou, E., Davilas, E., Sotiriou, V., Georgakopoulos, E., Georgakopoulou, S., Koliopanos, A., Aggelakis, F., Dardoufas, K., Agnanti, N.J., Karydas, I. et al. (2006). Cell-free DNA and RNA in plasma as a new molecular marker for prostate and breast cancer. *Ann NY Acad Sci* 1075, 235–243.

Park, E.S., Lee, J.S., Woo, H.G., Zhan, F., Shih, J.H., Shaughnessy, J.D. and Frederic Mushinski, J. (2007). Heterologous tissue culture expression signature predicts human breast cancer prognosis. *PLoS ONE* 2, e145.

Pawitan, Y., Bjohle, J., Amler, L., Borg, A.L., Egyhazi, S., Hall, P., Han, X., Holmberg, L., Huang, F., Klaar, S. et al. (2005). Gene expression profiling spares early breast cancer patients from adjuvant therapy: Derived and validated in two population-based cohorts. *Breast Cancer Res* 7, R953–R964.

Price, N.D. and Shmulevich, I. (2007). Biochemical and statistical network models for systems biology. *Curr Opin Biotechnol* 18, 365–370.

Price, N.D., Reed, J.L. and Palsson, B.O. (2004). Genome-scale models of microbial cells: Evaluating the consequences of constraints. *Nat Rev Microbiol* 2, 886–897.

Price, N.D., Schellenberger, J. and Palsson, B.O. (2004). Uniform sampling of steady-state flux spaces: Means to design experiments and to interpret enzymopathies. *Biophys J* 87, 2172–2186.

Price, N.D., Trent, J., El-Naggar, A.K., Cogdell, D., Taylor, E., Hunt, K.K., Pollock, R.E., Hood, L., Shmulevich, I. and Zhang, W. (2007). Highly accurate two-gene classifier for differentiating gastrointestinal stromal tumors and leiomyosarcomas. *Proc Natl Acad Sci USA* 104, 3414–3419.

Quackenbush, J. (2006). Microarray analysis and tumor classification. *N Engl J Med* 354, 2463–2472.

Ramaswamy, S., Tamayo, P., Rifkin, R., Mukherjee, S., Yeang, C.H., Angelo, M., Ladd, C., Reich, M., Latulippe, E. and Mesirov, J.P. (2001). Multiclass cancer diagnosis using tumor gene expression signatures. *Proc Natl Acad Sci USA* 98, 15149–15154.

Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S. et al. (2004). Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* 3, 1154–1169.

Ruvkun, G. (2001). Molecular biology. Glimpses of a tiny RNA world. *Science* 294, 797–799.

Sanchez-Carbayo, M., Socci, N.D., Lozano, J.J., Haab, B.B. and Cordon-Cardo, C. (2006). Profiling bladder cancer using targeted antibody arrays. *Am J Pathol* 168, 93–103.

Scherzer, C.R., Eklund, A.C., Morse, L.J., Liao, Z., Locascio, J.J., Fefer, D., Schwarzschild, M.A., Schlossmacher, M.G., Hauser, M.A., Vance, J.M. et al. (2007). Molecular markers of early Parkinson's disease based on gene expression in blood. *Proc Natl Acad Sci USA* 104, 955–960.

Simon, R. (2005). Roadmap for developing and validating therapeutically relevant genomic classifiers. *J Clin Oncol* 23, 7332–7341.

Singh, D., Febbo, P.G., Ross, K., Jackson, D.G., Manola, J., Ladd, C., Tamayo, P., Renshaw, A.A., D'Amico, A.V., Richie, J.P. et al. (2002). Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* 1, 203–209.

Smith, D.I. (2002). Transcriptional profiling develops molecular signatures for ovarian tumors. *Cytometry* 47, 60–62.

Solanky, K.S., Bailey, N.J., Beckwith-Hall, B.M., Davis, A., Bingham, S., Holmes, E., Nicholson, J.K. and Cassidy, A. (2003). Application of biofluid 1H nuclear magnetic resonance-based metabonomic techniques for the analysis of the biochemical effects of dietary isoflavones on human plasma profile. *Anal Biochem* 323, 197–204.

Song, S., Li, B., Wang, L., Wu, H., Hu, J., Li, M. and Fan, C. (2007). A cancer protein microarray platform using antibody fragments and its clinical applications. *Mol Biosyst* 3, 151–158.

Spentzos, D., Levine, D.A., Ramoni, M.F., Joseph, M., Gu, X., Boyd, J., Libermann, T.A. and Cannistra, S.A. (2004). Gene expression signature with independent prognostic significance in epithelial ovarian cancer. *J Clin Oncol* 22, 4700–4710.

Stark, C., Breitkreutz, B.J., Reguly, T., Boucher, L., Breitkreutz, A. and Tyers, M. (2006). BioGRID: A general repository for interaction datasets. *Nucleic Acids Res* 34, D535–D539.

Sun, B., Ranish, J.A., Utleg, A.G., White, J.T., Yan, X., Lin, B. and Hood, L. (2007). Shotgun glycopeptide capture approach coupled with mass spectrometry for comprehensive glycoproteomics. *Mol Cell Proteomics* 6, 141–149.

Tan, A.C., Naiman, D.Q., Xu, L., Winslow, R.L. and Geman, D. (2005). Simple decision rules for classifying human cancers from gene expression profiles. *Bioinformatics* 21, 3896–3904.

Thiele, I., Price, N.D., Vo, T.D. and Palsson, B.O. (2005). Candidate metabolic network states in human mitochondria. Impact of diabetes, ischemia, and diet. *J Biol Chem* 280, 11683–11695.

Thomassen, M., Tan, Q., Eiriksdottir, F., Bak, M., Cold, S. and Kruse, T.A. (2007). Prediction of metastasis from low-malignant breast cancer by gene expression profiling. *Int J Cancer* 120, 1070–1075.

Usui-Aoki, K., Shimada, K., Nagano, M., Kawai, M. and Koga, H. (2005). A novel approach to protein expression profiling using antibody microarrays combined with surface plasmon resonance technology. *Proteomics* 5, 2396–2401.

van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T. et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415, 530–536.

van de Vijver, M.J., He, Y.D., van't Veer, L.J., Dai, H., Hart, A.A., Voskuil, D.W., Schreiber, G.J., Peterse, J.L., Roberts, C., Marton, M.J. et al. (2002). A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347, 1999–2009.

Wang, X., Yu, J., Sreekumar, A., Varambally, S., Shen, R., Giacherio, D., Mehra, R., Montie, J.E., Pienta, K.J., Sanda, M.G. et al. (2005). Autoantibody signatures in prostate cancer. *N Engl J Med* 353, 1224–1235.

Wang, Y., Klijn, J.G., Zhang, Y., Sieuwerts, A.M., Look, M.P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M.E., Yu, J. et al. (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 365, 671–679.

Weigelt, B., Hu, Z., He, X., Livasy, C., Carey, L.A., Ewend, M.G., Glas, A.M., Perou, C.M. and Van't Veer, L.J. (2005). Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer. *Cancer Res* 65, 9155–9158.

Wilson, J.F. (2006). The rocky road to useful cancer biomarkers. *Ann Intern Med* 144, 945–948.

Xu, L., Tan, A.C., Naiman, D.Q., Geman, D. and Winslow, R.L. (2005). Robust prostate cancer marker genes emerge from direct integration of inter-study microarray data. *Bioinformatics* 21, 3905–3911.

Yang, S., Undar, A. and Zahn, J.D. (2006). A microfluidic device for continuous, real time blood plasma separation. *Lab Chip* 6, 871–880.

Zimmermann, M., Delamarche, E., Wolf, M. and Hunziker, P. (2005). Modeling and optimization of high-sensitivity, low-volume microfluidic-based surface immunoassays. *Biomed Microdevices* 7, 99–110.

RECOMMENDED RESOURCES

Website

www.systemsbiology.org Contains overviews of systems biology discipline, particularly in "Systems Biology in Depth" section of the website.

Article

Hood, L., Heath, J.R., Phelps, M.E. and Lin, B. (2004). Systems biology and new technologies enable predictive and preventative medicine. *Science* 306, 640–643.

CHAPTER

Electronic Medical Records in Genomic Medicine Practice and Research

Glenn S. Gerhard, Robert D. Langer, David J. Carey and Walter F. Stewart

INTRODUCTION

The implementation of electronic medical records (EMRs) is expected to substantially improve the quality and efficiency of health care and provide an important vehicle to advance patient-centered personalized care, the *sine quo non* of genomic medicine. In the United States, the use of EMRs in care delivery is expanding rapidly, especially among large integrated health delivery systems. The amount of clinically relevant genomic data and the number of resources devoted to research on genomic medicine are increasing in parallel. However, relatively few publications have yet addressed the use of EMRs in genomic medicine, although the need for such integration has been clearly established (Martin-Sanchez et al., 2004).

In this chapter, the role of EMRs in the practice of genomic medicine and their use for genomic research will be considered. The need for digital approaches to storing, processing, and using information is driven by the growing density of genomic data, whether derived from gene expression profiling, single nucleotide polymorphism (SNP) genotyping, or DNA sequencing, and from the pace of discovery. Moreover, the complexity of clinical decision-making will change commensurately as genetic data relevant to diagnosis and therapeutic interventions increases. Historically, the disciplines of medical (or clinical) informatics and biological informatics (or bioinformatics) have both evolved to address the needs of genomic medicine practice and research (Maojo and Kulikowski, 2003). Differences between the two communities have resulted in an attempt to integrate the approaches (Altman, 2000), with some proposing that health care organizations should be more proactive in collaborating with other public and private sector organizations in the development of an EMR containing both clinical and genomic data (Groen et al., 2005). These are laudable goals, but most of the work published thus far has come from those who have focused on the manipulation and analysis of genomic data for the clinical setting. However, the major factors driving the development of EMRs, such as quality measures, efficiency, and reimbursement, are not directly related to genomic medicine.

This chapter is written from the perspective of those involved in the use of EMRs in large integrated delivery systems that make use of data and populations for research. While the US health system is fragmented, the large health systems that are adopting EMRs are becoming increasingly integrated, especially in adopting and implementing practice standards (Orlova et al., 2005). A significant window of opportunity thus exists to incorporate various aspects of genomic medicine into the development of these emerging systems. The data relevant to genomic medicine are highly diverse and applicable to essentially all medical specialties. As a result, multiple approaches and solutions will likely need to be developed for the myriad of needs in genomic medicine research and practice (Stein, 2003; Sujansky, 2001).

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

EMRs AND GENOMIC MEDICINE CLINICAL PRACTICE

The EMR is a longitudinal electronic record of patient health information generated by encounters in any care delivery setting, encompassing patient demographics, progress notes, problem lists, medications, vital signs, past medical history, immunizations, laboratory data, and radiology reports. The EMR substitutes for traditional paper-based "charts". The EMR has the ability to generate a complete record of a clinical patient encounter, as well as supporting other care-related activities directly or indirectly, such as evidence-based decision support, quality management, and outcomes reporting. It may also be part of a solution to addressing a decades old problem of decreasing the long lag time that exists before evidence-based medical knowledge is routinely used in clinical care (Stewart et al., 2007). This may be of tremendous importance for the rapid progress being made in genomic medicine. In this section, the integration and use of genomics data in EMRs for the clinical practice of genomic medicine is discussed.

Example of an EMR

The Geisinger Clinic, a large integrated health system in northeastern and central Pennsylvania, initiated installation of an EMR (i.e., EpicCare) in 1996, which was completed (i.e., completely paperless operations) in all community practice sites and specialty clinics by 2001. The EMR contains patient information from a variety of sources that is routinely integrated into a common database and includes age, sex, height, weight, and other demographics; lifestyle data (e.g., smoking, alcohol, etc.); clinical measures (e.g., BP, pulmonary function, cardiac information); digital imaging (MRI, CT, X-ray); all orders (i.e., labs, prescriptions, imaging, procedures) that require at least one clinical indication (i.e., ICD-9 code); clinical notes, which are increasingly created using smart-sets, or structured protocols; and laboratory measures (including results of genetic and molecular tests). Patients have on-line access to their EMR in the EpiCare system through the My Geisinger Internet portal. With My Geisinger patients can securely access their EMR, schedule visits, and email their doctor. Such web-based communications represent future opportunities for genomic medicine research. To date, the Geisinger clinic EMR database contains information on more than 2.2 million patients.

Integration of Genomic Medicine Data into the EMR

EMRs make use of relational database structures and utilities to access and display data to facilitate medical care and clinical decision-making. Although most observers and users consider the EMR to be a single application, in reality there are multiple data feeds from disparate sources that comprise the EMR. Some data sources, for example laboratory information systems, are usually integrated with the main EMR application, while others, such as radiology, histopathology, and molecular data are incompletely integrated or are resident in entirely separate databases. Genomic data currently used for clinical practice are often generated by hospital, reference, or specialty laboratories that produce a wide variety of reporting formats, data types, and nomenclature (Ogino et al., 2007). Integration of these data into the EMR thus requires the ability to assemble the different data sources into an accessible format. Customized systems are sometimes developed in-house using on-site information technology expertise that are motivated by institution- and/or laboratory-specific needs and allow for flexibility and control over how the software evolves. However, the tendency is to invest significant resources into these in-house solutions, which sometimes underestimate the effort needed and fail to meet the intended goals. The choice of integration architecture and format for data representation are thus significant decisions and have a number of important issues associated with them (Haas et al., 2001).

The disparate sources that generate genomic data often report text-based data that are distributed through laboratory information systems into the EMR. These text results are commonly discrete, that is genotype A or genotype B, rather than continuous, although quantitative results are also produced, such as the number of trinucleotide repeats in the gene (*FMR1*) responsible for many cases of fragile X syndrome (Murray et al., 1997). Genomics results may also be reported in unique formats developed by diagnostic companies in niche markets. The disparate nature of the data and data sources further complicates the growing trend of integrating genomics data with other laboratory and/or clinically relevant data to generate a comprehensive clinical report. For example, cytogenetic and molecular remissions are recommended in reporting standards for acute myelogenous leukemia (Cheson et al., 2003). With the tendency of health care organizations towards the delivery of care in a service line structure and/or disease-specific care model (e.g., heart hospitals, neuroscience centers, etc.), more comprehensive and integrated reporting of genomic medicine results will likely be needed. This will apply to both genomic and other types of data-intensive results, such as proteomic and metabalomic data (Table 12.1).

Currently most genomic data generated for clinical care are in the form of a discrete result. However, as the cost of full DNA sequencing continues to decline (Bennett et al., 2005), it is likely that discrete and highly relevant-specific results will be provided along with a large amount of DNA sequence data that has no known clinical significance (Mitchell and Mitchell, 2007). Ultimately, complete genome sequencing may become the most cost effective means of identifying sequence variants (Hutchison, 2007). It would not be unexpected in the future if DNA sequence determination became the standard of care for every patient, even though most of the sequence data may not be relevant to clinical decisions. Prior to that, many smaller iterations of DNA sequence data will be determined, for example, comprehensive mutation detection for *BRCA1* and *BRCA2* (Myriad Genetics, Salt Lake City, Utah, "BRAC Analysis for *BRCA1* and *BRCA2*, Breast and Ovarian Cancer, Comprehensive Analysis"). In this analysis, full sequence determination in both forward and reverse directions of approximately 5400 base pairs comprising 22 coding exons and approximately 750 adjacent base pairs in the non-coding intervening sequences (introns) of the *BRCA1*

TABLE 12.1 Examples of genomic and other related data types potentially relevant for inclusion in the EMR

- DNA sequence data
- Single nucleotide polymorphism (i.e., SNP) genotypes
- Multiple nucleotide polymorphism (e.g., insertion, deletion, repeat) genotypes
- Cell and/or tissue-specific microarray gene expression profiles
- Proteomic (e.g., mass spectroscopy) profiles
- Glycomic (i.e., carbohydrate) structural data
- Lipomic (i.e., lipid) profiles
- Metabolic phenotypes

gene is made. For *BRCA2*, full sequence determination in both forward and reverse directions of approximately 10,200 base pairs comprising 26 coding exons and approximately 900 adjacent base pairs in the non-coding intervening sequence (intron) is performed. Similar sequence analysis is offered for several other gene or genes that are associated with specific inherited cancer syndromes. Such testing generates an enormous amount of extra information that is not currently clinically relevant. However, as new discoveries are eventually translated into practice, ready access to relevant DNA data on patients may become necessary.

How to maintain such dense sequencing data is thus an important decision (Mitchell and Mitchell, 2007). For example, placing every base sequenced into the patient's medical record may clutter the record with predominantly clinically insignificant data. Alternatively, only bases that differ from reference data may need to be tracked. Reference data may change, which may alter the interpretation of patient data. The American College of Medical Genetics (ACMG) has previously recommended a minimum amount of information for inclusion in DNA sequencing reports (ACMG, 2000). Important considerations include reporting of all of the bases sequenced, as well as linked sequence reference data. Such recommendations are likely to evolve as more clinical experience is acquired with sequence data.

Family History Data

An integral component of genomic medicine is the application of family history to clinical care. Interest in collecting family history data as a routine part of care delivery is growing, as knowledge advances in linking family history of disease to patient risk. Many common diseases, such as coronary heart disease, diabetes, several types of cancer, osteoporosis, and asthma, are known to occur more frequently in some families than in others; thus the identification of at-risk individuals is important for high quality medical care. Family history data are also necessary for research aimed at identifying environmental and/or genetic factors that may underlie these common diseases (Hopper et al., (2005)). However, the acquisition of such data by most physicians is inadequate, in part because tools do not exist for efficient, reliable, and valid collection and analysis of such data (Guttmacher et al., 2004). Structured family history data are almost universally absent in current EMRs.

The need for the development of better family history tools has been highlighted by projects at the Centers for Disease Control (CDC) and by the US Surgeon General's Family History Initiative (Wolpert, 2005). However, these efforts have not directly addressed the integration of tools into the real-world scenario of busy physicians and a multiplicity of health record systems, and do not provide an adequate breadth of data capture necessary for research. Various paper family history tools are available, as are various computer-based tools for drawing pedigrees (Trager et al., 2007). Several EMRs incorporate free text documentation of family history, while others allow for the development of family history templates and support checklists for taking and documenting family history. However, currently available tools are inadequate for many situations, especially primary care (Rich et al., 2004). The need for new tools is therefore apparent, and several characteristics of the ideal family history tool have been suggested, including patient-completed (e.g., paper, desktop, telephone, or web input), adapted to patient age, gender, ethnicity, common conditions, elicits-patient specific concerns, brief, understandable, easy to use, compatible with multiple clinical applications (e.g., paper, EMR, personal digital assistant), contains clinical decision support, and prioritizes based on clinical significance (Rich et al., 2004). No such electronic family history tools have yet been developed, despite the availability of suitable technologies.

EMRs, Genomic Medicine, and Clinical Decision Support

Genomic medicine is accelerating the growth of medical knowledge. In contrast, the means by which knowledge is translated into clinical practice have not evolved to keep up with the accelerating growth. Translation of knowledge into practice relies on a century-old tradition, the direct education of clinicians, an approach that was useful when knowledge was relatively limited. Today, it is not possible for most physicians to stay abreast of medical knowledge, especially genomics, and provide state-of-the-art care. The nuances in clinical decision-making in genomic medicine already render many care scenarios complex. Access to an EMR and the ongoing codification of medical knowledge (i.e., Clinical Practice Guidelines or CPGs) will be essential to addressing this growing translation gap.

CPGs greatly facilitate, but are not sufficient for translating knowledge to practice. Almost 2000 active CPGs exist in the US National Guideline Clearinghouse (Kozma, 2006) and an individual CPG may encompass dozens to hundreds of clinical recommendations. These recommendations will rapidly expand in the era of genomic medicine, and thus codification of knowledge will be essential to increasing its access. EMRs offer a platform to translate codified knowledge into real-time actionable processes. Genomic data will need to be accessed with other patient data located in disparate locations within the EMR and evaluated in relation to a rule set. Real-time actionable recommendations will need to be created and be supported by an integrated and intuitive visual display of information, such as a set of orders or recommendations for the physician or other care provider.

An example of clinical decision-making using an EMR potentially relevant to genomic medicine is a process with a specific focus on cardiovascular risk management in primary care being developed at the Geisinger Health System. The process involves a decision support rules "engine", external to the EMR, in which patient data such as labs, questionnaire data for phenotyping, and other clinical measures (Figure 12.1) are extracted in real time and evaluated in relation to clinical rules. The process can be used to generate a span of outputs including orders, clinical notes, and medication lists, etc. (Figure 12.2). The current model involves rules for smoking cessation, control of LDL, control of blood pressure, and management of other cardiovascular risk factors. Together, the combination of rules for different risk factors translates into more than 100,000 possible ordersets. The process itself demonstrates that it is technically feasible to develop and implement a comprehensive set of rules for state-of-the-art management of a clinical condition. Genetic variants reported to predispose to coronary heart disease (Samani et al., 2007) may impact this process and could greatly increase the number of ordersets.

Figure 12.1 Screenshot of EMR (Epic Systems, Corporation, Verona, WI). The software uses a tab-based design for access to medical record information including encounters, clinic notes, lab and imaging results, orders, medications, and procedures. These data can be extracted and used to automatically generate orders or recommendations, such as laboratory tests or imaging procedures for health maintenance or disease management.

Figure 12.2 Screenshot of EMR (Epic Systems, Corporation, Verona, WI). Examples of output from problem list, immunizations, allergies, medications, and health maintenance fields.

EMRs AND GENOMIC MEDICINE RESEARCH

The EMR offers an exciting opportunity for genomic medicine research, both as a tool and as a data resource. It also represents an economically efficient means of obtaining phenotypic data and biosamples for generating genotypic data. Initial large-scale efforts to use genomic and clinical information for research, for example, deCode Genetics in Iceland, required the conversion of medical information from paper records into an electronic format (Hoffman, 2007). Even current large-scale efforts, such as the UK BioBank (Ollier et al., 2005), are based upon a mix of paper and electronic sources. The EMR thus represents a potentially large increase in efficiency for obtaining phenotypic data, and as described below, can also be an extremely efficient tool for patient recruitment and biosample acquisition.

EMRs and Genome-Wide Association Studies

The continued acceleration in genotyping technologies has enabled a dramatic increase in the use of genome-wide association studies (GWAS) to identify genetic variants underlying a variety of clinical phenotypes. However, the very large sample sizes and the logistics of collecting phenotype data and biological samples often require multi-center studies with coordinating centers and are very costly and time consuming to complete. Moreover, multiple replication studies are essential to verify initial GWAS findings. The long time period and substantial costs in initiating population-based studies limits the scope of such approaches.

As an alternative, biorepositories are being developed to provide the resources necessary for GWAS. In Iceland (Hakonarson et al., 2003), the United Kingdom (Ollier et al., 2005), and elsewhere (Austin et al., 2003a), efforts are being made to link biobanking activities to national health records. In the United States, however, there are challenges with devising a national biobanking strategy (Austin et al., 2003b; Hsieh, 2004; McCarty et al., 2007; Sanner and Frazier, 2007). The US population is very heterogeneous, highly mobile, and health care is highly fragmented. Information on patients is highly dispersed and not in an easily accessible form. Hence, GWAS and related validation studies are largely conducted in conventional single project formats (i.e., *de novo* data collection), where there is a long lead-time between data acquisition and analysis.

In the United States, integrated delivery systems that use EMRs may be critical to addressing logistical challenges to conducting large-scale GWAS studies. While health care in the United States is highly fragmented, approximately 10% of care is provided by integrated delivery systems. Nationwide, these systems are undergoing a quiet revolution through the introduction of EMRs. Efforts to standardize data elements for research purposes and data sharing are being pursued by the larger systems (Hornbrook et al., 2005). There may be substantial opportunities to accelerate genomics research in the United States by leveraging the growing number of systems that are both adopting EMRs and creating biorepositories. However, there are many unanswered questions regarding whether such resources can be effectively used for genomics research. At the individual system level, there are questions regarding the quality, specificity, and completeness of data for phenotyping. In addition, little is known about system-level challenges that influence the quality and completeness of data captured during clinical encounters (e.g., impacts upon workflow, technical, and business priorities). Moreover, in some systems, even if quality data are available, there may be substantial technical, ethical, or policy barriers that limit access or use. Finally, GWAS and follow-up replication studies will require collaboration of multiple systems to achieve sample sizes that are sufficiently large, but there are enormous data quality, technical, and policy challenges to pooling data across institutions.

Using the EMR to Access Populations for GWAS

Case-control studies and family based studies have been the two primary designs for GWAS, but the economic efficiencies of the case-control format conducted within health systems are significant. Patient populations and inherent selection bias also impact GWAS, which may vary substantially depending on the source (e.g., primary care, specialty care, outpatient, inpatient). Primary care populations tend to be similar to general population samples with the exception that younger individuals (especially males) and individuals from lower middle-income strata (i.e., not eligible for Medicaid and no access to health care benefits through work) are under-represented. Selection bias inherent to specialty care samples depends on the extent to which there are competing providers in the same market and whether a provider has local, regional, national, or international reach. Selection bias increases directly in relation to market reach. These potential biases will be reflected in clinical data available in EMRs, depending upon where and when the system was installed.

Genetic heterogeneity in a population may also arise due to population structure and recent admixture and may confound the results of genetic association studies in unrelated individuals, leading to a potential excess of both false positive and false negative results (Ziv and Burchard, 2003). The degree of confounding is dependent in part on how genetically different the subgroups are within the population. For example, among many contemporary populations, recent admixture between two or more genetically distinct subgroups has added to the genetic complexity. In the United States, this complexity is particularly pertinent for admixed populations such as Latino- and African-Americans (Parra et al., 1998). Within indigenous continental population groups, European subpopulations appear to be the most similar to each other (Bowcock et al., 1994; Mountain and Cavalli-Sforza, 1997). The ability of an EMR to distinguish among ethnicities within its patient base may impact the usability of the data for GWAS.

The Geisinger MyCode Biorepository

The creation of biorepositories that are closely linked with EMRs may be an extremely efficient approach to genomic medicine research, especially for GWAS that require DNA samples from large populations with corresponding robust phenotypic data. The challenges to establishing such biorepositories involve complex institutional, legal, and social issues that also must be addressed. However, the EMR opens new opportunities to address these concerns, such as patients' safety, rights, informed consent, and privacy.

The Geisinger Clinic MyCode biorepository project is perhaps the first large-scale biobanking initiative built around an EMR. In this project, the EMR is used as a tool for identifying and recruiting patients, obtaining blood samples, and retrieving phenotype data. The primary goal of the MyCode project is to bank blood/serum/DNA samples from patients on a large-scale coupled with access to their EMR data. MyCode was initiated in 2007 to leverage the resources of the Geisinger Health System, an integrated delivery system offering healthcare services to residents of 31 of Pennsylvania's 67 counties, with a significant presence in central and northeastern Pennsylvania between Pittsburgh and Philadelphia. Census data indicate that in most counties, the out-migration rate is less than 1% per year, thus providing an extremely attractive population for longitudinal studies. The population within the Geisinger Clinic catchment area is relatively homogeneous. Based on available historical, demographic, and self-reported ethnicity data, the Geisinger population is almost entirely Caucasian of mixed European ancestry, with only small contributions from other continental population groups, such as African, Asian, or Latino. The relatively lower degree of genetic variation among European subpopulations, and the length of time for intermingling of European subgroups in Pennsylvania, also serves to minimize population stratification. Such a population straddles genetically homogeneous populations such as Iceland or Finland with strong founder effects, and admixed populations, now present in many urban areas.

The MyCode patient recruitment process (Figure 12.3) begins with a search of the EMR for patients scheduled to visit an outpatient or designated specialty clinic in which a consenter will be stationed. Eligible patients are identified through a daily EMR report, while consent status (e.g., consent/did not consent/ not interested) is verified in the patient's EMR to prevent unnecessary requests for participation at future appointments. Patients are invited to participate in MyCode through direct interaction with a project consenter. After consent, a standing order for the additional MyCode blood draw is placed in

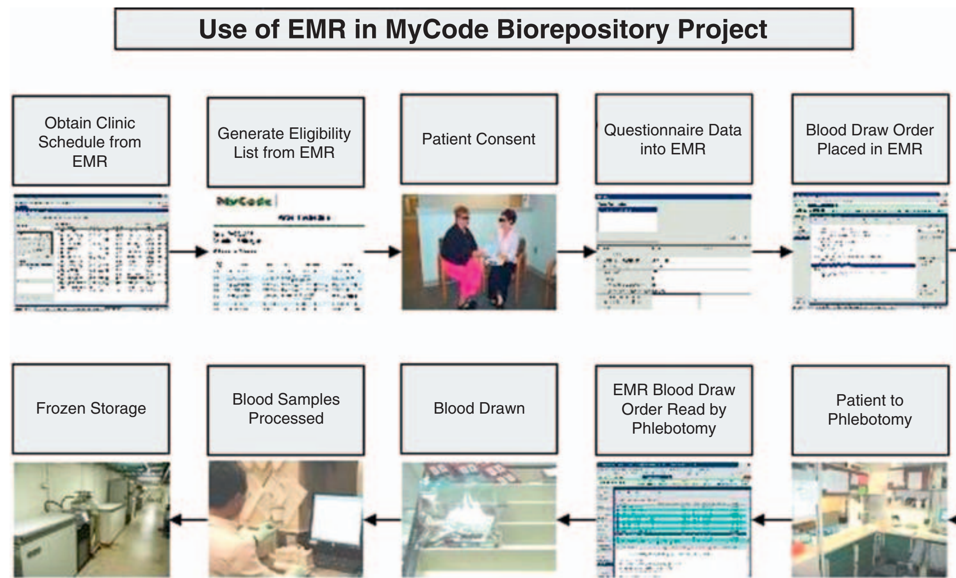


Figure 12.3 Use of the EMR in the MyCode Biorepository Project. The EMR functions as a tool at multiple points in the patient recruitment and consenting processes, and for acquisition of blood samples. The EMR also serves as the repository for phenotype data.

the EMR such that it is appended to the new order for blood work ordered by a physician at a future visit. With this strategy, blood is not specifically drawn for research which has resulted in high patient participation rates, as has been found in other studies (Sanner and Frazier, 2007). The sample collection process is highly efficient because it leverages the EMR (i.e., maintaining a standing order with an automatic protocol to activate the order), the laboratory infrastructure (e.g., standard protocols for phlebotomy, data systems, transportation of blood samples to a central location, etc.), and a passive low cost means of obtaining samples. Blood samples are transported from clinics across the Geisinger coverage region using the existing Geisinger Lab courier service to a central research laboratory that processes the samples for the MyCode biorepository. The EMR serves as the electronic enabler of the entire MyCode process.

The approach to creating the MyCode biorepository mitigates several methodological challenges with GWAS, including possible genotype errors, phenotype misclassification, and selection bias. While genotyping platforms are robust, DNA quality can affect genotyping results. Frozen blood and/or DNA are collected using the same protocol and processed using the same methodology. Errors in labeling patient samples are minimal, since an integrated clinical lab quality protocol is used for tracking samples from the point of phlebotomy to freezer storage. Phenotype misclassification is minimized through the use of careful case and control definition. Selection bias is minimized through the selection of cases and controls from the same population base.

Example of EMR Data Extraction

The EMR is a powerful and convenient source of phenotypic data for GWAS and genomic medicine research. A reference dataset has been created with a record for each patient in the Geisinger EMR population that includes demographics, physical measurements, personal medical history, health behaviors, laboratory tests, surgical and medical procedures, and medications prescribed. The dataset was developed to reflect the top 80% of elements most likely to be used for standard research purposes. Core data were extracted from the EMR using imbedded routines (known as Clarity), provided by the software vendor. Clarity tables can be manipulated using standard queries in SQL applications. The Clarity tables containing the specific elements in the data dictionary for a particular domain (e.g., physical measurements, or endocrine-related lab) are identified, and those fields are extracted into a text file. Typically, all instances of an element are pulled (e.g., the dataset contains all blood pressure measurements, or all glucose lab for a patient), not just the most recent or the first. The text file is read into SAS/STAT software (SAS Institute Inc., Cary, NC) and mapped to pre-defined fields. The resulting files are documented, cleaned and stored on a dedicated directory controlled by the biostatistics team. Analysis files are created by merging the appropriate data and applying any algorithms necessary for derived values.

As an example, this extraction of core data from the EMR can be used to obtain body mass index (BMI) data on over 270,000 adults (Table 12.2). Analyses were limited to adult patients 20 years of age and older. The mean and median BMIs were 28.4 and 27.2, respectively. Slightly over 30% of patients had a BMI ≤25 while 7.8% had a BMI of 40+. Patients with a BMI of 40+ were three times more likely to be female and more likely to be between 40 and 59 years of age (i.e., with 46.2% versus 29.6%) compared to those with a BMI ≤25. The relatively smaller proportion above age 60 and the gender bias may suggest a survival effect for this subgroup.

TABLE 12.2 Percent of adult patients by BMI and specific phenotypes

Phenotype	BMI Category			
	≤25 n = 80,861	>25, <30 n = 85,626	>30, <40 n = 83,298	40+ n = 21,187
Male	31.2	46.2	42.2	27.9
Female	68.8	53.8	57.8	72.1
Age <40	45.4	28.2	25.1	30.2
Age 40–59	29.6	36.5	40.7	46.2
Age 60+	25.0	35.3	34.2	23.6
Diagnosis of HTN	39.5	59.9	72.2	78.5
Diagnosis of LoHDL	18.8	28.2	39.8	47.8
Diagnosis of T2D	7.8	12.8	21.1	32.1
Diagnosis of HTN and LoHDL	13.3	23.4	34.4	42.5
Diagnosis of HTN and T2D	6.2	11.1	19.0	29.0
Diagnosis of LoHDL and T2D	5.3	9.0	15.3	23.6
Diagnosis of HTN, LoHDL, and T2D	4.9	8.3	14.4	22.2
No Diagnosis of HTN, LoHDL, or T2D	35.4	20.2	11.0	6.1

The prevalence of Type II diabetes (T2D), hypertension (HTN), and low HDL levels (LoHDL), and combinations of these, for each BMI subgroup was also determined. A patient was defined as having T2D or HTN if the condition was on the EMR Problem List (i.e., ICD-9 linked searchable fields) or if they had two encounters within 24 months where the condition(s) were listed as a reason for visit. HTN was also designated if there was any record of diastolic blood pressure of 90 mm Hg or more, or a systolic blood pressure of 140 mm Hg or more, or if the patient was taking an antihypertensive medication. T2D was designated if any of the following conditions were true: a random blood sugar of 200 or more, a hemoglobin A1C of 7% or higher, or if there was a prescription for insulin or an oral hypoglycemic medication. LoHDL was designated if the patient was male and had an HDL ≤35 or was female with an HDL ≤45 mg/dl. In general, prevalence of HTN, LoHDL, and T2D increase substantially in relation to BMI category. A diagnosis of all three conditions was found for 22.2% of those with a BMI of 40+, more than four times higher than the prevalence among those with normal BMI. In contrast, fully 35.4% of individuals with a BMI ≤25 had one of these conditions (as opposed to only 6.1% of patients with a BMI of 40+).

2009 UPDATE

The use of EMR data for genomic medicine research and practice is still in its early stages, but it is expanding rapidly and is evolving as a major platform to enable genomic medicine research. Our group published the first genomics paper based upon a comprehensive database of phenotypic information derived from an electronic medical record (Wood et al., 2008), supporting the feasibility of this approach and providing an initial example of an EMR-based genomic medicine research study. An important issue for such studies, as was emphasized, is the quality of the data, which may be influenced by practice variation among staff and clinicians (Yamada, 2008). Despite the potential disadvantages, the economic and temporal efficiency at which genotype–phenotype associations and replication studies can be performed with existing databases derived from EMRs coupled with biobanked blood/DNA samples is very promising.

Such promise has led to the initiation of large-scale biobanks at several health care organizations with long-standing electronic health records, similar to the Geisinger MyCode biorepository project described in this chapter. These projects plan to exploit the phenotypic information contained in electronic health records, but vary in their specific approaches. In one case, an "opt-out" model is used for patient consenting that is allowed for a de-identified mirror image of the EMR, a "synthetic derivative", to be linked to DNA extracted from leftover blood clinical samples (Roden et al., 2008). Algorithms for sample handling and procedures for de-identification were developed and validated in order to ensure acceptable error

rates. A strong advantage of this approach is the high rate of sample acquisition from the "opt out" consenting approach. In another large-scale biobanking effort, DNA samples will be sought from a patient population consisting of diverse social and genetic backgrounds (Nature News, 2009). Other projects have also been announced, so the next several years should see a blossoming of biorepositories in the United States and elsewhere.

The US National Institutes of Health has also realized the potential value of EMRs and has funded the electronic medical records and genomics (eMERGE) network (https://www.mc.vanderbilt.edu/victr/dcc/projects/acc/index.php/Main_Page), a five-member consortium formed to develop, disseminate, and apply approaches that combine DNA biorepositories with EMR systems for large-scale, high-throughput genetic research. The eMERGE network is designed to address whether EMRs can serve as resources for complex genomic analysis of disease susceptibility and therapeutic outcomes across diverse patient populations. In addition, the consortium will also address social and ethical issues, such as privacy and confidentiality. Issues involving the confidentiality, privacy, and security of EMR data for genomic studies have been addressed by others as well (Castle and DeBusk, 2008; McGuire et al., 2008).

Progress has also been made on the use of EMRs to collect family history data. The US Surgeon General's tool, "My Family Health Portrait" (https://familyhistory.hhs.gov), was re-issued with interoperability in an EMR environment using existing standards including the HL7 Family History Model, LOINC, SNOMED-CT, and HL7 Vocabulary. In addition, the tool includes information from the minimum core dataset for family health history as recommended by the American Health Information Community (Feero et al., 2008). At Geisinger, a pilot project was initiated to use the EMR web-portal, MyGeisinger, to collect family history information. A novel aspect of this approach is the automatic generation of an email alert message to a provider for patients indicating they have relevant family members with histories of premature heart attack, colon cancer, and breast cancer. The use of the EMR allows for the collection of actionable information relevant to preventative genomic medicine.

2009 UPDATE REFERENCES

Castle, D. and DeBusk, R. (2008). The electronic health record, genetic information, and patient privacy. *J Am Diet Assoc* 108(8), 1372–1374.

Feero, W.G., Bigley, M.B., Brinner, K.M., and Family Health History Multi-Stakeholder Workgroup of the American Health Information Community. (2008). New standards and enhanced utility for family health history information in the electronic health record: An update from the American Health Information Community's Family Health History Multi-Stakeholder Workgroup. *J Am Med Inform Assoc* 15(6), 723–728.

McGuire, A.L., Fisher, R., Cusenza, P., Hudson, K., Rothstein, M.A., McGraw, D., Matteson, S., Glaser, J. and Henley, D.E. (2008). Confidentiality, privacy, and security of genetic and genomic test information in electronic health records: Points to consider. *Genet Med* 10(7), 495–499.

Nature News (2009). Health organization lays plans for major biobank. *Nature* 457(7225), 16.

Roden, D.M., Pulley, J.M., Basford, M.A., Bernard, G.R., Clayton, E.W., Balser, J.R. and Masys, D.R. (2008). Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin Pharmacol Ther* 84(3), 362–369.

Wood, G.C., Still, C.D., Chu, X., Susek, M., Erdman, R., Hartman, C., Yeager, S., Blosky, M.A., Krum, W., Carey, D.J. et al. (2008). Association of chromosome 9p21 SNPs with cardiovascular phenotypes in morbid obesity using electronic health record data. *Genomic Med* 2(1–2), 33–43.

Yamada, Y. (2008). The electronic health record as a primary source of clinical phenotype for genetic epidemiological studies. *Genomic Med* 2(1–2), 5.

ACKNOWLEDGEMENTS

The authors wish to thank the Geisinger Clinic and the Pennsylvania State Department of Community and Economic Development for support of the MyCode project.

REFERENCES

ACMB Laboratory Practice Committee Working Group (2000). ACMG recommendations for Standards for interpretation of Sequence Variations. *Genet Med* 2(5), 302–303.

Altman, R.B. (2000). The interactions between clinical informatics and bioinformatics: A case study. *J Am Med Informat Assoc* 7(5), 439–443.

Austin, M.A., Harding, S. et al. (2003a). Genebanks: A comparison of eight proposed international genetic databases. *Community Genet* 6(1), 37–45.

Austin, M.A., Harding, S.E. et al. (2003b). Monitoring ethical, legal, and social issues in developing population genetic databases. *Genet Med* 5(6), 451–457.

Bennett, S.T., Barnes, C. et al. (2005). Toward the 1,000 dollars human genome. *Pharmacogenomics* 6(4), 373–382.

Bowcock, A.M., Ruiz-Linares, A. et al. (1994). High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368(6470), 455–457.

Cheson, B.D., Bennett, J.M. et al. (2003). Revised Recommendations of the International Working Group for Diagnosis, Standardization of Response Criteria, Treatment Outcomes, and Reporting Standards for Therapeutic Trials in Acute Myeloid Leukemia. *J Clin Oncol* 21(24), 4642–4649.

Groen, P., Wine, M. et al. (2005). Genomic Information Systems and Electronic Health Records (EHR). *Virtual Medical Worlds Monthly*. (October).

Guttmacher, A.E., Collins, F.S. et al. (2004). The family history—more important than ever. *N Engl J Med* 351(22), 2333–2336.

Haas, L.M., Schwarz, P.M. et al. (2001). DiscoveryLink: A system for integrated access to life sciences data sources. *IBM Syst J* 40(2).

Hakonarson, H., Gulcher, J.R. et al. (2003). deCODE genetics, Inc. *Pharmacogenomics* 4(2), 209–215.

Hoffman, M.A. (2007). The genome-enabled electronic medical record. *J Biomed Informat* 40(1), 44–46.

Hopper, J.L., Bishop, D.T. et al. (2005). Population-based family studies in genetic epidemiology. *Lancet* 366(9494), 1397–1406.

Hornbrook, M.C., Hart, G. et al. (2005). Building a virtual cancer research organization. *J Natl Cancer Inst* (35), 12–25.

Hsieh, A. (2004). A nation's genes for a cure to cancer: Evolving ethical, social and legal issues regarding population genetic databases. *Columbia J Law Soc Probl* 37(3), 359–411.

Hutchison, C.A., III (2007). DNA sequencing: Bench to bedside and beyond. *Nucleic Acids Res* 35: 6227–6237.

Kozma, C.M. (2006). The National Guideline Clearinghouse. *Manag Care Interface* 19(5). 43, 51.

Maojo, V. and Kulikowski, C.A. (2003). Bioinformatics and medical informatics: Collaborations on the road to genomic medicine?. *J Am Med Informat Assoc* 10(6), 515–522.

Martin-Sanchez, F., Iakovidis, I. et al. (2004). Synergy between medical informatics and bioinformatics: facilitating genomic medicine for future health care. *J Biomedl Informat* 37(1), 30–42.

McCarty, C.A., Nair, A. et al. (2007). Informed consent and subject motivation to participate in a large, population-based genomics study: The Marshfield Clinic Personalized Medicine Research Project. *Community Genet* 10(1), 2–9.

Mitchell, D.R. and Mitchell, J.A. (2007). Status of clinical gene sequencing data reporting and associated risks for information loss. *J Biomed Informat* 40(1), 47–54.

Mountain, J.L. and Cavalli-Sforza, L.L. (1997). Multilocus genotypes, a tree of individuals, and human evolutionary history. *Am J Human Genet* 61(3), 705–718.

Murray, J., Cuckle, H. et al. (1997). Screening for fragile X syndrome. *Health Technol Assess* 1(4), 1–71, i-iv.

Ogino, S., Gulley, M.L. et al. (2007). Standard mutation nomenclature in molecular diagnostics: Practical and educational challenges. *J Mol Diagn* 9(1), 1–6.

Ollier, W., Sprosen, T. et al. (2005). UK Biobank: From concept to reality. *Pharmacogenomics* 6(6), 639–646.

Orlova, A.O., Dunnagan, M. et al. (2005). Electronic health record-public health (EHR-PH) system prototype for interoperability in 21st century healthcare systems. *AMIA Annu Symp Proc*, 575–579.

Parra, E.J., Marcini, A. et al. (1998). Estimating African American admixture proportions by use of population-specific alleles. *Am J Hum Genet* 63(6), 1839–1851.

Rich, E.C., Burke, W. et al. (2004). Reconsidering the family history in primary care. *J Gen Intern Med* 19(3), 273–280.

Samani, N.J., Erdmann, J. et al. (2007). Genomewide association analysis of coronary artery disease. *N Engl J Med* 357(5), 443–453.

Sanner, J.E. and Frazier, L. (2007). Factors that influence characteristics of genetic biobanks. *J Nurs Scholarsh* 39(1), 25–29.

Stein, L.D. (2003). Integrating biological databases. *Nat Rev Genet* 4(5), 337–345.

Stewart, W.F., Shah, N.R. et al. (2007). Bridging the inferential gap: The electronic health record and clinical evidence. *Health Aff (Millwood)* 26(2), w181–w191.

Sujansky, W. (2001). Heterogeneous database integration in biomedicine. *J Biomed Informat* 34(4), 285–298.

Trager, E.H., Khanna, R. et al. (2007). Madeline 2.0 PDE: A new program for local and web-based pedigree drawing. *Bioinformatics* 23(14), 1854–1856.

Wolpert, C.M. (2005). Surgeon General launches new public health campaign: The family history initiative. *JAAPA* 18(1), 20–22.

Ziv, E. and Burchard, E.G. (2003). Human population structure and genetic association studies. *Pharmacogenomics* 4(4), 431–441.

RECOMMENDED RESOURCES

James M. Walker (Editor), Eric J. Bieber (Editor), Frank Richards (Editor), *Implementing an Electronic Health Record System* (Health Informatics) Publisher: Springer 2004.

Implementing an Electronic Health Record System addresses the range of issues and opportunities that implementing an electronic health records system (EHR) poses for any size of medical organization-from the small one-man operation to a large healthcare system. The book is divided into sections on preparation, support, implementation and a summary and prospects section, enabling the clinician to define the framework necessary to implement and evaluate a clinically effective EHR system. With the increasing involvement of clinicians in the day-to-day running of the practice, interest is now focused on EHR as a key area for improving clinical efficiency. This book uniquely provides the guidance a clinical team needs to plan and execute an effective EHR system within any clinical setting.

URL: http://www.hmoresearchnetwork.org/

CHAPTER

Online Health Information Retrieval by Consumers

Mark S. Boguski

INTRODUCTION

In an age of personalized medicine, nothing represents the *zeitgeist* more than individual consumers using the Internet and World Wide Web to seek medical and health information. According to surveys by the Jupiter organization and Harris Interactive, 71% of people who use the Internet also used it to seek health information in 2007, and this percentage, which represents an estimated 160 million people in the United States, had increased by 37% since 2005 (Anonymous, 2007b; Levy, 2007).

Consumers perform health information search (HIS) and retrieval (HIR) for themselves as well as for friends and family. Studies have shown that most of these consumers do not later discuss the information with a health-care provider and that, for many people, the Internet may be the primary or even sole source of health information. (Fox and Fallows, 2003). Seventy percent of people who obtain health information online say that it has influenced a decision about their treatment.

Clearly, it is important for health care professionals to understand how their patients find health information and the pitfalls associated with this activity. Indeed, given the challenges that consumers face in obtaining quality health care information from Internet sources, health-care providers will increasingly be in a position where they have to act as reviewers of information and as educators of patients who bring this information to their attention.

CHARACTERISTICS OF CONSUMER SEARCHES FOR HEALTH INFORMATION

A 2007 study by the Pew Internet & American Life Project (Fox, 2006) showed that the top 10 reasons American adults search online for health information are:

- a specific disease or medical condition;
- a particular medical treatment or procedure;
- diet, nutrition and supplements;
- exercise or fitness;
- prescription and non-prescription drugs;
- a particular doctor or hospital;
- alternative medicine or treatments;
- health insurance;
- mental health issues and
- environmental health hazards.

Zeng and Tse (Zeng and Tse, 2006) reported that, in performing searches, consumers use query terms that consist of "every-day language, technical terms (with or without knowledge of the underlying concepts) and various explanatory models, all influenced by psychosocial and cultural variations..." According to Lorence and Spink (Lorence and Spink, 2004), lay terminology is only partly effective in retrieving useful health

Copyright © 2010, Elsevier Inc.
All rights reserved.

information and often produces irrelevant or misleading information because, as reported by Zeng et al. (Zeng et al., 2006), "… the terms and concepts used by consumers often do not accurately reflect their information needs and therefore do not constitute effective queries."

Problems with such *consumer health vocabularies* may occur at various levels (Zeng et al., 2002):

- *Lexical* (form-level) mismatches (e.g., misspellings)
- *Semantic* misunderstanding (e.g., incomplete or misinterpretation of abbreviations or acronyms, over-generalization, redundancy)
- Misleading mental models, conceptual models that consumers employ either incorrectly and/or in a manner different than health care professionals.

Form-level (lexical) mismatches need no further explanation, as examples are easy to imagine. In the case of semantic mismatches, searches with general search engines may yield a large amount of irrelevant information. Consider the abbreviation "MS", which a consumer might use to mean *multiple sclerosis* but could also be interpreted to mean *Microsoft*, *mass spectrometry* or the state of *Mississippi* (Romacker et al., 2006). In another example, "CHF" would be interpreted as *congestive heart failure* by a physician whereas an international banker would interpret CHF as a unit of currency, the Swiss franc.

Other groups (Tse and Soergel, 2003) have similarly characterized problems with consumer health vocabularies on multiple levels, namely:

- *Shared forms/different concepts* (e.g., the phrase "the results of the [diagnostic] test were negative" could mean "unfavorable" in the mind of a patient but is usually a positive step forward in the mind of a physician.)
- *Different forms/shared concepts* (e.g., "blood cancer" is a lay term that a physician might refer to as a "hematologic malignancy." Likewise when a physician refers to "metastatic disease", the patient would say "the cancer has spread.")
- *Different forms/different concepts* (e.g., "Miracle cure" rather than a "statistically unlikely" or "idiopathic" remission of a disease, emergence from coma, etc.).

The most challenging HIS/HIR problems of all undoubtedly arise at the conceptual level because it is intimately tied to an individual's level of *health literacy* (McCray, 2005). Concept-level mismatches might include notions outside the framework of mainstream medicine such as homeopathy, acupuncture and other alternative medicine concepts. Another example is when a patient may focus on a body part (anatomical location) or symptom when a clinician might view the same problem in terms of pathophysiology or from a sub-specialty perspective. Even mainstream medical concepts are subject to significant cultural influences (Payer, 1996).

There have been a number of attempts and exploratory studies to map lay terminology to *controlled vocabularies* in a number of formal coding systems and comprehensive collections of medical terms such as *CTP, SNOMED, MeSH, ICD-9* and the *UMLS Metathesaurus* (Shortliffe et al., 2003). Furthermore, *ontologies* have been employed to perform *semantic* expansion on queries that may be too general or simplistic (see (Spasic et al., 2005) and below). The idea is to take queries consisting of consumer health terms and to translate or reformulate them into professional terms and qualifiers with the aim of improving query meaning, *precision* and *recall*. So far such experiments have met with only minor or mixed success (Plovnick and Zeng, 2004; Zeng et al., 2006). There is also the challenge of "back translating" from professional to consumer language the results of a search using a *reverse medical dictionary*. (See also Chapter 10.)

Two additional reasons for suboptimal performance of consumer HIS/HIR is short query length and low complexity. Consumer HIS/HIR queries tend to be very short and too general to be effective. Indeed one study found that 63% of consumer queries consisted of only a single word and only 10% contained more than two words (Zeng et al., 2002). Also, consumers are generally unaware of *stop words* that add little or nothing to search performance. Thus, limited knowledge of both medical vocabulary and query string search principles contributes to the construction of simplistic, ineffective queries. Another important aspect is the inability or unwillingness to construct complex queries (such as Boolean combinations of terms) even when such "advanced search" options are provided via simple check boxes. Users seem to prefer (or at least have become accustomed to) search and retrieval as an iterative process during which their goals are refined, focused or revised again and again (Fredin and Prabu, 1998) until either a satisfactory answer is obtained or they abandon the effort altogether due to frustration or fatigue. A precise, unambiguous and well-formed query would ideally produce the answer(s) one is looking for in the first iteration. This ideal assumes that the information sources containing the information exist and are accessible.

WHAT AND WHERE ARE CONSUMERS SEARCHING?

Equally important to how searches are formulated by consumers are the sources of the information that are being searched. As an editorial in the *British Medical Journal* put it: "…the internet has vastly increased the availability of information, but often what it offers is untailored, incomplete, irrelevant, and plain wrong" (Jones, 2003) although actual cases of harm have been difficult to document (Crocco et al., 2002). Online health information is problematic as it encompasses everything from evidence-based, peer-reviewed content in professional journals to advertiser-sponsored content and personal testimonials (Table 13.1). A consumer's online HIS/HIR experience may even be greatly influenced by non-substantive characteristics of a site such as an attractive and professional-looking appearance even though the quality of the underlying information is unknown.

Sixty-six percent of consumer HIS/HIR begin with a general search engine such as Google or Yahoo! (Fox, 2006), and the rest will start with one of a number of emerging or established

TABLE 13.1 Comparisons of Selected Online Health Information Sites

	Libraries, Publishers	**e-Magazines**	**Search portals**	**Social Networks**
Content	Curated, Aggregated	Commissioned, Aggregated	Distributed	User-generated
Audience	Healthcare Professionals	General public	General public	General public
Stage	Pre-Internet	Dot-com era	Beta	Pre-alpha or alpha
Revenue	Gov't-sponsored, Subscription, Advertising	Gov't-sponsored, Advertising, Services portal	Advertising	Advertising Services?
Examples	National Library of Medicine http://www.nlm.nih.gov/ Reed-Elsevier http://www.elsevier.com/wps/find/homepage.cws_home Macmillan/Nature http://us.macmillan.com/splash/publishers/nature.html Thompson http://www.thompson.com/public/library.jsp?cat=HEALTHCARE	National Institutes of Health http://health.nih.gov/ Healthline http://www.healthline.com/ WebMD http://www.webmd.com/default.htm RevolutionHealth http://www.revolutionhealth.com/	Healia http://www.healia.com/healia/ Mamma http://www.mammahealth.com/ Medstory http://www.medstory.com/ Kosmix http://www.kosmix.com/health	TauMed http://www.taumed.com/ Trusera http://www.trusera.com/corp

specialty sites that are surveyed in Table 13.1. These sites fall into four general categories (online libraries, e-magazines, search and social networking) that largely reflect their historical origins and/or stages of development. Although a complete description of the origins and development of these sites is beyond the scope of this chapter, one pioneering organization deserves special mention.

Online medical libraries date back to mid-1960s, when the US National Library of Medicine (NLM) experimented with teletypewriter machines and satellite communications (Schoolman and Lindberg, 1988). Searches at this time were not interactive but rather conducted by "batch processing." In the early 1970s, interactive searches of NLM's MEDLINE became possible through dedicated remote terminals in 10 regional and 14 large academic medical libraries (Atlas, 2000). During this era, users of these systems were trained librarians who learned specific query commands and retrieval protocols. It was not until after the advent of personal computers that online search of the medical literature became accessible to consumers. In 1986 a computer application named "Grateful Med," that ran on IBM-compatible DOS computers was produced by NLM, was distributed at a cost of $9.95 through the National Technical Information Service of the US Department of Commerce (Schoolman and Lindberg, 1988). Using Grateful Med and a 300 bits-per-second modem, users could use dial-up telephone lines to NLM computers and receive literature citations. The current descendents of these pioneering efforts are PubMed (Wheeler et al., 2007) for the professional literature and MedlinePlus (Miller et al., 2004) for consumer information.

The Dot-com era from 1995 to 2001 (2007) was characterized by experimentation with new sites and services, including commercial entities supported by sponsored advertising. Compared with online "libraries," which maintained content and user interfaces more oriented toward health care professions than consumers, Dot-com era sites introduced magazine-type articles and layouts focused on non-professional users, enabled by the development of web browsers such as NCSA Mosaic (later named Netscape Navigator (2007)). Somewhat prior to this, sites providing comprehensive ("horizontal") Internet search were mainly being used by academic researchers and computer professionals (e.g., (Boguski and Ouellette, 1995)). But gradually, at first, search engines of the time (e.g., AltaVista (2007)) began to index content that expanded into the consumer realm. The ascendance of horizontal Internet search by companies like Yahoo! (2007) and Google (2007) has recently led to the emergence of several start-up companies applying a similar approach to deep *vertical* (specialized) health care content on the web (Table 13.1). The latest experimentation with consumer-oriented, medically-oriented websites involves social-networking, following the example of more horizontal (although demographically differentiated) social-networking communities and sites (e.g., MySpace (2007) and Facebook (2007)).

In July 14, 2007 article in the *Wall Street Journal*, Borzo reported running a variety of searches on several vertical health sites and noted that many features were not self-evident and required repeated experimentation to uncover. Borzo's overall conclusion from this limited, non-scientific study was that conducting a useful search requires the consumer to run multiple queries on several sites and to then compare the results and reach their own consensus on the adequacy and usefulness of the retrieved information. These observations are consistent with more formal research (Fox, 2006). Borzo did not examine, however, the underlying sourcing and quality of the information. When one examines this aspect, a perplexing assortment of both public and propriety

TABLE 13.2 Survey* of Content of Selected "Vertical" Health Information Search Sites

Site	Primary content
Healia www.healia.com/healia/	www.pubmed.gov/ www.clinicaltrials.gov/
Mamma www.mammahealth.com/	www.emedicinehealth.com/ www.healthatoz.com/ www.mayoclinic.com/ www.medem.com/ www.medicinenet.net/ medlineplus.gov/ www.nhsdirect.nhs.uk/ www.webmd.com/
Medstory www.medstory.com/	http://www.breastcancer.org/ http://online.wsj.com/public/us
Kosmix www.kosmix.com/health	www.nlm.nih.gov www.americanheart.org www.cdc.gov www.mayoclinic.com health.nih.gov www.cancer.org www.revolutionhealth.com www.womentowomen.com www.4women.gov www.healthline.com

*Survey date was November 16, 2007.

information from both consumer-oriented sources as well as sources designed for medical professionals is revealed (Table 13.2).

Only a tiny fraction of Internet health information sites publish any sourcing and date-stamped information or other information quality indicators (Anonymous, 2007). Published guidelines for quality assurance and quality control are rare and consumers are left to evaluate the quality of the information based on whether or not they consider the site to be a trusted "brand." Given this situation, and the difficulty non-professionals have in constructing effective queries (see previous section), the best advice on Internet HIS/HIR that one can give to the consumer at the present time is to validate their findings with a health care professional. The US National Library of Medicine provides information for consumers to finding health information and includes a "Guide to Healthy Web Surfing" at their MedlinePlus site (Table 13.1).

PERSONALIZED GENOMICS FOR CONSUMERS

Personalized genomics involving DNA polymorphism scans (e.g., see (Weber, 2006)) is currently in a tumultuous gestational stage that will ultimately lead to a new way to teach and practice medicine (Childs et al., 2005 (see Chapter 1).) Consider the following examples:

- A father uses commodity DNA sequencing technologies and publicly-available medical databases to investigate the elusive cause of his daughter's genetic illness (Maher, 2007).
- A controversial scientist publishes his autobiography which is partially based on the complete sequence of his own genome (Venter, 2007).
- A Harvard geneticist launches a *Personal Genome Project* (Table 13.3) to encourage medical altruism and self-knowledge, makes his own biological material available for study and encourages others to volunteer to do likewise (Church, 2005).
- Several start-up companies (one of which is backed by Google) form to provide genotyping services as direct-to-consumer (DTC) businesses (Winslow, 2007) and Table 13.3.
- The cover story of *WIRED* magazine asserts that "a new $1000 DNA test can tell how you'll live – and die" (Goetz, 2007).

The ethical, legal and social issues arising in this milieu certainly signify a brave new world well beyond the scope of this chapter. But what are the medical issues? Kohane and colleagues (Kohane et al., 2006) consider the situation "a threat to genomic medicine" because the volume and complexity of genotypes

Table 13.3 Selected Sites of Interest for Personal Genomics*

Site	Services Offered	Cost
Not-for-Profit		
www.MyDaughtersDNA.org	"a forum for those searching for explanations and the help of the interested community of geneticists, patients, physicians, scientists and family members"	Not applicable
Personal Genome Project www.pgen.us	"focuses on the practical issues of recruiting and informing volunteers… a test bed for personalized medicine and new ways of interfacing with the research subjects"	Not applicable
Commercial Businesses		
23andMe Personal Genome Service www.23andme.com	"personal insight into ancestry, genealogy, and inherited traits"	Not available
Navigenics www.navigenics.com	"a personalized genetic analysis that, combined with relevant health and wellness information, enables a far more personalized health strategy for each individual"	$2500
deCODE genetics www.decode.com	"Subscribers … can take their genome and examine it in the context of the literature"	$985 (Introductory, promotional price)

*Information current as of November 16, 2007.

and their statistical, phenotypic associations will undoubtedly lead to a plethora of incidental, "abnormal" findings that will be pursued at great cost but little benefit by the patients and their physicians. A person's genotype will become an "incidentalome" – analogous to the *incidentalomas* (Mirilas and Skandalakis, 2002) recognized by a previous generation of physicians but unbelievably more complex. Kohane and colleagues call for several key actions including the creation of information systems for "estimating and explaining the risks associated with various incidental genomic findings" (Kohane et al., 2006). They envisioned such systems as being used by medical professionals "…in the clinic and at the bedside." However, it is likely that consumer versions will lead the way as patients "surf the web" trying to understand the implications of their genotypes and manage their personal health accordingly. Dr. Google has office hours 24/7.

2009 UPDATE

A New York Times' article by Tara Parker-Pope entitled *You're sick. Now what? Knowledge is power*, discusses the use of the Internet by consumers seeking health information (Parker-Pope, 2008). In this article, Dr. Marisa Weiss, founder of breastcancer.org, states: "I don't think people have a choice – it's mandatory…The time you have to spend with your doctor is getting progressively shorter, yet there's so much more to talk about. You have to prepare for this important meeting." This situation represents a great potential for enhancing the value of clinical encounters between patients and providers and improving the overall efficiency of the healthcare system, provided that the central challenge of locating and retrieving reliable and personalized health information can be met.

The sources that consumers might consult in preparation for important meetings with their physicians remain highly variable and are rapidly changing. For example, of the new health search portals listed in Tables 13.1 and 13.2, the *MammaHealth* search service is no longer available, *MedStory* has not developed further, and *Kosmix* has split its health coverage into an odd assortment of subcategories such as Health Law, Fashion Health and films, magazines, and even rock bands that have the term "health" in their titles. The general search engines, particularly Google and Microsoft Live Search, have improved their medical search algorithms using a mechanism for query refinement (see Table 13.4) and by apparently "hard-coding" some links to credible, authoritative sources.

Treatment Information

Information on drugs and other treatments from the US Food and Drug Administration (FDA) has historically been designed not for patients but for "learned intermediaries," that is physicians and other healthcare professionals (Goetz and Growden, 2008). Recent studies by the Institute of Medicine (IOM) of the US National Academy of Sciences, however, have stressed the need to "help the public get the accurate, science-based

TABLE 13.4 Results from query term "diabetes"

Google			
Refine results for diabetes			
Treatment	Tests/diagnosis	For patients	From medical authorities
Symptoms	Causes/risk factors	For health professionals	Alternative medicine
Microsoft live search			
Health results for diabetes			
Conditions • Symptoms • Drugs & substances • Personal health			
Treatment • Diagnosis • Alternative medicine • Nutrition			

information they need to use medicines and foods to improve their health" and specifically recommended "that Congress enact legislation to establish a new FDA advisory committee on communication with patients and consumers" (Baciu et al., 2007). Moreover, the same report noted that "patients historically have been left out of the loop in much of the communication…" and also noted "…a shift in the formerly passive role of the patient…" A separate but related study from the IOM reported that "patients often leave the doctor's office with very little information about what drugs they are taking, what they are used for, and how they should be taken," and "few patients are given any sort of printed material about their medications" (Hernandez, 2007).

As a partial solution to these problems, the National Library of Medicine, working with the FDA, provides two web pages (dailymed.nlm.nih.gov/dailymed/about.cfm and www.nlm.nih.gov/medlineplus/druginformation.html), but these sites are extremely limited in their search capabilities, permitting only browsing or alphabetical searching by the first letter of a drug's name.

Personally Controlled Electronic Health Records

Within the past two years, Microsoft Corporation (Lohr, 2007), Google (Lohr, 2008), and a consortium of major employers called Dossia (McWilliams, 2007) have launched services to provide portable, electronic, personally controlled health records (PCHR) to consumers. With the motivation to reduce the risk of medical errors and spare the expense of missing records and unnecessary treatment (Goetz, 2007), it is hoped that doctors and medical practices will be largely converted to these from paper records by 2014, although this goal may be a bit too optimistic (Steenhuysen, 2008).

Electronic health records have the potential to permit consumers to customize or personalize their medical Internet search activities, in much the same manner in which online travel agencies such as *Orbitz* can tailor their products and services to individuals based on voluntarily submitted personal preferences.

Clinical and Recreational Genomics

In September 2008, the personal genotyping company 23andMe lowered its price for a genome-wide, single nucleotide polymorphism (SNP) scan from $999 to $399 and held a marketing event (dubbed "spit party" in reference to its saliva sample collection procedure) that was reported in the *Style Section* of the *New York Times* (Salkin, 2008). In this article, Dr. Alan Guttmacher, acting Director of the US National Human Genome Research Institute, has quoted a saying: "People think if you have the money to spend on this, why not by a test instead of a model train for Christmas." DTC genomics has leaped far outside the domain of learned intermediaries and it will undoubtedly be some years before the medical profession can catch up by assimilating genomics into medical education, training, and practice (Boguski, 2008). In the meantime, it's *caveat emptor* for consumers. Kaye (2008) discusses many of the important ethical, social, and legal implications of DTC genomics and its regulation.

The principal consumer genomics companies, 23andMe, deCODEme and Navigenics, currently genotype approximately 500,000 to 1,000,000 SNPs and report associations of up to 100 diseases and conditions ("traits," "characteristics"). 23andMe and deCODEme use the Illumina human 1M duo bead chip platform for their assays (www.illumina.com), whereas Navigenics uses the Affymetrix Gnome-wide SNP Array 6.0 technology (www.affymetrix.com). Thirty of these diseases or conditions are common to either two or all three of the companies (Tables 13.5 and 13.6). Each company has its own heuristic methods for determining and reporting risks of disease associated with these polymorphisms (e.g., relative risk, lifetime risk); the associations themselves are derived from published literature.

A number of people (including this author) have had themselves genotyped by all three of these companies in order to compare and contrast their services and features, and to assess the consistency of results. Davies (2008) had himself genotyped by the three major companies and also a fourth company, SeqWright (www.seqwright.com), that entered this marketplace in January 2008 and, like Navigenics, uses the 500,000 SNP Affymetrix GeneChip platform. Davies reported general agreement among results obtained from the different companies, but also noted some disagreements or discrepancies and their possible causes. A detailed analysis of my experiences is in preparation, but a preliminary account is available as a "Grand Rounds" presentation (Boguski, 2009).

TABLE 13.5 Disease Markers Detected by 23andMe, deCODEme and Navigenics

Diseases covered by all three companies[a]	SNP associations[b]
Age-related macular degeneration	rs1061147, rs547154, rs3750847
Atrial fibrillation	
Breast cancer	rs1219648, rs3803662
Celiac disease	rs2187668
Colorectal cancer	rs2187668
Crohn's disease	rs2066844, rs2066845, rs2066847, rs11209026, rs2241880, rs7714584, rs3197999, rs11190140, rs17234657, rs1893217, rs11805303, rs10210302, rs10761659
Glaucoma	
Heart attack	rs2383207
Lung cancer	rs8034191
Multiple sclerosis	rs6897932, rs3135388
Obesity	rs3751812
Prostate cancer	rs1447295, rs6983267, rs10505483, rs1859962, rs4430796
Psoriasis	rs10484554, rs3212227, rs11209026
Restless leg syndrome	rs3923809
Rheumatoid arthritis	rs6457617, rs11203366, rs2476601, rs3890745, rs2327832, rs3761847
Type 2 diabetes	rs7903146, rs1801282, rs5219, rs4402960, rs1111875, rs4712523, rs13266634, rs10012946, rs2383208

TABLE 13.6 Conditions Detected by Personal Genomics Companies

Diseases and other phenotypes covered by two out of three companies	SNP associations
Abdominal aortic aneurysm	
Alcohol flush reaction	rs671
Alzheimer's disease	
Asthma	
Bitter taste perception	rs713598
Intracranial aneurysm	
Lactose intolerance	rs4988235
Lupus (systemic lupus erythematosis)	rs9888739, rs7574865, rs2187668, rs10488631
Nicotine dependence	
Osteoarthritis	
Peripheral arterial disease	
Type 1 diabetes	rs3129934, rs3087243, rs1990760, rs3741208, rs1893217, rs2476601, rs3184504, rs725613
Venous thromboembolism	rs6025, i3002432

[a]Information current as of January 15, 2009.
[b]SNPs included in table are from the 23andMe website, specifically those pages that are accessible without a subscription. SNPs are referenced by their dbSNP ("rs") numbers – unique identifiers that can be used to locate extensive information about SNPs and their associations at www.ncbi.nlm.nih.gov/projects/SNP/. SNPedia (www.snpedia.com) is another useful resource.

As described earlier, the consumer genotyping companies report on up to 100 diseases and conditions that have been linked to genetic polymorphisms in published, genome-wide association studies (GWAS). However, there are many more than 100 published GWAS associations, and each company selects those it will report on according to their own criteria. For example, Davies describes the "decision heuristic" used by Navigenics to select diseases that it will cover (Davies, 2008). However, each of the companies also makes the primary genotype data available, and one company, SNPedia, has developed a software tool they call *Promethease* which allows one to generate their own reports based upon the primary data (Cariaso and Lennon, 2009).

Over the next several years, the technology platforms for personal genomics are expected to shift from SNP genotyping to complete DNA sequencing as so-called "Next Generation Sequencing" technologies continue to advance toward the "$1000 genome" (Mardis, 2006; Pollack, 2008).

2009 UPDATE REFERENCES

Baciu, A., Stratton, K. and Burke, S.P. (eds). (2007). *The Future of Drug Safety: Promoting and Protecting the Health of the Public*. National Academies Press, Washington DC.

Boguski, M.S. (2008). Genomics and clinical medicine. *N Engl J Med* 359(18), 1969–1970.

Boguski, M.S. (2009). *Clinical and Recreational Genomics: Personal Experiences with Direct-to-Consumer Genotyping*. Brigham and Womens Hospital, Boston, MA.

Cariaso, M. and Lennon, G. (2009). SNPedia.http://www.snpedia.com/index.php?title=Promethease.

Davies, K. (2008). Keeping score of your sequence. *Bio-IT World* 7, 16–21.

Goetz, T. (2007). Physcian, upgrade thyself. *New York Times*. http://www.nytimes.com

Goetz, R.B. and K.R. Growdon (2008). A defense of the learned intermediary doctrine. *Food Drug Law J* 63(2), 421–438.

Hernandez, L.M. (ed.). (2008). *Standardizing Medication Labels: Confusing Patients Less*. National Academies Press, Washington DC.

Kaye, J. (2008). The regulation of direct-to-consumer genetic tests. *Hum Mol Genet* 17(R2), R180–R183.

Lohr, S. (2007). Microsoft offers system to track health records. *The New York Times*. http://www.nytimes.com

Lohr, S. (2008). Google offers personal health records on the web. *The New York Times*. http://www.nytimes.com

Mardis, E.R. (2006). Anticipating the 1000 dollar genome. *Genome Biol* 7(7), 112.

McWilliams, G. (2007). Employers back development of web-based health records. *The Wall Street Journal*.

Parker-Pope, T. (2008). You're sick. Now what? Knowledge is power. *The New York Times*. http://www.nytimes.com

Pollack, A. (2008). The race to read genomes on a shoestring, relatively speaking. *New York Times*. http://www.nytimes.com

Salkin, A. (2008). When in doubt, spit it out. *The New York Times*. http://www.nytimes.com

Steenhuysen, J. (2008). US doctors making slow shift to digital records. *Reuters*. http://www.reuters.com

REFERENCES

Anonymous. (2007a). CDC Wonder DATA2010: Focus Area 11-Health Communication, from http://www.wonder.cdc.gov/data2010/.

Anonymous (2007b). Harris Poll Shows Number of "Cyberchondriacs" – Adults Who Have Ever Gone Online for Health Information – Increases to an Estimated 160 Million Nationwide. Harris Interactive.

Atlas, M.C. (2000). The rise and fall of the medical mediated searcher. *Bull Med Libr Assoc* 88(1), 26–35. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10658961

Boguski, M.S. and Ouellette, B.F.S. (1995). *Internet Basics for Biologists. Current Protocols in Molecular Biology*. John Wiley & Sons, New York.

Childs, B., Wiener, C. and Valle, D. (2005). A science of the individual: Implications for a medical school curriculum. *Annu Rev Genomics Hum Genet* 6, 313–330. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16124864

Church, G.M. (2005). The personal genome project. *Mol Syst Biol* 1, 0030. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16729065

Crocco, A.G., Villasis-Keever, M. and Jadad, A.R. (2002). Analysis of cases of harm associated with use of health information on the internet. *JAMA* 287(21): 2869–2871. http://www.ncbi.nlm.nih.gov/pubmed/12038937?ordinalpos=1&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_RVDocSum

Fox, S. (2006). Online Health Search 2006. Washington, DC, Pew Internet & American Life Project.

Fox, S. and Fallows, D. (2003). Internet Health Resources. Washington, DC, Pew Internet & American Life Project.

Fredin, E.S. and Prabu, D. (1998). Browsing and the hypermedia interaction cycle a model of self-efficacy and goal dynamics. *J Mass Comm Quarterly* 75(1), 35–54. http://www.eric.ed.gov/ERICWebPortal/custom/portlets/recordDetails/detailmini.jsp?_nfpb=true&_&ERICExtSearch_SearchValue_0=EJ569971&ERICExtSearch_SearchType_0=no&accno=EJ569971

Goetz, T. (2007). Your life decoded. *Wired* 283, 256–265.

Jones, G. (2003). Prescribing and taking medicines. *BMJ* 327(7419), 819. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14551062

Kohane, I.S., Masys, D.R. and Altman, R.B. (2006). The incidentalome a threat to genomic medicine. *JAMA* 296(2), 212–215. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16835427

Levy, M. (2007). Assessing consumers demand for health vertical search engines. *Jupiter Research Industry Focus Health*.

Lorence, D.P. and Spink, A. (2004). Semantics and the medical web a review of barriers and breakthroughs in effective healthcare query. *Health Info Libr J* 21(2), 109–116. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15191602

Maher, B. (2007). His Daughter's DNA. *Nature* 449(7164), 772–776.

McCray, A.T. (2005). Promoting health literacy. *J Am Med Inform Assoc* 12(2), 152–163. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15561782

Miller, N., Tyler, R.J. and Backus, J.E.B. (2004). MedlinePlus® the National Library of Medicine® brings quality information to health consumers. *Library Trends* 53(2), 375–388. http://hdl.handle.net/2142/1735

Mirilas, P. and Skandalakis, J.E. (2002). Benign anatomical mistakes incidentaloma. *Am Surg* 68(11), 1026–1028. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12455801

Payer, L. (1996). *Medicine and Culture*. Holt Paperbacks.

Plovnick, R.M. and Zeng, Q.T. (2004). Reformulation of consumer health queries with professional terminology a pilot study. *J Med Internet Res* 6(3), e27. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15471753

Romacker, M., Grandjean, N., Parisot, P., Kreim, O., Cronenberger, D., Vachon, T., Peitsch, M.C. (2006). The UltraLink an expert system for contextual hyperlinking in knowledge management. In *Computer Applications in Pharmaceutical Research and Development* (S. Ekins, ed.), Wiley Interscience, 729–753.

Schoolman, H.M. and Lindberg, D.A.B. (1988). The information age in concept and practice at the National Library of Medicine. *Annals of the American Academy of Political and Social Science* 495, 117–126.

Shortliffe, E.H., Perreault, L.E., Wiederhold, G. and Fagan, L.M. (2003). *Medical Informatics Computer Applications in Health Care and Biomedicine*. Springer, New York.

Spasic, I., Ananiadou, S., McNaught, J. and Kumar, A. (2005). Text mining and ontologies in biomedicine making sense of raw text. *Brief Bioinform* 6(3), 239–251. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16212772

Tse, T. and Soergel, D. (2003). Exploring medical expressions used by consumers and the media an emerging view of consumer health vocabularies. *AMIA Annu Symp Proc* 674-8. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14728258

Venter, J.C. (2007). *A Life Decoded: My Genome My Life*. Viking, New York.

Weber, J.L. (2006). Clinical applications of Genome Polymorphism Scans. *Biol Direct* 1, 16. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16756678

Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Edgar, R., Federhen, S. et al. (2007). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 35(Database issue), D5–D12. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17170002

Winslow, R. (2007). Is there a heart attack in your future. *New York Times*.

Zeng, Q., Kogan, S., Ash, N., Greenes, R.A. and Boxwala, A.A. (2002). Characteristics of consumer terminology for health information retrieval. *Methods Inf Med* 41(4), 289–298. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12425240

Zeng, Q.T., Crowell, J., Plovnick, R.M., Kim, E., Ngo, L. and Dibble, E. (2006). Assisting consumer health information retrieval with query recommendations. *J Am Med Inform Assoc* 13(1), 80–90. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16221944

Zeng, Q.T. and Tse, T. (2006). Exploring and developing consumer health vocabularies. *J Am Med Inform Assoc* 13(1), 24–29. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16221948

WIKIPEDIA REFERENCES

(2007). AltaVista. Wikipedia.

(2007). Dot-com bubble. Wikipedia.

(2007). Facebook. Wikipedia.

(2007). Google. Wikipedia.

(2007). Mosaic (web browser). Wikipedia.

(2007). MySpace. Wikipedia.

(2007). Yahoo! Wikipedia.

This page intentionally left blank

Translational

Section 3

14. Translational Genomics: From Discovery to Clinical Practice

15. Pharmacogenetics and Pharmacogenomics

16. Clinical Implementation of Clinical Implementation of Translational Genomics

17. The Role of Genomics in Enabling Prospective Health Care

18. From Sequence to Genomic Medicine: Genome Policy Considerations

19. Federal Regulation of Genomic Medicine

20. Economic Issues and Genomic Medicine

21. Public Health Genomics

This page intentionally left blank

CHAPTER

Translational Genomics: From Discovery to Clinical Practice

Geoffrey S. Ginsburg

INTRODUCTION

The use of genomic technologies to improve medical decision-making so as to achieve the goal of personalized medicine is poised to transform health care (Ginsburg and McCarthy, 2001; Ginsburg et al., 2005; Snyderman and Williams, 2003; Willard et al., 2005). New knowledge from the study of genomes and their by-products now has permitted the development of predictors of disease predisposition ("who is at risk"), prognosis ("who to treat") and therapeutic response ("how to treat") in individual patients and thus provides an opportunity to employ these predictors in present day practice. Many of these opportunities are highlighted in this book in which an astounding number of genome-based discoveries are defined with the potential to fundamentally alter how medicine is practiced in cardiovascular disease, oncology, metabolic disease, neuropsychiatric disorders, infectious disease and diseases with a basic inflammatory component.

The rate of genome-based discovery is stunning. However, despite the vast amount of genomic information now in the scientific literature, the translation of basic scientific findings from the genome and its derivative products, RNA and protein, to daily use in medical practice has been slow (Graham et al., 2006; Sung et al., 2003). It now takes an average of 17 years for 14% of scientific discoveries to enter day-to-day practice (Westfall et al., 2007). For genomic discoveries, the path is undoubtedly more complex.

The research paradigm for translational genomics is inconsistent and nonsystematic; multiple, and in some cases non-linear, pathways to translation of genomics to clinical medicine exist. For example, diagnostic companies have traditionally "cherry-picked" a narrow repertoire of new genomics-based molecular diagnostics from academic laboratories and the literature. The pharmaceutical industry has developed "targeted therapies" such as trastuzamab, imatinib and erlotinib that use molecular diagnostic tests – some of which are based on genetic variation (Lynch et al., 2004) – to identify individual patients who are likely to benefit from the drug. The lack of a clear system of regulatory oversight of genetic and genomic tests has led to gaps and ambiguity in the pathways of translation, an area that the US Secretary's Advisory Committee on Genetics, Health and Society has recently addressed (SACGHS, 2007). Personal genomics and direct-to-consumer avenues are being opened to bring genomics directly to the public, thus circumventing the traditional health care delivery and regulatory channels (Feero et al., 2008; Hunter et al., 2008; Janssens et al., 2008; McGuire et al., 2007).

These issues speak to the need for a comprehensive agenda for translational research to move human genome discoveries into health practice. This chapter will focus on paradigms

Copyright © 2010, Elsevier Inc.
All rights reserved.

for translational genomics to human health as well as the various opportunities, challenges, and strategies that will enable the effective translation of human genome information to clinically relevant actions and outcomes.

A ROADMAP FOR TRANSLATION

Khoury and colleagues (2007) have developed a framework for the acceleration of human genome discoveries into health care and disease prevention. This is a logical first step in guiding genomic innovation to health care and more widespread use. The framework (Table 14.1) contemplates four phases of translational research from discovery to public health impact. T1 research seeks to move a basic genome-based discovery into a candidate health application (e.g., the association of a gene mutation with a health outcome). This phase of research starts after gene discovery and provides the early evidence that genomic information might support predictive testing, screening, diagnostic, prognostic or pharmacogenomic testing. T1 research will usually take advantage of observational studies or clinical trials where the correlation of the genomic information with health characteristics or outcome can first be made. T2 research assesses the value of a genomic application for health practice, leading to the development of evidence-based guidelines (e.g., prospective or retrospective assessment of the predictive value of the mutation test for the health outcome). This phase of research usually begins after the analytical validity of the test has been established and progresses to establish clinical utility, or impact on clinical decision-making. T3 research examines the pathways for delivery, dissemination, and diffusion of genomic testing into practice (e.g., defining the barriers to the use of the test in relevant populations). Public–private partnerships will usually be required to facilitate T3 research, and this research phase will also involve policy makers, regulatory bodies, professional organizations and insurers. Finally, T4 research assesses the impact on human health (e.g., outcomes studies across a broad and diverse population). This area of research is about the results of the implementation, as opposed to the implementation research itself (T3).

TABLE 14.1 The continuum of translation research in genomic medicine: Types of research

Translation research phase	Notation	Types of research
T1	Discovery to candidate health application	Phases I and II clinical trials; observational studies
T2	Health application to evidence-based practice guidelines	Phase III clinical trials; observational studies; evidence synthesis and guidelines development
T3	Practice guidelines to health practice	Dissemination research; implementation research; diffusion research; Phase IV clinical trials
T4	Practice to population health impact	Outcomes research (included many disciplines); population monitoring of morbidity, mortality, benefits and risk

Adapted from Khoury et al. (2007).

T1 research is illustrated in many of the clinical chapters in this book that highlight genomic discoveries with potential disease relevance. The molecular signatures described by Dave et al. (2006), Bullinger and Valk (2005) and Potti et al. (2006) to stratify and number of cancer phenotypes based on their molecular signature are illustrative. Examples of T2 research are found in Marcom et al. (2008) and Potti et al. (2006), in which a genomic-guided clinical trial design is described that uses a microarray-based molecular profile to assign patients to the various treatment arms in the trial. The clinical outcomes in the trial are dependent on the genomic signatures and the evidence for the genomic application to humans will be determined by these prospective studies. Examples of T3 research are described in Khoury et al. (2007) as surveys of physicians, internists and oncologists on the knowledge of genetic testing for susceptibility for breast and ovarian cancers. This information would provide support for the basis of strategies to enhance the dissemination of innovative genomic tools and information. T4 research is ongoing in areas such as *BRCA1* testing for hereditary breast and ovarian cancers or for HNPCC testing in colon cancer (see Chapters 35 and 36). In these early cases, the value of such testing is being assessed on a population level in terms of survival, quality of life and economic value. We are clearly in the early days of T2–T4 research.

WHERE CAN GENOMICS HAVE IMPACT IN THE CONTINUUM OF HEALTH AND DISEASE?

DNA-based approaches (SNP detection, copy number variation and sequencing) (see Chapters 3–5) can now be used to quantify one's predisposition, susceptibility, and risk for complex diseases. These measures can be made during health and even at birth or at any time point over a person's lifetime, as these "stable" genomic measures do not change from the time of conception. Individual SNPs or multi-SNP panels are emerging from genome-wide approaches that might be used as part of health risk assessment (see e.g.: Bare et al., 2007; Zheng et al., 2008). DNA variation may also provide information about the possibility of being relatively protected from disease development

(Cox et al., 2007), as well as one's sensitivity or resistance to certain medications (Eichelbaum et al., 2006) and ability to metabolize nutrients in our diets (Ordovas and Corella, 2004).

All of this can and probably should be done early in life, such that a strategy to maintain health can be formulated well in advance of the development of potentially detrimental lifestyle habits and exposures. The "dynamic" components of the genome (gene-, protein- and metabolite expression; see Chapter 2) that are responsive to environmental stimuli, lifestyles, diets and pathogens are rapidly improving capabilities to predict and intervene at an individual level. Transcriptional profiles, protein expression and levels of metabolites combined with dynamic imaging modalities should provide more precise ways to screen individuals who are at high risk for developing a disease for its earliest molecular manifestations while the disease is subclinical (Seo and Ginsburg, 2005). This same type of information may also provide a definitive diagnosis and a molecular classification of a disease state that foretells prognosis.

For example, today a *her2/neu*-positive breast cancer ascribes that patient to a more aggressive form of the disease and directs care to a much different course than a *her2/neu* negative cancer (see Chapter 35). Similarly the selection of drugs can be guided both by the patients underlying genetic make up and by the molecular architecture of the disease in the individual (see Chapter 15). The emerging picture is one of applied genomics *throughout* an individual's lifetime, from health through death, that enables assessment of disease predisposition, screening, diagnosis, prognosis, therapeutic selection with precision that until now has been unachievable (Figure 14.1). "Strategic health planning" (see Chapter 17 and Snyderman and Yoediono, 2006) and disease prevention should be possible and will allow a shift in the current paradigm of care from the time when disease is manifest (rightward in Figure 14.1) to proactive personalized predictive care at a more cost-effective time in the disease life cycle (leftward in Figure 14.1).

TRANSLATIONAL GENOMICS: ENABLING COMPETENCIES

The pharmaceutical industry has a clearly defined and decision-gated series of steps to move a potential therapeutic from discovery to the clinic. Translational genomics may also benefit from defining the processes, the key competencies, and capabilities that support the directional movement of genomics from discovery to human health applications (Khoury et al., 2007). Figure 14.2 is a potential (and admittedly simplified) schema for moving genomic discoveries from bench to bedside from the perspective of needed and enabling capabilities and competencies. Academic centers, industry and government research groups right benefit from developing organizing principles to move a potential genomic discovery into practice through a variety of stages similar to what the pharmaceutical industry has done for drug development candidates. Below and in the ensuing chapters, we will consider these capabilities in greater detail.

Centralized Biobanking and Biorepositories

Biobanks – centralized, institutional repositories for biological samples – are among the most important enabling resources in

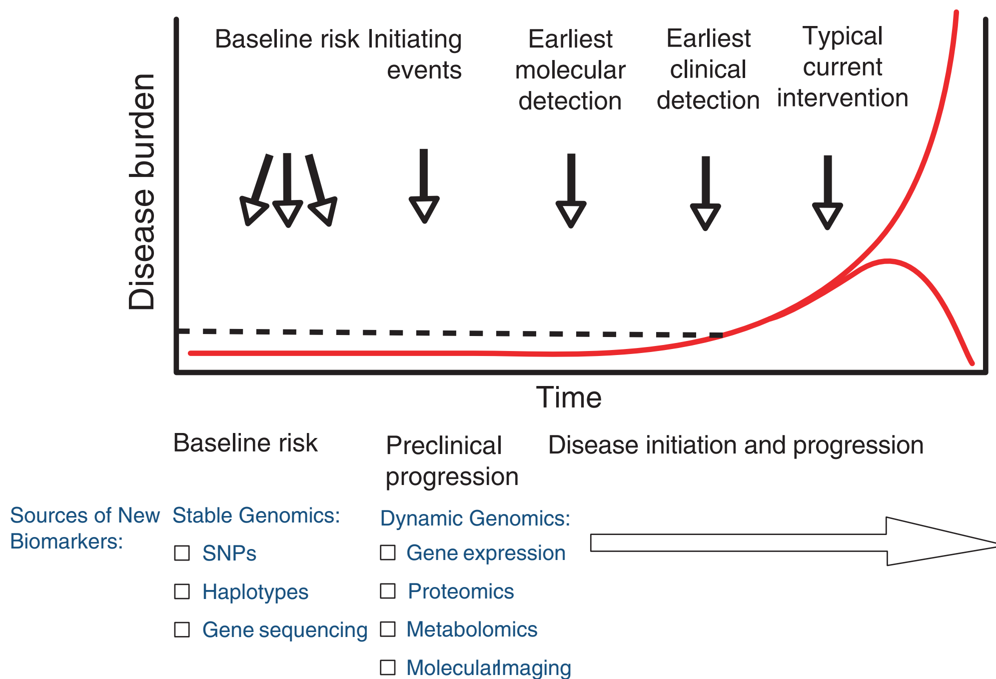


Figure 14.1 Use of genomic markers predict, prognose, diagnose, treat and monitor health and disease (adapted from Snyderman and Yoediono, 2006).

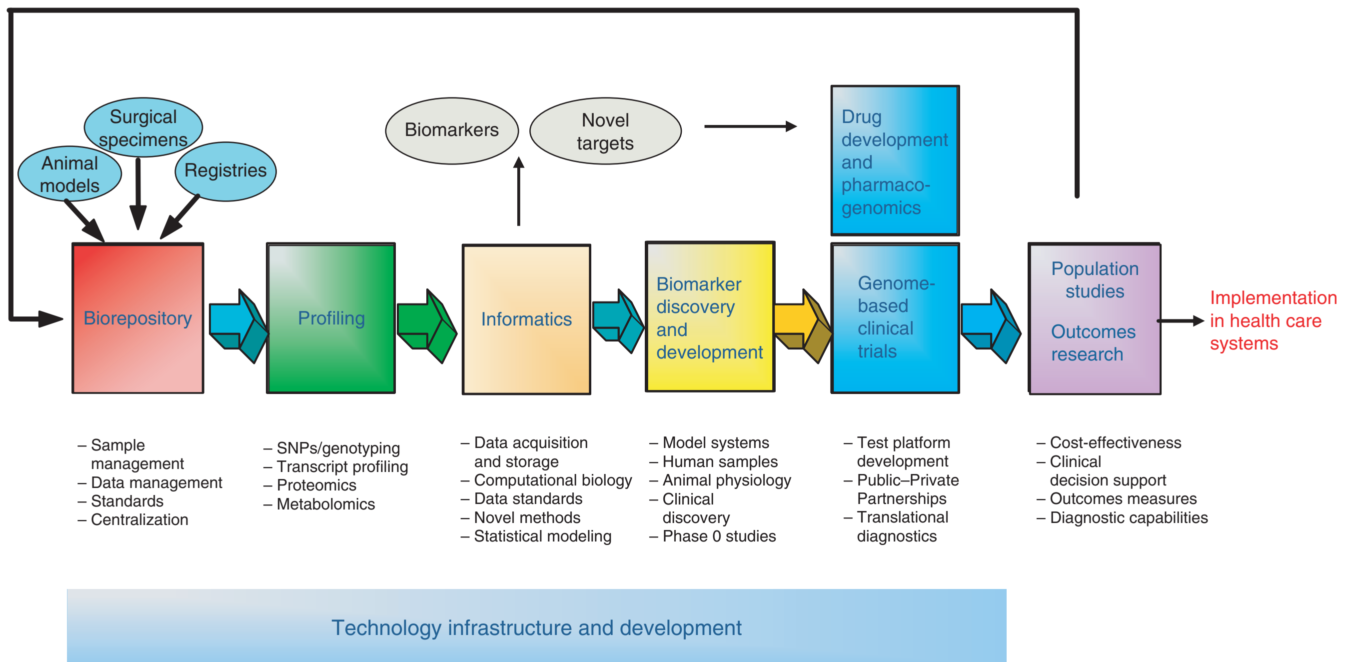


Figure 14.2 Enabling capabilities and competencies to translation genomic-based discoveries into health care applications. Based on Willard et al., 2005.

support of genomic medicine (Ginsburg et al., 2008). The study of complex diseases has been greatly facilitated by the availability of well-annotated DNA, blood and tissue specimens for high-throughput genomic analyses. Scale and scope characterize biobanks today from previous tissue repositories in Departments of Pathology or held by individual investigators. Today biobanks manage hundreds of thousands to millions of samples and require robotics and advanced informatics tools. Key capabilities for efficient and effective biobanking include centralized sample and data management systems and clearly define standards for handling of samples and their associated data. Equally important is the ability of biobanks to provide support for longitudinal studies enabling sample and data capturing in the entire disease cycle from pre-disease onset, to disease diagnosis, and clinical outcomes.

Several existing and developing biobanking models highlight developing trends in biobanking organization and management. The Singapore experience is particularly instructive in that the establishment of a national biobank was coupled with simultaneous development of the entire biomedical research infrastructure at an accelerated pace. Key considerations involve professional acceptance, custodianship, security and access. Data privacy for subjects is of paramount importance and applications of informatics technologies may provide the requisite security. Once properly implemented, biobanks may not only accelerate medical research but also contribute to public health.

Genomic "Profiling"

There are many sources of genomic data (see Chapters 3–9) that can now be applied to an integrated view of genomic epidemiology. A competency in genomic profiling technologies is clearly needed to optimize the value of clinical samples and data to develop novel predictive models of disease states and outcomes. The information provided by today's clinical and biochemical markers of disease falls well short of an adequately describing disease complexity and heterogeneity (West et al., 2006). The advent of the full complement of technologies spawned by the Human Genome Project capable of interrogating complex diseases such as cancer, cardiovascular disease, obesity and diabetes mellitus provides opportunities for acquiring quantitative data that can match the complexity of the disease. Genomic technologies can now potentially capture data that do provide this complexity, identifying discrete subsets of disease that have not been previously recognized. Profiles and patterns that identify new subclasses of tumors, such as the distinction between acute myeloid leukemia and acute myeloid leukemia (Bullinger and Valk, 2005), or Burkitt's lymphoma from diffuse B cell lymphomas (Dave et al., 2006), without prior knowledge of the classes. More recently several genomic signatures that go beyond disease classification have been discovered and validated that predict prognosis and response to therapy for many solid tumors and hematologic malignancies (Potti et al., 2006; Staudt, 2003). Not far behind, although lacking the throughput of RNA expression analyses, are whole-proteome analyses of tissue homogenates or serum using surface-enhanced laser desorption ionization–time-of-flight or liquid chromatograph/mass spectrometry/mass spectrometer technologies (Caprioli, 2005; see Chapter 8).

The capability to discern structure in this data – in the form of patterns of gene, protein, or metabolite expression that provide

snapshots of gene activity in a cell or tissue sample at a given instant of time that can then be used to describe a phenotype – is transforming biology from an observational molecular science to a data-intensive quantitative genomic science. The dimension and complexity of such data provide opportunities for uncovering patterns and trends that can distinguish subtle phenotypes in ways that traditional methods cannot (West et al., 2006). In an integrated view of genomic epidemiology, all of these sources of data will be used to develop signatures and models that classify disease and predict outcomes (Figure 14.3). To fully realize the clinical potential of genome-scale information requires a dramatic shift in the way complex, large-scale data are captures, viewed, analyzed and used.

Data Acquisition, Standards and Bioinformatics

All of the exciting medical, biological and technological advances described in this book are limited by requirements for intensive data acquisition, storage and computational capacity. Individuals with skills in this arena are in critically short supply. Informatics is a fundamental competency for the era of genomic medicine. Unless the nomenclature of the current "Tower of Babel" is improved, we will miss the opportunity for applying our newfound knowledge, because these measurements will not be compatible when attempts are made to aggregate them. Currently, most data collection efforts are relatively isolated and many different groups are working on controlled vocabularies and data standards (Hanauer et al., 2007).

In addition to nomenclature standards that allow observations to be compared across observers, the methods for exchanging data must be standardized (see Chapter 12 and Klein, 2002). Health Level Seven has essentially solved this problem for common laboratory measurements that can be measured as a number or coded output. Unfortunately, such standards are only now emerging for clinical data, and considerable variation exists in the purposes, stated or perceived, for this effort, including better individual patient care, quality measurement, research and improved administrative efficiency and billing. An effective and unifying cross-platform system will be needed that will allow collaborations among industry, academia, the NIH and other partners (see below).

There are fundamental "bioinformatics blocks" in translation of genomics that must be overcome (see Chapters 10, 12, and 13). First, there is an enormous amount of data at the level of fundamental biological measurements. Second, at the level of the patient in the intact health care system, a major challenge exists in integrating disparate electronic health records (EHRs). Third, there are further challenges in integrating research, clinical and demographic/societal data into a common format. And lastly, there must be ways to link the clinical and molecular information, so that analyses that yield both cross-sectional and longitudinal patterns relevant to disease sub-classification can be carried out in a reproducible fashion.

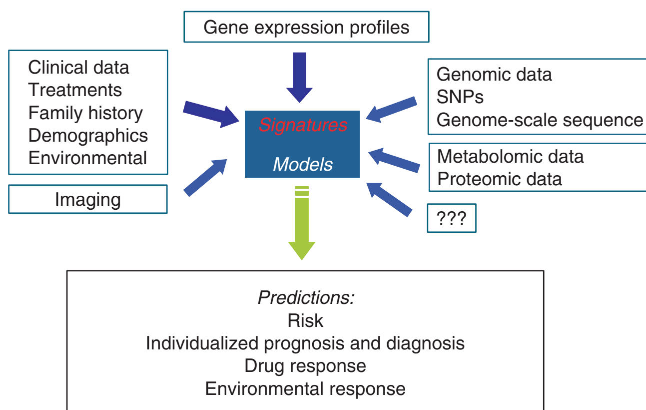


Figure 14.3 Integration of clinical, genomic, imaging, and other data to develop signatures and models to predict disease and response to drugs and environmental stimuli.

Biomarker Discovery and Development

The discovery and development of physiologic and disease biomarkers has provided a valuable set of tools to assess disease outcomes (Downing, 2000; Goodsaid et al., 2008). Biomarker research has also greatly improved our knowledge and understanding of mechanisms of etiology and disease pathogenesis. This will require access to robust model systems, human samples and exquisitely curated phenotypic data. Indeed, a genomic signature that characterizes a disease subset also provides a wealth of biological information through the members of genes, gene transcripts or proteins that constitute such a signature, either as markers of disease or as target for potential therapeutics. This approach has been particularly well developed in cancer research, although the driving concept of the gene–environment interaction has been widely applied in medical research.

Genome-Based Clinical Trials

Fundamental to defining the clinical validity and clinical utility of genomic biomarkers is the genomics-based clinical trial (see e.g., Marcom et al., 2008). In this setting, genetic and genomic biomarkers are fundamental to the trial design and the selection of patients for the trial (inclusion criteria) or for assignment to a specific experimental arm of the study. The basic principles of experimental design can also be applied to genomic studies. The task is to design a study in as economical a way as possible and simultaneously to optimize the information content of the experimental data. Clinical studies that serve to enrich a trial with a particular genomic biomarker or to test that a particular genomic subset of patients respond to therapy are required to advance translational genomics. Often these studies will both "validate" the biomarker – showing that it indeed identifies the relevant group of patients – and at the same time show its clinical utility – that the patient group selected by the biomarker indeed responds differently to the intervention in the trial. Key to enabling genome-based clinical trials is the interface between

discovery genomics with the platforms for developing a genomic assay that is ready for clinical use. In many cases, this will require key public–private partnerships to facilitate the transition from discovery research to human applications. In addition, there are several open issues in the design of genomic studies that need further investigation, such as selection bias at the patient recruitment level, tissue specimen sampling procedures, assays, and time point selection; in addition lack of calibration, poor reproducibility of molecular signatures across different studies and technical variation between different chip technologies also are in need of greater study as these novel study designs are developed.

Integration of Genomics into Drug Discovery and Development

Currently, the pharmaceutical industry is facing challenges in terms of the low number of new chemical entities (NCEs) that are approved each year, despite soaring research and development (R&D) costs. A main cause for low pharmaceutical productivity is pipeline attrition due to failure of NCEs to demonstrate efficacy or the emergence of safety issues during or after development. In parallel to genomic-guided clinical trials, pharmaceutical studies that use biomarkers for trial enrichment or for pharmacogenomics will be another pathway for genomics to reach the clinic in the form or diagnostic/therapeutic combinations. The practice of genomics and related technologies in transforming drug discovery and development has provided sophisticated and novel new approaches for target identification, target validation, the drug screening process and the understanding of the mechanism of action of targets and drugs (Stoughton and Friend, 2005). In drug development, the process by which new drugs are tested in model organisms and in humans, genomics has provided a means to understand heterogeneous disease populations at the gene level. This has enabled the selection of subjects for clinical trials with more homogeneous disease characteristics, thereby enabling a much better assessment of drug activity. In addition, the understanding of drug metabolism on a genetic level has offered a means to provide drugs with a better safety profile for certain patient populations.

Pharmacogenetics and Pharmacogenomics

The impact of the genome on our ability to predict drug response is one of the most promising and fertile areas of genomic and personalized medicine (see Chapter 15). *Pharmacogenetics* is the study of genetic variation that ultimately gives rise to the variable responses in individuals to any given drug treatment. More recently, pharmacogenetics has provided an explanation as to why certain individuals do not respond, or respond differently to a given drug treatment. *Pharmacogenomics* utilizes genomic technology to understand the effects of all relevant genes on the behavior of a drug or conversely the effect of a drug on gene expression. Pharmacogenomics, like pharmacogenetics, has rapidly embraced genomic technologies to identify molecular patterns of response, drug disposition and drug targets, yielding molecular biomarkers of drug response both of which have great potential to positively affect the area of medicine. Grossman and Goldstein (Chapter 15) outline the scope of study designs for pharmacogenomics studies, depending on specific features related to the study population, underlying disease, phenotypes of interest (i.e., drug efficacy, safety and dosing) and drug compound being tested. Considerations related to the realization of a valid genetic effect into an acceptable diagnostic kit are reviewed, providing analytical tools objectively evaluating the validity, utility and cost-effectiveness, as well as ethical properties, associated with a specific drug application.

Translational Diagnostics

Translational diagnostics is “the subfield of translational medicine concerned with diagnostic methods and information.” Many of the translational issues affecting genomics as whole also affect this area of innovation. These include communication and education of researchers, institutional support for translational research activities, the regulatory environment, cost and reimbursement issues, the need for bioinformatics and complex data handling capabilities, intellectual property and licensing protection and the willingness to embrace new paradigms for diagnostic and therapeutic strategies (Horig and Pullman, 2004; Horig et al., 2005; Humes, 2005; Marincola, 2003; Sonntag, 2005).

Integration of Genomic Testing into Clinical Practice

A final hurdle that genome-based medicine faces is the integration of testing into day-to-day clinical practice. Despite the optimism that genomic testing might bring to improving the practice of medicine, there are barriers that must be overcome for their seamless integration into clinical practice. The incorporation of genetics and genomics into patient management guidelines has largely failed to occur thus far, perhaps because researchers, diagnostic firms and the regulatory authorities are still seeking to establish methodologies by which to judge their effectiveness because practicing clinicians and guideline writers are still working to understand how such new tests fit into current models of care and risk assessment, and because payers are just beginning to foresee new pressures to cover the additional costs. Indeed, population studies and outcome studies will be critical for this final phase of clinical uptake, implementation into guidelines and into reimbursement strategies.

To address some of these issues, a framework has been proposed to assist in genetic testing evaluation, consisting of six areas (Table 14.2): technical and operational excellence, diagnostic capability, impact on diagnostics and/or prognostics, impact on therapeutic strategy, cost-effectiveness and health outcomes (Douglas and Ginsburg, 2008). It is clear that the physician community faces many of the same issues the payer and diagnostics groups do as well. Therefore, it makes sense to consider developing the partnerships that bring the stakeholders together to jointly address these translational obstacles.

TABLE 14.2 Framework for evaluating genetic tests*

Technical capability and operational excellence	● development of analytic standards ● computational capabilities to acquire, store and analyze large datasets
Diagnostic capability	● define the clinical situation(s) that may be informed by genomic data ● validation of genomic associations or predictors in distinct and separately ascertained cohorts
Impact on diagnostics and/or prognostics	● development of clinical decision support platforms and "just-in–time" solutions in the practice environment ● linkage to authoritative knowledge sources
Impact on therapeutic strategy	● evaluate genomic testing in the context of physicians' confidence in decision-making ● develop educational strategies with test implementation ● assess physician behavior
Cost-effectiveness	● assessment of system wide costs resultant health care utilization and outcomes ● assess patient, provider, payer and government perspectives
Health outcomes	● incorporation into guidelines driven by the evidence base ● carry out prospective randomized trials of usual care versus genomics guided care ● development of an educational strategy aimed at the patient

*Modified from Douglas and Ginsburg (2008).

DEVELOPING ENVIRONMENTS THAT FOSTER TRANSLATIONAL GENOMICS TO HEALTH APPLICATIONS

It is clear that academicians and industry scientists are on convergent paths to personalized medicine using genomic technologies and information. Specialized centers for genomic and personalized medicine in academic health systems can be instrumental in integrating, facilitating and catalyzing the needs of both the academic and industry stakeholders by providing:

- Access to patients, patient data, and molecular and biological data that drive the development and exploration of genomic information and its link to clinical outcomes.
- The scientific foundation for novel biomarker discovery for both disease and drug response based on mechanism.
- An environment for innovation that fuels the development of novel translational strategies.
- A vehicle for aligning the efficiency and quality metrics of patient care with the goals of personalized medicine.
- The infrastructure for the types of public–private partnerships required for executing genomic assay-guided clinical trials, and, finally,
- A place to engage in a dialog and research on the key issues challenging the translation of genomics into personalized health care (Haga and Willard, 2006): education, facilitating clinical trials, information systems and integration into practice.

To bring about the transformation in health care promised by the genome sequence and its derivatives will require assembling diverse stakeholders focused on the application and translation of genomics with a goal of improving the health of individuals and driving efficiency in health care. Specialized centers housing basic genome science laboratories, clinical researchers, informaticians, clinicians, health policy makers and in partnership with industry (pharmaceutical and diagnostic companies), and with health systems that will enable the scientific output of the genome to cross the chasm between bench and bedside.

Just a few years ago, academic medical centers alone were the primary mechanism to foster translational research and facilitate collaboration between researchers and clinicians. In the United States, the key ingredients for this effort are in place through the NIH Roadmap and the Clinical and Translational Science Awards (NIH, 2007). This NIH initiative will eventually fund up to 60 academic health centers and their associated clinical practice partners, with significant infrastructure capability in most of the areas above. These large centers are being bolstered by multiple grants and contracts that are intended to provide avenues for innovative approaches to scientific problems ranging from biomedical informatics to community-based research networks. Similar programs are under way in Canada, the United Kingdom, Singapore, Germany and other countries.

In particular, it will be important for governmental and foundational funding agencies to anticipate the critical imperative for interdisciplinary work (Figure 14.4). Academic health systems will need to consciously steer faculty activities in this direction to reap the benefits of the enormous societal investment in the biomedical enterprise that has made this clinical scientific revolution possible. Industry will need to engage earlier in the discovery process and assist in guiding the research agenda.

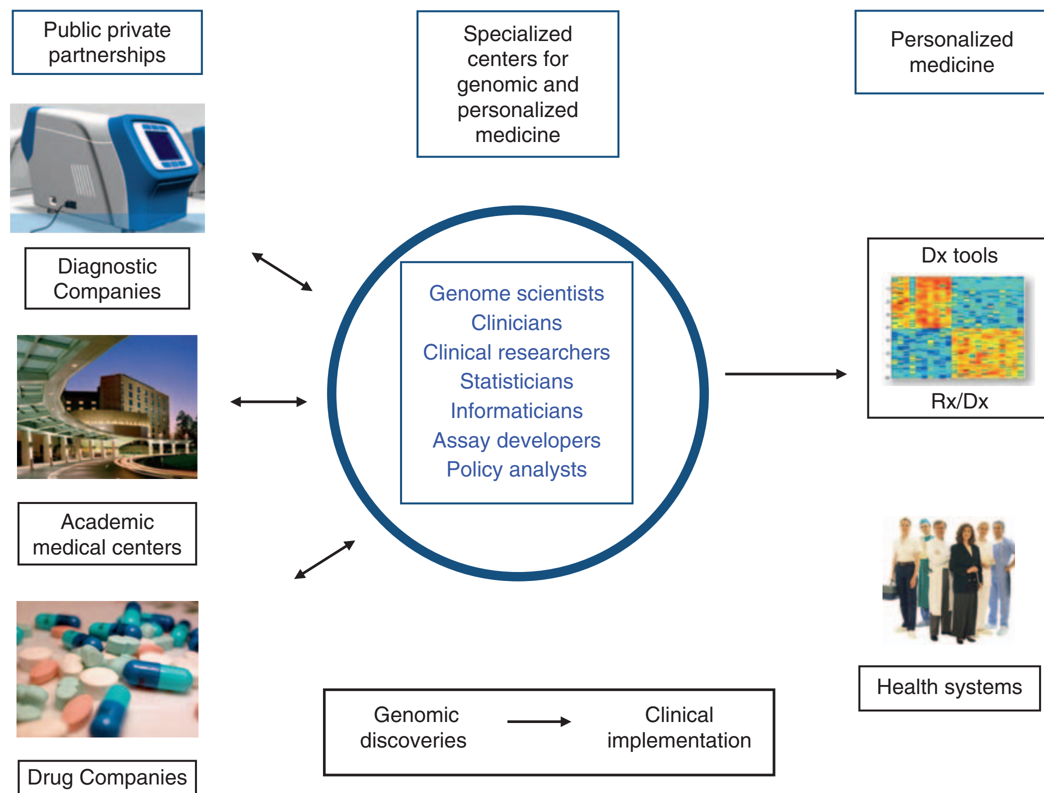


Figure 14.4 A framework for an organization to translate genomics to personalized medicine through public–private partnership and academic health centers.

Regulatory agencies and payers will need to prepare for this fundamental shift in the basis for assessing diagnostic and therapeutic technologies and their reimbursement. Effective planning for this coming shift in the way human disease and its prevention and treatment are viewed through the genomic lens, particularly the systems aspects of creating interdisciplinary hubs for the development of translational genomics, will result not only in an advanced clinical scientific discipline, but improved health for people worldwide.

2009 UPDATE

There continue to be significant advances in translational genomics. Progress has been made in developing the infrastructure that supports developing and disseminating genome information, the informatics tools for clinical use of this information, as well as the definition and development of the evidence required for clinical adoption. Indeed the labels of more therapies are including genetic and genomic information to facilitate prescribing by physicians – a major advance for pharmacogenetics. Personal genomics represents a new and potentially disruptive innovation in translating genomics, and the stakeholder community is developing guidance as to how best to use this information in the care of patients.

Tools and Catalogs of Genome Information

The toolbox for achieving the goals of genomic and personalized medicine is becoming increasingly high throughput and low cost. In April 2008, the sequencing of the genome of Nobel scientist James Watson was completed (http://jimwatsonsequence.cshl.edu/cgi-perl/gbrowse/jwsequence/). It took 13 years and $3B to complete the first draft sequence of the human genome; however, Watson's sequence of the human genome cost only $1M and took two months to complete. Advances in genomic sequencing are occurring exponentially and the $1000 genome is likely within reach. The NHGRI (US NIH) announced several initiatives aimed at taking full advantage of advanced sequencing technologies – Medical Resequencing (http://www.genome.gov/15014882), 1000 Genomes (http://www.genome.gov/26524516), and the Human Microbiome project (http://nihroadmap.nih.gov/hmp/members.asp) are among them – that will generate large publically accessible databases of human variation in health and disease. In parallel to the acceleration of sequencing technologies, a

number of important public–private partnerships have begun to generate publically available data and frameworks to support large genome-wide association studies.

- The *Genetic Association Information Network* (*GAIN*) (http://www.genome.gov/19518664) was established as a public–private partnership to support genome-wide association studies. The effort has linked comprehensive genotyping data on roughly 2000 participants to their clinical records and deposited the information in the publicly accessible dbGaP (http://www.ncbi.nlm.nih.gov/entrez/query/Gap/gap_tmpl/about.html).
- The *Biomarker Consortium* (http://www.biomarkers-consortium.org/) was set up as an industry–government partnership to discover and develop biomarkers for risk for developing disease, to assist in diagnosis or to guide treatment.
- The NIH launched the *Genes, Environment, and Health Initiative* (http://www.gei.nih.gov/) to develop data on how genetic variation affects an individual's response to environmental factors.
- *The Cancer Genome Atlas* (http://cancergenome.nih.gov/) to explore the entire spectrum of genomic changes involved in human cancer, starting with brain, lung and ovarian cancer. The first results were reported this year for glioblastoma (Cancer Genome Atlas Research Network, 2008) and the first data for ovarian cancer were recently released (http://tcga-data.nci.nih.gov/tcga/homepage.htm).

Developing a Nationwide Biospecimen Repository

There remains a compelling need for centralized biobanks and development of standards of biospecimen banking that are universally accepted and applied (Ginsburg et al., 2008). No organized effort is underway; however, in the United States, the Clinical and Translational Science Awards consortium has initiated several projects to achieve this goal, including one aimed at defining the data standards and IT platform to create a "virtual" national biorepository, as well as a second to articulate the cultural barriers and regulatory best practices. These data will lay the groundwork for a future initiative of national biospecimen banking.

Advances in Health Care Information Technology

Health care information technology (HIT) development is essential to record genetic profiles of disease susceptibility and drug response, to gather evidence for improved outcomes, and to aid in the discovery of new gene-disease associations through formal studies and the collection of data in everyday clinical encounters. HIT systems will also be important in providing clinicians not trained in genomics with critical information and decision-making tools (Kawamoto et al., 2009). Genomic and personalized medicine will require a sound HIT infrastructure to support clinical decisions to deliver full benefit to the patient and to the health system as a whole. This effort remains woefully behind the need. Sobering is a study that reported that only 19% of hospitals and 13% of solo physicians have implemented fully operational EHRs. Large group practices (<50 physicians) are relatively ahead of the curve, with 57% having adopted EHRs, but adoption remains incomplete overall (Jha et al., 2006). Surprisingly, large segments of the health care profession remain primarily on a paper-based system. The reasons for this rest largely on two key issues: the cost of implementing a health care information system and the lack of a uniformly accepted standard in the way data is stored and shared. Interoperable HIT is a goal; how this will be accomplished remains unclear. In the United States, the Obama administration has prioritized nationwide implementation of standardized HIT, and there is reason to hope that a changeover to EHRs will occur within a matter of years with government driving the process.

Where are We Today in Translating Genomics to Health care?

Despite the clear advances in technology that bring genomic information closer to physicians, patients, and the public, looming even closer are issues that are outside of the sphere of the genome sciences and more in the realm of genome policy. Scheuner and colleagues carried out an extensive meta-analysis of 68 studies seeking to integrate genomic medicine to the clinical management of chronic diseases (Scheuner et al., 2008). The study posed the following questions: "What are the outcomes of genomic medicine? What is the current level of consumer understanding about genomic medicine and what information do consumers need before they seek services? How is genomic medicine best delivered? What are the challenges and barriers to integrating genomic medicine into clinical practice?" The authors concluded that gaps in provider and public education, addressing privacy concerns, building an evidence base, and developing the appropriate cost models for genomic medicine in health care delivery will need to be addressed in order to facilitate the translation of genomic innovations into clinical practice. These findings are not surprising given the spectrum genome policy issues that have emerged as the genome sciences have matured (Haga and Willard, 2006); nonetheless, this is the first comprehensive analysis of data on these issues and provides a basis to address them.

Evidence Evaluation

The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative (http://www.cdc.gov/genomics/gtesting/egapp.htm), established by the US National Office of Public Health Genomics at the Centers for Disease Control and Prevention, supports the development and implementation of a rigorous, evidence-based process for evaluating genetic tests and other genomic applications for clinical and public health practice in the United States. EGAPP recently published its recommendations for the use of CYP450 polymorphisms

in the management of selective serotonin reuptake inhibitors (EGAPP, 2007), gene expression profiles and outcomes in breast cancer (EGAPP 2009a), and UGTA1A testing in colorectal cancer patients treated with irinotecan (EGAPP 2009b).

Building the Evidence: Genome-Based Clinical Trials

Personalized approaches to cancer therapy will likely depend on genomic assays. Full-transcriptome assays using microarrays have the advantage of providing multiple prognostic and predictive signatures in one assay. Duke University's Institute for Genome Sciences & Policy has created a unique clinical infrastructure with the objective of obtaining full genome expression data on cancer samples as a clinical assay for use in a prospective trials (www.genomestohealth.org). One of these trials – "Performance of Genomic Expression Profiles to Direct the Use of Preoperative Chemotherapy for Early Stage Breast Cancer" – is a prospective trial validating genomic signatures for predicting response to doxorubicin (A) or docetaxel (T) treatment in HER2 negative cancers (Marcom et al, 2008). Fresh-frozen cores are reviewed by the study pathologist for tumor content. RNA is then extracted and probe generated to hybridize to an Affymetrix U133Plus2.0 microarray. Microarray data quality is determined using summary metrics for U133Plus2.0 arrays and principal component analysis (PCA) plots. The data is then used for predicting A or T sensitivity. This is an important example of a series of innovative clinical trials underway to develop the prospective evidence in breast, lung, prostate, ovarian, colon, and hematologic malignancies that a genomic signature derived from a patient's tumor can predict prognosis, as well as response and resistance to conventional cytotoxic chemotherapy.

Pharmacogenetics

Information on genetic risk for adverse events is now beginning to appear on drug labels. For example, testing for the HLA-B*5701 allele is recommended for the initiation of abacavir therapy to reduce the risk of hypersensitivity reactions (Mallal et al., 2008). In a randomized control trial, testing for this allele showed a 50% reduction in the risk for hypersensitivity. Testing is also recommended for HLA-B*1502 for Asians who are being prescribed carbamazepine to identify those at high risk for Stevens Johnson syndrome (FDA, 2009). A recent study of warfarin pharmacogenetics found that a clinical/genetics algorithm better predicted maintenance doses compared to a clinical algorithm without genetics, particularly for the outliers (International Warfarin Pharmacogenetics Consortium, 2009). In 2008, the FDA published a comprehensive list of biomarkers on drug labels (http://www.fda.gov/cder/genomics/genomic_biomarkers_table.htm), laying the groundwork for the clinical implementation of pharmacogenetic testing in clinical practice.

Personal Genomics

"The genome is out of the bottle" (Hunter et al., 2008). Direct-to-consumer marketing of personal genomics services has raised concerns about the scientific foundation for such tests. There is a growing need for clarity of the evidence base required for both the clinical validity and utility of personal genomics. A recent NIH/CDC workshop made the following recommendations: (1) develop industry-wide scientific standards, (2) develop a multidisciplinary research agenda, including observational studies and clinical trials, (3) enhance credible knowledge synthesis and information dissemination to both providers and consumers of health cares, (4) link scientific evidence to recommendations for use of personal genomics, and (5) expand the concept of clinical utility to include personal health utility and developing appropriate metrics for evaluation (http://www.scgcorp.com/PersonalGeno2008/agenda.htm).

2009 UPDATE REFERENCES

Cancer Genome Atlas Research Network. (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455(7216), 1061–1068.

Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group. (2007). Recommendations from the EGAPP Working Group: Testing for cytochrome P450 polymorphisms in adults with nonpsychotic depression treated with selective serotonin reuptake inhibitors. *Genet Med* 9(12), 819–825.

Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group. (2009a). Recommendations from the EGAPP Working Group: Can tumor gene expression profiling improve outcomes in patients with breast cancer? *Genet Med* 11(1), 66–73.

Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group. (2009b). Recommendations from the EGAPP Working Group: Can UGT1A1 genotyping reduce morbidity and mortality in patients with metastatic colorectal cancer treated with irinotecan? *Genet Med* 11(1), 15–20.

Food and Drug Administration. Information on carbamazepine (marketed as Carbatrol, Equetro, Tegretol, and generics). http://www.fda.gov/cder/drug/infopage/carbamazepine/default.htm (accessed January 30, 2009).

Ginsburg, G.S., Burke, T. and Febbo P. (2008). Centralized biorepositories for academic medical centers and health systems. *JAMA* 19(299), 1359–1350.

Haga, S. and Willard, H. (2006). Defining the spectrum of genome policy. *Nature* 7, 966–972.

Hunter, D.J., Khoury, M.J. and Drazen, J.M. (2008). Letting the genome out of the bottle – Will we get our wish? *N Engl J Med* 358(2), 105–107.

International Warfarin Pharmacogenetics Consortium, Klein, T.E., Altman, R.B., Eriksson, N., Gage, B.F., Kimmel, S.E., Lee, M.T.,

Limdi, N.A., Page, D., Roden, D.M., Wagner, M.J. et al. (2009). Estimation of the warfarin dose with clinical and pharmacogenetic data. *N Engl J Med* 360(8), 753–764.

Jha, A.K., Ferris, T.G., Donelan, K., DesRoches, C., Shields, A., Rosenbaum, S. and Blumenthal, D. (2006). How common are electronic health records in the United States? A summary of the evidence. *Health Affairs* 25, w496–w507.

Kawamoto, K., Lobach, D.F., Willard, H.F., Ginsburg, G.S. (2009). National clinical decision support infrastructure to enable the widespread and consistent application of genetics and genomics in healthcare. *BMC Med Inform Decis Mark.* 9, 17–31.

Mallal, S., Phillips, E., Carosi, G., Molina, J.M., Workman, C., Tomazic, J., Jägel-Guedes, E., Rugina, S., Kozyrev, O., Cid, J.F. et al. (2008). HLA-B*5701 screening for hypersensitivity to abacavir. *N Engl J Med* 358, 568–579.

Marcom, P.K., Anders, C.K., Ginsburg, G.S., Potti, A. and Nevins, J. (2008). Utilization of genomic signatures for personalized treatment of breast cancer. In *Pharmacogenetics of Breast Cancer*, Leyland-Jones, B., ed., MacMillian Publishing Solutions.

Scheuner, M., Sieverding, P. and Shekelle, P. (2008). Delivery of genomic medicine for common chronic adult diseases: A systematic review. *JAMA* 299(11), 1320–1334.

REFERENCES

Bare, L., Morrison, A., Rowland, C., Shiffman, D., Luke, M., Iakoubova, O., Kane, J., Malloy, M., Ellis, S., Pankow, J. et al. (2007). Five common gene variants identify elevated genetic risk for coronary heart disease. *Genet Med*(10), 682–689.

Bethke, L., Sullivan, K., Webb, E., Murray, A., Schoemaker, M., Auvinen, A., Kiuru, A., Salminen, T., Johansen, C., Christensen, H.C. et al. (2008). The common D302H variant of CASP8 is associated with risk of glioma. *Cancer Epidemiol Biomarkers Prev* 17(4), 987–989.

Bullinger, L. and Valk, P. (2005). Gene expression profiling in acute myeloid leukemia. *J Clin Oncol* 23(26), 6296–6305.

Caprioli, R. (2005). Deciphering protein molecular signatures in cancer tissues to aid in diagnosis, prognosis, and therapy. *Cancer Res* 65(23), 10642–10645.

Cox, A., Dunning, A., Garcia-Closas, M., Balasubramanian, S., Reed, M., Pooley, K., Scollen, S., Baynes, C., Ponder, B., Chanock, S. et al. (2007). A common coding variant in CASP8 is associated with breast cancer risk. *Nat Genet* 39(3), 352–358.

Dave, S., Fu, K., Wright, G., Lam, L., Kluin, P., Boerma, E., Greiner, T., Weisenburger, D., Rosenwald, A., Ott, G. et al. (2006). Lymphoma/leukemia molecular profiling project. Molecular diagnosis of Burkitt's lymphoma. *N Engl J Med* 354(23), 2431–2442.

Douglas, P., Ginsburg, G. (2008). Clinical genomic testing: Getting it right. *J Cardiovasc Translational Res* 1(1), 17–20.

Downing, G.J. (2000). *Biomarkers and Surrogate Endpoints: Clinical Research and Applications*. Elsevier, New York.

Easton, D.F., Pooley, K.A., Dunning, A.M., Pharoah, P.D., Thompson, D., Ballinger, D.G., Struewing, J.P., Morrison, J., Field, H., Luben, R. et al. (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 447(7148), 1087–1093.

Eichelbaum, M., Ingelman-Sundberg, M. and Evans, W. (2006). Pharmacogenomics and individualized drug therapy. *Annu Rev Med* 57, 119–137.

Feero, W., Guttmacher, A. and Collins, F. (2008). The genome gets personal–almost. *J Am Med Assoc* 299(11), 1351–1352.

Fellay, J., Shianna, K.V., Ge, D., Colombo, S., Ledergerber, B., Weale, M., Zhang, K., Gumbs, C., Castagna, A., Cossarizza, A. et al. (2007). A whole-genome association study of major determinants for host control of HIV-1. *Science* 317(5840), 944–947.

Frayling, T.M., Timpson, N.J., Weedon, M.N., Zeggini, E., Freathy, R.M., Lindgren, C.M., Perry, J.R., Elliott, K.S., Lango, H., Rayner, N.W. et al. (2007). A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 316(5826), 889–894.

Ginsburg, G., Burke, T. and Febbo, P. (2008). Centralized biorepositories for genetic and genomic research. *J Am Med Assoc* 299(11), 1359–1361.

Ginsburg, G.S. and McCarthy, J.J. (2001). Personalized medicine: Revolutionizing drug discovery and patient care. *Trends Biotechnol* 19(12), 491–496.

Ginsburg, G.S., Konstance, R.P., Allsbrook, J.S. and Schulman, K.A. (2005). Implications of pharmacogenomics for drug development and clinical practice. *Arch Intern Med* 165(20), 2331–2336.

Goodsaid, F., Frueh, F. and Mattes, W. (2008). Strategic paths for biomarker qualification. *Toxicology* 245(3), 219–223.

Graham, I.D., Logan, J., Harrison, M.B., Straus, S.E., Tetroe, J., Caswell, W. and Robinson, N. (2006). Lost in knowledge translation: Time for a map?. *Journal of Continuing Education in the Health Professions* 26, 13–24.

Gregory, S.G., Schmidt, S., Seth, P., Oksenberg, J.R., Hart, J., Prokop, A., Caillier, S.J., Ban, M., Goris, A., Barcellos, L.F. et al. (2007). Interleukin 7 receptor alpha chain (IL7R) shows allelic and functional association with multiple sclerosis. *Nat Genet* 39(9), 1083–1091.

Gudbjartsson, D.F., Arnar, D.O., Helgadottir, A., Gretarsdottir, S., Holm, H., Sigurdsson, A., Jonasdottir, A., Baker, A., Thorleifsson, G., Kristjansson, K. et al. (2007). Variants conferring risk of atrial fibrillation on chromosome 4q25. *Nature* 448(7151), 353–357.

Haga, S. and Willard, H. (2006). Defining the spectrum of genome policy. *Nature Rev Genetics* 7, 966–972.

Hanauer, D., Rhodes, D., Sinha-Kumar, C. and Chinnaiyan, A. (2007). Bioinformatics approaches in the study of cancer. *Curr Mol Med* 7(1), 133–141.

Horig, H. and Pullman, W. (2004). From bench to clinic and back: Perspective on the 1st IQPC translational research conference. *Journal of Translational Medicine* 2, 44–51.

Horig, H., Marincola, E. and Marincola, F. (2005). Obstacles and opportunities in translational research. *Nat Med* 11(7), 705–708.

Humes, H. (2005). Translational medicine and the National Institutes of Health road map: Steep grades and tortuous curves. *J Lab Clin Med* 146, 51–54.

Hunter, D., Khoury, M. and Drazen, J. (2008). Letting the genome out of the bottle – will we get our wish? *N Engl J Med* 358(2), 105–107.

Janssens, A., Gwinn, M., Bradley, L., Oostra, B., van Duijn, C. and Khoury, M. (2008). A critical appraisal of the scientific basis of commercial genomic profiles used to assess health risks and personalize health interventions. *Am J Hum Genet* 82(3), 593–599.

Khoury, M., Gwinn, M., Yoon, P., Dowling, N., Moore, C. and Bradley, L. (2007). The continuum of translation research in genomic medicine: How can we accelerate the appropriate integration of human

genome discoveries into health care and disease prevention? *Genet Med*(10), 665–674.

Klein, G. (2002). Standardization of health informatics – results and challenges. *Meth Inform Med* 41(4), 261–270.

Lowe, C.E., Cooper, J.D., Brusko, T., Walker, N.M., Smyth, D.J., Bailey, R., Bourget, K., Plagnol, V., Field, S., Atkinson, M. et al. (2007). Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the IL2RA region in type 1 diabetes. *Nat Genet* 39(9), 1074–1082.

Lynch, T., Bell, D., Sordella, R., Gurubhagavatula, S., Okimoto, R., Brannigan, B., Harris, P., Haserlat, S., Supko, J., Haluska, F. et al. (2004). Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 350(21), 2129–2139.

Ma, X., Zabriskie, N.A., Hau, V.S., Chen, H., Tong, Z., Yang, X., Gibbs, D., Katz, B.J., Luo, L., Pearson, E. et al. (2007). Genetic association of LOXL1 gene variants and exfoliation glaucoma in a Utah cohort. *Cell Cycle* 7(4).

Marcom, P.K., Anders, C.K., Ginsburg, G.S., Nevins, J.R. and Potti, A. (2008). Utilization of genomic signatures for personalized treatment of breast cancer. In: "Pharmacogenetics of Breast Cancer", B. Leyland-Jones, ed., Macmillan Publishing Solutions (in press).

Marincola, F. (2003). Translational medicine: A two-way road. *Journal of Translational Medicine* 1, 1–2.

McGuire, A., Cho, M., McGuire, S. and Caulfield, T. (2007). Free in PMC medicine. The future of personal genomics. *Science* 317(5845), 1687.

Moffatt, M.F., Kabesch, M., Liang, L., Dixon, A.L., Strachan, D., Heath, S., Depner, M., von Berg, A., Bufe, A., Rietschel, E. et al. (2007). Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 448(7152), 470–473.

NIH (2007). http://www.ncrr.nih.gov/clinical_research_resources/clinical_and_translational_science_awards/

Ordovas, J. and Corella, D. (2004). Nutritional genomics. *Annu Rev Genom Hum Genet* 5, 71–118.

Parkes, M., Barrett, J.C., Prescott, N.J., Tremelling, M., Anderson, C.A., Fisher, S.A., Roberts, R.G., Nimmo, E.R., Cummings, F.R., Soars, D. et al. (2007). Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat Genet* 39(7), 830–832.

Potti, A., Mukherjee, S., Petersen, R., Dress, A.H., Bild, A., Koontz, J., Kratzke, R., Watson, M., Kelley, M., Ginsburg, G. et al. (2006). A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. *N Engl J Med* 355(6), 570–580.

Remmers, E.F., Plenge, R.M., Lee, A.T., Graham, R.R., Hom, G., Behrens, T.W., de Bakker, P.I., Le, J.M., Lee, H.S., Batliwalla, F. et al. (2007). STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N Engl J Med* 357(10), 977–986.

Samani, N.J., Erdmann, J., Hall, A.S., Hengstenberg, C., Mangino, M., Mayer, B., Dixon, R.J., Meitinger, T., Braund, P., Wichmann, H.E. et al. (2007). Genomewide association analysis of coronary artery disease. *N Engl J Med* 357(5), 443–453.

Sandhu, M.S., Weedon, M.N., Fawcett, K.A., Wasson, J., Debenham, S.L., Daly, A., Lango, H., Frayling, T.M., Neumann, R.J., Sherva, R. et al. (2007). Common variants in WFS1 confer risk of type 2 diabetes. *Nat Genet* 39(8), 951–953.

Secretary's Advisory Committee on Genetics, Health, and Society (2007). Draft Report On US System of Oversight of Genetic Testing.

Seo, D. and Ginsburg, G. (2005). Genomic medicine: Bringing biomarkers to clinical medicine. *Curr Opin Chem Biol*(4), 381–386.

Sklar, P., Smoller, J.W., Fan, J., Ferreira, M.A., Perlis, R.H., Chambert, K., Nimgaonkar, V.L., McQueen, M.B., Faraone, S.V., Kirby, A. et al. (2008). Whole-genome association study of bipolar disorder. *Mol Psychiatr* 6, March 4 [Epub ahead of print]

Snyderman, R. and Williams, R.S. (2003). Prospective medicine: The next health care transformation. *Acad Med* 78(11), 1079–1084.

Snyderman, R. and Yoediono, Z. (2006). Prospective care: A personalized, preventative approach to medicine. *Pharmacogenomics* 7(1), 5–9.

Sonntag, K. (2005). Implementations of translational medicine. *Journal of Translational Medicine* 3, 33–35.

Staudt, L. (2003). Molecular diagnosis of the hematologic cancers. *N Engl J Med* 348(18), 1777–1785.

Stoughton, R.B. and Friend, S.H. (2005). How molecular profiling could revolutionize drug discovery. *Nat Rev Drug Discov* 4(4), 345–350.

Sung, N., Crowley, W., Genel, M., Salber, P., Sandy, L., Sherwood, L., Johnson, S., Catanese, V., Tilson, H., Getz, K. et al. (2003). Central challenges facing the national clinical research enterprise. *J Am Med Assoc* 289(10), 1278–1287.

Tenesa, A., Farrington, S.M., Prendergast, J.G., Porteous, M.E., Walker, M., Haq, N., Barnetson, R.A., Theodoratou, E., Cetnarskyj, R., Cartwright, N. et al. (2008). Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet* 40(5), 631–637.

van Heel, D.A., Franke, L., Hunt, K.A., Gwilliam, R., Zhernakova, A., Inouye, M., Wapenaar, M.C., Barnardo, M.C., Bethel, G., Holmes, G.K. et al. (2007). A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet* 39(7), 827–829.

West, M., Ginsburg, G., Huang, A. and Nevins, J. (2006). Embracing the complexity of genomic data for personalized medicine. *Genome Res*(5), 559–566.

Westfall, J., Mold, J. and Fagnan, L. (2007). Practice-based research – "Blue Highways" on the NIH roadmap. *J Am Med Assoc* 297(4), 403–406.

Willard, H., Angrist, M. and Ginsburg, G. (2005). Genomic medicine: Genetic variation and its impact on the future of health care. *Phil Trans Roy Soc Lond B Biol Sci* 360(1460), 1543–1550.

Wohlsen, M. (2008). http://news.yahoo.com/s/ap/20080331/ap_on_hi_te/trolling_for_genes_startups

Yates, J.R., Sepp, T., Matharu, B.K., Khan, J.C., Thurlby, D.A., Shahid, H., Clayton, D.G., Hayward, C., Morgan, J., Wright, A.F. et al. (2007). Complement C3 variant and the risk of age-related macular degeneration. *N Engl J Med* 357(6), 553–561.

Zheng, S., Sun, J., Wiklund, F., Smith, S., Stattin, P., Li, G., Adami, H., Hsu, F., Zhu, Y., Bälter, K. et al. (2008). Cumulative association of five genetic variants with prostate cancer. *New Engl J Med* 358(9), 910–919.

CHAPTER

Pharmacogenetics and Pharmacogenomics

Iris Grossman and David B. Goldstein

INTRODUCTION

Western medicine has been extremely successful in producing mass pharmaceutical solutions by following the "one-drug-fits-all" paradigm. Indeed, the majority of individuals plotted on a Gaussian distribution show reasonable efficacy and safety profiles, justifying the overall validity of this approach. However, alarming reports are amassing on the frequency and gravity of adverse drug responses, as well as inadequate efficacy, clearly signaling the need to shift from "blockbuster" to more individualized treatment models. The objective of pharmacogenetic research is to identify a genetic marker, or a set of genetic markers, that serve as prognostics for the fashion in which a given person is likely to respond to a given medicine. The appropriate formulation, dose, and regimen are presumed to be predicted, at least partially, by genetic determinants, thus serving as diagnostics that can be tested prior to initiation of drug treatment. In this fashion, treatment would be allocated only, or mainly, to individuals who are expected to benefit from it, thus reducing dramatically the suffering and costs associated with adverse drug reactions (ADRs) and inadequate efficacy. As summarized by Wilkinson, "most major drugs are effective in only 25–60% of patients, and more than two million cases of ADRs occur annually in the US, including 100,000 deaths" (Wilkinson, 2005).

It should be noted that two terms exist in the scientific literature, *pharmacogenetics* and *pharmacogenomics*, each with its own connotation and a range of alternative and poorly defined distinctions. Here, we use the term *pharmacogenetics* (PGx), which has historical priority, according to its broadest meaning – relating heritable variation to inter-individual variation in drug response (Goldstein et al., 2003). It should be mentioned that the term *pharmacogenomics* is often used to describe the study of differential gene-expression profiles in tissues of interest and their relation to drug-response phenotypes. While such expression signatures may be affected by underlying genetic variation, non-inherited environmental determinants play an important role as well. We therefore focus here solely on inherited sequence variation (i.e., single nucleotide polymorphisms, SNPs), insertions–deletions, copy number variations, micro-satellites, *Alu* repeats, and so on) and its association with treatment response end-points. A comprehensive review of gene-expression signatures, their association with response to chemotherapy, and the search for underlying genetic determinants is provided in Nevins and Potti (2007).

The complexity of genetic relationships and its interplay with drug-response phenotypes has been recently acknowledged by the US Food and Drug Administration (FDA) in a rather crude and controversial decision to stratify patients' populations by race in the case of severe heart failure adjunct treatment with BiDil (Meadows, 2005). Although little beneficial effect has been registered in the general population, BiDil treatment showed a dramatic reduction in death and hospitalization in African

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

Americans. Indisputably, this stratification criterion is simplistic, biased, and ethically provocative, but it illustrates the concept of inherent differences between individuals in the general population and serves as a bridging step from indiscriminant allocation of therapeutics to genetic-based stratification into risk/benefit subgroups. This example also illustrates the fact that the popular description of PGx as aspiring to "individually tailor" therapy is an exaggeration. The field is *not* centered around each individual specifically, but rather seeks to increase homogeneity among patients to an acceptable efficacy and safety level. In other words, PGx aspires to generate an individualized guide to proper therapeutic option among predictable segments of the patient population (Roses, 2004).

The discipline of PGx is closely tied to human genetics and genomics as a whole, but focuses on specific phenotypes of interest. While disease predisposition research in the realm of non-Mendelian traits is faced with extremely complex etiology, dictated by interplay between various pathways in the human body, it is assumed that pharmacogenetic phenotypes are largely governed by a limited number of molecules (an oligogenic structure), most of which interact with the drug as pharmacokinetic or pharmacodynamic entities. Although genetic factors associated with underlying disease etiology may impact treatment response features, these effects are hypothesized to have relatively modest sizes comparison to genetic variations in molecules that come in physical contact with the drug upon its administration. Moreover, the phenotypic changes that are induced by drugs are clearly and reproducibly caused by treatment exposure, which renders them less difficult to relate to genotype in comparison to studies of other complex traits. Thus, PGx research is equipped with the enormous advantage of being more readily and fruitfully realized in clinical practice. In addition, the (potential) direct application of pharmacogenetic findings in the daily management of disease burden prompts this field to be attractive, not only to academic circles but also to the pharmaceutical industry, patients, health care providers, regulators, and health systems as a whole.

Any treatment for which inter-individual variability is significantly larger than intra-individual diversity indicates an underlying genetic effect associated with response variables. In order to maximize the potential benefit from PGx research, it is helpful to focus and prioritize the analyzed therapies as follows (Table 15.1):

- Chronic disease states: in these cases disease burden on both the patients and health systems translates into compromised patients' quality of life, continuous suffering, low medication adherence rates, polypharmacy, and ultimately considerable care management costs. Therefore, management of chronic diseases comprises an attractive aim for optimization of available treatments and development of new therapeutic solutions (e.g., hypertension, diabetes, asthma, and so on).
- Narrow therapeutic index drugs (i.e., those with little difference between toxic and therapeutic doses) require close and constant dosage adjustment procedures to achieve therapeutic levels and minimize toxicity (usually by repeated measurements of therapeutic drug monitoring (TDM) protocols). Utilization of predictive genetic tests that may identify individuals at risk – either prior to treatment initiation or during the initial dose adjustment period – will save patients' lives, while diminishing substantially the costs of care management (e.g., warfarin, 6-mercaptopurine, and so on).

TABLE 15.1 Circumstances defining high priority targets for pharmacogenetic investigation

Circumstances	Rationale	Examples
The underlying condition being treated is chronic, incurable, and/or recurrent	Arrest the progression of the underlying disease, reduce patients' suffering, and decrease disease management burden on the health system	Epilepsy (Tate et al., 2005) Multiple sclerosis (Grossman et al., 2007)
Current clinical practice of the underlying condition follows a "trial and error" paradigm		Asthma (Weiss et al., 2006) HIV (Telenti and Goldstein, 2006) Hypertension (Arnett et al., 2006)
The studied drug has a narrow therapeutic index	Save patients' lives and decrease hospitalization costs associated with ADRs	Warfarin (Aquilante et al., 2006) TPMP (van den Akker-van Marle et al., 2006)
The studied drug is associated with severe, life-threatening side-effects		Herceptin (Krejsa et al., 2006)
The predictive value and clinical utility of the studied drug are high	Develop accurate and useful clinical diagnostic tools	Abacavir (Warren et al., 2006) TPMT (van den Akker-van Marle et al., 2006)
Rescue drugs post Phase II development	Market drugs to patient subpopulations that are predicted to benefit from the treatment, protecting potential non- or adverse responders	GW320659 (Spraggs et al., 2005)

- Disease states for which current practice follows a "trial-and-error" paradigm due to absence of predictive markers of efficacious and safe treatment regimens often require long adjustment periods, thus resulting in months or even years of incrementing doses and drug switching until a successful course of treatment is identified. Predictive genetic tests may assist in reducing this elongated adjustment period, resulting in increased compliance rates, and reduction of side effects, disease management burden and costs.
- As an integral part of the drug research and development process, rescuing formulations efficacious in subgroups of the patient populations, but failing to exhibit overall benefit in the all-inclusive cohort.

Noteworthy is the fact that most of the current examples in the PGx literature, including those discussed here, investigate marketed drugs retrospectively, when the drug and its clinical responses are available to be studied by academic investigators. A comprehensive review of PGx research along the drug development process with examples from discovery through phase III trials is presented elsewhere (Roses et al., 2006).

PHARMACOGENETIC STUDIES: FROM CONCEPT TO PRACTICE

The Ultimate Goal: Definition of the Key Research Question

A PGx study aims at identifying genetic determinants associated with drug–response features. The latter focus mainly on treatment efficacy, dosing, and safety. The definition of each of these elements is not trivial and is tightly dependent upon the specific parameters of a given dataset: the underlying disease and relevant clinical follow-up measurements dictate the available tests and clinical information that is not only relevant, but also obtainable, especially with respect to retrospective data acquisition. For example, in an investigation of the response to multiple sclerosis, treatment efficacy may be determined by clinical observations evaluated by the Extended Disability Status Scale (Kurtzke, 1983) (EDSS), reduction in relapse rate, severity of relapses, and/or MRI findings estimating T1-weighted hypointense lesions and central nervous system (CNS) atrophy measures (Grossman et al., 2007). Even a phenotype as "simple" as dosing may be regarded by various experts in a host of different ways. For instance, dosing phenotypes relevant to treatment with anti-psychotics for the treatment of schizophrenia are often assessed by "maintenance dose," which may reflect halving the drug dose that was proven effective during the acute phase (Gaszner and Makkos, 2004) or a titration regimen starting from minimum doses determined experimentally up to a recommended target dose on average (Perry and Hradek, 2004). When considering phenotypes of treatment safety recombinant protein therapeutics (including supplements to, or blockers of, endogenous proteins, immunomodulatory molecules and small, highly selective, targeted binding proteins), as opposed to small-molecule drugs, introduce an additional dimension requiring consideration of their potential immunogenicity (Krejsa et al., 2006). Whatever the case may be, the phenotypes to be tested must be determined prior and blinded to acquisition of results and are required to be precise, consistent, obtainable in the current investigation, reproducible and clinically relevant.

Tailor-Design of Your Study by Available Resources and Specific Goals

Researching the genetic determinants underlying a phenotype of interest is a task largely pre-defined by the specific characteristics of the studied phenomena. The investigator has no control over attributes such as phenotype penetrance; frequency of the (unknown) causative polymorphism(s); number of involved and modifier genes; genotype relative risk; phenocopy and genocopy effects; linkage disequilibrium extent between the typed markers and the causative polymorphism(s); and so on. Notwithstanding, an array of tools are available at the disposal of the pharmacogenetic researcher aimed to better the chances of identifying genotype–phenotype associations by sophisticated tailor fitting each individual project (Figure 15.1):

- collection of a large enough sample size as reflected by power estimations (Chapman et al., 2003; Kelly et al., 2005; Singer et al., 2005);
- distinct and precise, yet simple and broad, phenotype definitions;
- skilled marker selection (as detailed later); and
- appropriate choice of study design.

Classic Case–Control Study Design

The classic most "simplistic" study design, which is still considered the basic and fundamental association test pursued in genetic literature, is the case-control study. Originally, this analysis was used for testing significant differences in exposure without attempting to quantify the risk associated with exposure: *Statistically, do more lung cancer patients have a history of smoking than controls?* Rather than *by how many times does smoking increases the risk of lung cancer?* Case-control studies contain only group-level information and may determine the association between potential cause and effect on an *individual* basis. In pharmacogenetic investigations, cases and controls do not refer to health states or existence of drug exposure (affected versus unaffected, treated versus untreated), but rather to subpopulations within a source population of affected *and* treated individuals: positive responders versus non-responders, or patients manifesting ADRs versus safely treated individuals. Investigation of quantitative traits (as opposed to qualitative traits) is traditionally pursued by linear regression methods assuming that the *Y*-axis variable is a continuous measurement (e.g., optimized drug doses prescribed to schizophrenia patients [Grossman et al., 2008]). In order to avoid bias due to confounding effects inherent in the study populations while employing qualitative, as well as quantitative, phenotype definitions, and so as to allow follow-up replication analysis within the total or the remaining trialed population (Figure 15.2), it is

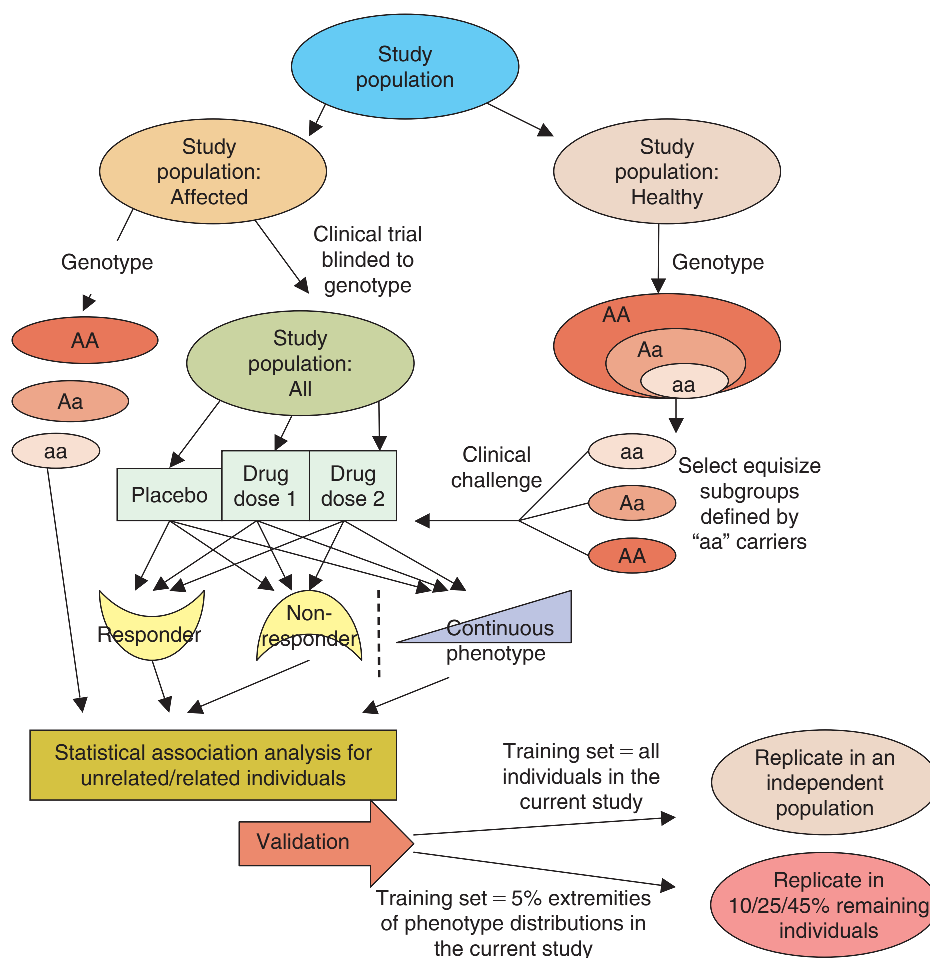


Figure 15.1 Pharmacogenetic study design alternatives in a simplified case, genotyping a single nucleotide polymorphism in a study population. In research settings pharmacogenetic study populations may either be healthy or affected subjects. *A priori* genotyping may be used as inclusion criteria and in several models may be used to reduce study population, when equisize groups of genotype carriers (aa, aA, and AA) are randomized to treatment and subsequent analysis. In clinical trial settings, individuals are randomized to treatment groups (active drug, parallel dose comparison arms, gold standard comparator drug, and/or placebo) blinded to genotype. Only once trial is completed and efficacy and safety end-points are clinically analyzed does statistical analysis take into account genetic underlying effects. Similar approach can be utilized in adaptive trial design for interim analysis. It is expected that some genetic associations will reach statistical significance at random and thus it is vital to validate any association results in either the replication set within the current cohort, or in an independent population.

becoming acceptable to employ case–control designs in which cases are chosen as patients exhibiting the lowest 5% of a phenotype distribution, while controls represent the 95th percentile of this distribution. Contrasting these two subgroups in the initial analysis enables focusing on the most extreme representations of the phenotype of interest with a lenient significance level, accentuating the roles that causative factors (genetic and environmental) play in exhibition of these phenotype states, and expecting most of the findings to reflect false-positive associations. In the second stage, designed to replicate and validate the obtained results, individuals represented by the remainder of the distribution space are split into cases and controls. Analysis can be performed at this level by using a stringent Bonferroni-corrected significance level as a "replication-based analysis" or by combining test statistics from both stages and applying significance level corrected only for the number of markers tested, titled "joint-analysis" (Skol et al., 2006). The above-described alternatives of one-stage versus two-stage association study designs are summarized in Figure 15.2.

Familial Study Design

Familial and twin genetic association studies allow refined elucidation of genetic associations while inherently correcting for environmental effects and risk factors (Endrenyi et al., 1976; O'Reilly et al., 1994). The task of collecting appropriate related cohorts suffering from the same conditions and exposed to the same pharmacological agents is often not trivial (although has been demonstrated to be helpful in the case of nicotine

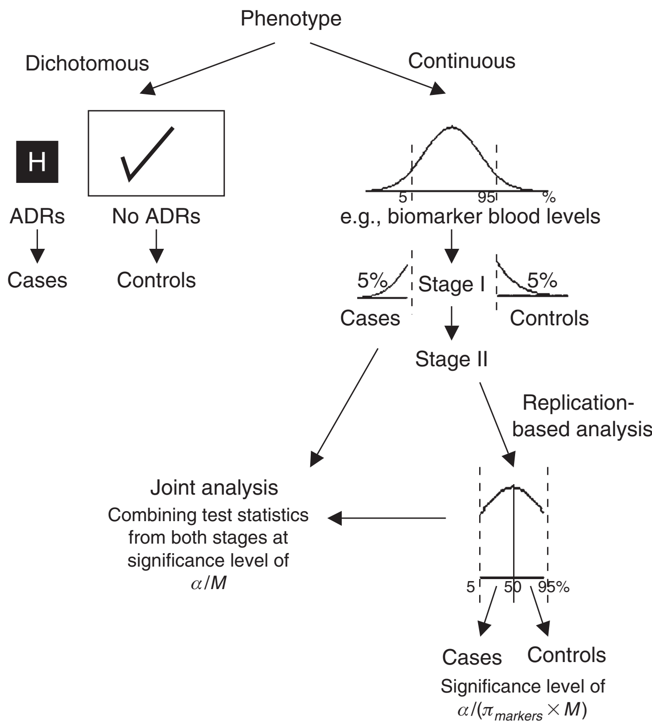


Figure 15.2 One-stage versus two-stage case–control genetic association analyses. The phenotype of interest in pharmacogenetic studies can be categorized as either continuous or dichotomous. When occurrence of adverse drug reactions is analyzed, the natural statistical model calls for a case-control design. However, efficacy end-points tend to display a continuous distribution, such as measurements of biomarker blood levels. A two-stage design would split the study population into (Stage I) a case–control analysis contrasting the 5% extremities of the Normalized distribution; thereafter (Stage II) would employ a replication-based analysis of the remaining 90% of the distribution with adjusted requirements for significance level. A joint analysis will then ensue, combining test statistics from both stages at a corrected significance level (Skol et al., 2006). ADR, adverse drug reaction; α, study false-positive rate (type I error rate); M, number of genotyped markers; $\pi_{markers}$, the proportion of markers found to be associated in Stage I and followed up in Stage II.

metabolism [Swan et al., 2004] and alcohol exposure [Heath and Martin, 1994]). However, implementation of healthy volunteer challenge studies (see later) within related cohorts may be highly effective (Kendler et al., 1988).

Healthy Volunteer Challenge Design

Pharmacogenetic investigations, as opposed to disease predisposition genetic analyses, bear the extraordinary advantage of inducing the phenotype of interest in any given individual – including in healthy volunteers. In other words, in order to study the effects of drug intake in a genetic context, one can either investigate affected individuals prescribed the drug in question for the therapeutic purpose of controlling their condition (the traditional approach) or administer the drug to healthy consenting individuals specifically for the purpose of PGx research. This unique option allows not only the rapid recruitment of study population to be screened for genetic variation but also enables specific tailoring of the study design to address PGx questions, rather than clinical trial end-points.

In practice, researchers assess the impact of pre-selected polymorphisms on the PGx phenotypes (e.g., efficacy, safety, dosing, and pharmacokinetic properties) of the studied drug, by prospective clinical analysis of healthy volunteers chosen conditioned on their genotype profile. In this manner, Kirchheiner et al. (2004) identified a total of 25 healthy males carrying one of three genotypes of interest (*CYP2D6* ultra, extensive, and poor metabolizers) out of more than 1000 healthy volunteers in a PGx study of Mitrazapine and clinically challenged this miniature group of interest for clinical measurements. This cohort size matched power estimates for detecting at least a 1.5-fold effect on drug clearance differences between subgroups. Similar designs have been published for a variety of pharmaceutical applications and bear a great potential for simplification of PGx research in agents with favorable safety profiles.

Placebo and Multiple Dosing Cohort Comparisons

PGx association studies in general, and those of medications for psychiatric and immunological diseases in particular, are characterized by a poor signal-to-noise ratio: approximately one-third of the patients enrolled in efficacy trials may respond to placebo treatment. The placebo "response" in randomized clinical trials includes such statistical artifacts as regression to the mean, drift in measurement of the response over time, and bias of expectations by both patients and evaluators, as well as real effects such as spontaneous recovery, a tendency to seek treatment outside the study, and the response to additional attention and concern involved with participating in clinical trials (Singer et al., 2005). It has been shown that placebo "response" strongly affects statistical power of association studies and that employing a placebo-treated control group in the PGx analysis enables elucidation of specific drug-induced genetic effects, differentiating these from confounding effects related to underlying disease progression and severity (Grossman et al., 2007). Similarly, employment of trial arms administering different drug doses may shed light on the genetic underlying mechanisms and their magnitude (Risner et al., 2006).

MARKER SELECTION – STRATEGY AND APPLICATION

Candidate Genes Strategy Versus Whole-Genome Scans

A comprehensive search for genetic influences on drug response would involve examining all genetic differences in a large number of affected individuals and controls. This will eventually

(and probably not too far ahead) be achievable by complete genome re-sequencing. However, current practice is limited to a systematical test of common genetic variants (those with 2–5% minor allele frequency [MAF] and higher). These variants explain most of the genetic diversity in human populations and may be sampled taking advantage of the genetic property termed linkage disequilibrium (LD) (see also Chapter 4). LD reflects shared ancestry of contemporary chromosomes, measuring the degree of non-random co-occurrence of two (or more) polymorphic sites along a stretch of genomic sequence. Polymorphisms showing a high degree of LD may be considered redundant, eliminating the necessity to type all SNPs in this correlated group. It suffices to genotype a representative SNP and virtually represent the entire variation captured by the whole group, given the LD threshold imposed on the model.

Traditionally, studies of common traits have been studied by two different strategies: family-based linkage studies across the entire genome and population-based association studies of individual candidate genes (usually only a handful). Although there have been notable successes, progress has been slow due to the inherent limitations of both methods: linkage analysis has low power except when a single locus explains a substantial fraction of disease, and association studies of one or a few candidate genes examine only a small fraction of the "universe" of sequence variation in each patient. Progress in technology and the development of sophisticated statistical tools have paved the way toward the application of Whole Genome Association Studies (WGAS) in a population-based cohort (see Chapter 4). This progress makes it possible to study the full scope of possible genetic determinants involved with phenotypes of drug response in a hypothesis-free fashion (i.e., unbiased approach). However, the quest for the Holy Grail, that is, the identification of causal polymorphisms underlining the phenotype of interest, remains a key unanswered question even when interrogating the entire genome.

While many approaches have been suggested in the literature as to how to decipher between true associations and false-positive results, no consensus is likely to be established in the next few years (for recommendations on replication design in large-scale association studies see the NCI-NHGRI Working Group on Replication in Association Studies [2007]). Expert genetic, biological, and pharmaceutical skills will be needed to establish a prioritization scheme by which result interpretation will reflect not only statistical plausibility but also scientific rationale. Independent comparative studies will eventually give rise to the most successful strategies that should be pursued. Hence, the strategy of choice for years to come will progressively rely on a single universal tagging set that will be used in all whole-genome association studies. This would not only ensure comparability for replication efforts within complex traits but also allow direct comparison of the role of the same variants as risk factors for different conditions. Led by this vision, The National Center for Biotechnology Information (NCBI) is currently managing the submission, storage, and access for clinical phenotype measures and associated whole-genome genotype data for several different programs, making this data as widely available to the research community as possible while protecting the privacy of the participants as defined by consent agreements for individual projects.

Application of whole-genome scans for PGx phenotypes is currently being undertaken by various academic and pharmaceutical researchers. One pioneering example depicts a WGAS investigating drug-induced liver injury ADRs and represents the incorporation of PGx in pharmaceutical R&D (Kindmark et al., 2007).

Pharmacokinetics- Versus Pharmacodynamics-Based Gene Selection

Once administered, a drug goes through a series of interactions in the body that are responsible not only for the therapeutic effect conveyed by this agent, but also for side effects and bystander alterations to other compounds processed in conjunction. As a whole, all interactions related to modifications acted upon the drug by the host body are termed "drug pharmacokinetics" and include all aspects of drug absorption, distribution, metabolism, and elimination (ADME). The most prominently studied molecule group within the sphere of pharmacokinetics has indisputably been drug-metabolizing enzymes (DMEs). These enzymes exist in every eukaryotic cell and in most, if not all, prokaryotes and are thought to have evolved in humans to neutralize xenotoxins and/or to control concentrations of signaling molecules in endogenous pathways (Nebert and Dieter, 2000). The fact that modern drugs have been introduced to the *in vivo* environment very recently in evolution uniquely excludes selection forces from having impacted the functional genetic properties of DMEs, thus potentially permitting whopping effects on drug-response phenotypes (Nebert and Dieter, 2000). In the early years of PGx research and up until very recently, genes encoding pharmacokinetic properties of drugs have been the focus of research (for a comprehensive table of drug–drug interactions based on cytochrome P450 enzymatic activity see [Flockhart, 2007]). This has been due to the scientific expectation of marked clinical effects underlined by diminished or absent capacity of the body to obliterate drugs or transform prodrugs into active compounds, but also due to the pragmatic constraint posed by considerable gaps in our understanding of, and technological methods for, investigating genes encoding drug pharmacodynamic features. Pharmacodynamics relates to any modifications acted upon the host body by the drug, that is, drug targets (receptors or enzymes) and their downstream signal pathways.

With the advent of the Human Genome Project, the HapMap project (see below) and incredible progress in genotyping techniques and costs, it is currently feasible to interrogate known functional, as well functionally as-yet-unknown, genetic variants in pharmacodynamic candidate genes. It is now becoming increasingly evident that polymorphisms within genes involved in the pharmacodynamic properties of drugs may in fact be more important and clinically relevant than variants of ADME genes for non-Mendelian drug-response phenotypes. Examples supporting this notion include reports on the relative contribution of variants in the *VKORC1* and *CYP2C9* genes to warfarin

response (Aquilante et al., 2006); *SCN1A* and *CYP2C9* to phenytoin response (Tate et al., 2005); and *5-HTR$_{2A}$* and *CYP2D6* to paroxetine response (Murphy et al., 2003).

Pathway-Based Analyses and Search for Epistatic Associations

Advances in haplotype tagging strategies have made it possible to represent variation in entire pathways relevant to drug response economically. The focus in pathway-based PGx investigation is further facilitated by the accumulation of knowledge regarding pathways relevant to drug response (e.g. PharmGKB Pathways). Although promising, pathway-based analyses face considerable challenges. Most fundamentally, the huge number of tests conducted when analyzing large SNP datasets, multiple phenotype definitions, and application of a diverse spectrum of parametric and non-parametric statistical methodologies pose considerable barriers on power, statistical design, and interpretation. Increasing the population sample size, applying permutation-based multiple correction methods, and seeking independent replication are obvious solutions dealing with these difficulties. A further complication is that all components of all pathways that a drug acts on are rarely understood completely (Need et al., 2005).

As PGx investigations rapidly progress toward a pathway-centered approach, considering dozens and hundreds of genes in each instance, or assigning prior probability scores of whole-genome scans founded on the biological processes involved with a given drug's PGx, it is progressively evident that research should embrace methods that take into account possible interactions between genetic variants along a pathway of interest. It is very likely that, for instance, genetic-based conformational changes brought forth in one molecule along a drug-related pathway will require complementary structural, quantitative, or functional changes in a downstream (or upstream) molecule to produce a detectable sum total effect on phenotype measurements. This genetic interaction is termed "epistasis" and it is yet unclear how best, computationally and statistically, to surmount the complexity arising from the combinatorial magnitude of all the possible interactions involved within a pathway. Moreover, the high dimensionality of the resulting analysis overwhelms traditional analysis methods, a problem exponentially prohibitive for large candidate gene studies and whole-genome investigations. The Multifactor Dimensionality Reduction (MDR) approach (Wilke et al., 2005) has been suggested as one potential solution for this problem, although sophisticated analytical approaches capable of modeling high order interactions and sizable datasets will have to be developed in order to handle genome-wide and large-scale candidate gene studies.

Gene-expression and proteomic profiling may in the future provide additional knowledge and complementary diagnostics associated with pathways relevant to drug response, although at this time it is difficult to differentiate incidental differences in expression levels from those that are causally relevant to drug response (Need et al., 2005). In conclusion, thanks to technological and computational advances, it is no longer justifiable to look at the effects of single genes in isolation, but rather incorporate the milieu of potential effects more comprehensively describing the sum total of a studied phenotype.

HapMap – A Public Resource for Marker Selection

The International HapMap Project was launched in October 2002 to create a public, genome-wide database of common human sequence variation, providing information needed as a guide to genetic studies of clinical phenotypes (The International HapMap Project, 2003). Four different ethnic populations were systematically genotyped for each SNP variant as follows: (i) 90 individuals (30 parent–offspring trios) from Yoruba, Ibadan, Nigeria; (ii) 90 individuals (30 trios) from Utah, USA (the Centre d'Etude du Polymorphisme Humain collection); (iii) 45 non-related Han Chinese from Beijing, China; and (iv) 44 non-related Japanese from Tokyo, Japan.

The HapMap project set out to describe patterns of association between typed polymorphisms (LD) in the above-described four population samples in order to identify a minimum set of SNPs ("tagging SNPs") sufficient to represent all common variants in the human genome. A publicly available software for tagging has been incorporated into the HapMap website (description and a more detailed version are available in de Bakker et al. [2005]), enabling any given user to select a desirable set of tag SNPs based on a variety of criteria (such as population ethnicity, force inclusion of SNPs with known functional effects or which have previously been described in the literature, and so on). The chief criterion for tagging capacity employed is the r^2 threshold. This is a study-independent measure of SNP utility that has become a leading standard for evaluating performance of marker sets. It represents the correlation coefficient between any observed marker and a putative causal allele, reflecting the expected drop in non-centrality of an association test statistic under specified conditions (Pe'er et al., 2006). It is customary to enforce a minimum pair-wise r^2 threshold of 0.7–0.8.

Analyses increasingly attest to the remarkable performance of tagging (Conrad et al., 2006; de Bakker et al., 2006), indicating that the HapMap samples provide an appropriate resource for selecting globally useful tags (although African populations remain partially unrepresented by tagging approaches due to their high diversity). Thanks to the fact that all phases of the HapMap database have been publicly available ever since the projects' inception, both commercial companies and academic laboratories routinely implement HapMap data in their drug research and development programs. In the realm of pharmacogenetics, for instance, elucidation of differential response to, and dosing of, warfarin anti-coagulation treatment as governed by *VKORC1* was performed using the HapMap data (Rieder et al., 2005). Another example, specifically designed to examine the capacity of HapMap tools to discover and validate PGx phenotypes, was recently published (Jones et al., 2007). In this study, HapMap SNPs in the thiopurine methyltransferase (*TPMT*) gene region (tags) were tested for association with cell line absolute levels of TPMT activity. It was expected that a variant of biochemically and clinically validated relevance to thiopurine plasma levels and induced toxicity will rank as most significantly associated in all analyses. Indeed, this SNP and others in high LD with it tested positively, out of a total of 66 HapMap

SNPs in the *TPMT* gene region. However, an agnostic genome-wide approach to assess genomic predictors of TPMT phenotype in this dataset ranked *TPMT* haplotypes at 97th place out of the 17,542 genes tested. The latter result demonstrates that whole-genome tagging approaches are useful screening tools for the highest few percent of associated genes; nonetheless, ultimately biology, biochemistry, and clinical expertise are required to decipher true associations from false-positives.

Validation of Association Results

Any results detected in "first stage" genetic analyses require further validation. This is usually accomplished by two lines of investigation: functional analysis and replication studies. Functional studies are aimed at revealing direct causality between the reported genetic variant and measurable functional modifications elicited in the encoded molecule by employing animal models (e.g., knockout mice) or *in vitro* investigations. Replication studies are performed in independent patient populations mimicking as closely as possible the study design settings and testing only the top priority fraction of SNP associations detected in the first stage. Replication analysis serves not only for confirmation of true association between a genetic marker and a phenotype of interest, but it is also a means for better assessing the true magnitude of the detected effect, as first publications tend to report inflated effects due to the phenomenon termed "the winner's curse" (Ioannidis et al., 2001).

Prospective pharmacogenetic analyses compromise an important step toward the validation of genetic markers as predictors of clinical management in "real-life" settings (Grossman, 2007). Although few such studies have been conducted thus far, it is becoming generally accepted that prospective analyses comprise necessary and integral evidence of the potential utility of genetic tests in routine health care management. Initiatives already prospectively monitoring patients for PGx outcomes include several studies aimed at analysis of cost-effectiveness of *HLA-B**5701* genotyping in abacavir-treated HIV-positive patients prone to suffer from hypersensitivity reaction (Hughes et al., 2004, 2007; Rauch et al., 2006; Zucman et al., 2007); a genotype-guided clinical cancer trial for management of rectal cancer by irinotecan therapy based on analysis of a polymorphic tandem repeat in the 5′-untranslated region of the *TYMS* enhancer region, resulting in 2–9 copies of a 28-bp repeated sequence (McLeod et al., 2005); and genotype-guided dosing analyses of warfarin therapy (Hillman et al., 2005).

FROM BENCH TO BEDSIDE: INTEGRATION OF PHARMACOGENETIC TESTING INTO CLINICAL PRACTICE

Analytic Assessment of Pharmacogenetic Tests and Their Utility in Clinical Practice

Demonstration of unambiguous association between genotype(s) and drug-response features in controlled retrospective datasets, albeit promising, does not guarantee the usefulness of routine genetic testing for the clinical decision-making process in the daily clinical environment. A whole host of factors need to be further assessed in order to predict and determine the cost-effectiveness and practical utility of a pharmacogenetic diagnostic test. This evaluation should follow recommendations such as those devised by the Centers for Disease Control (CDC, 2006), including four components: Analytic Validity; Clinical Validity; Clinical Utility; and Ethical, Legal, and Social implications. Each of these components focuses on a different aspect of test performance and applicability in real-life settings, starting at the laboratory performing the test, through clinical studies proving its benefits, to more subjective evaluation of managed care burden associated with introduction of the test to routine practice and reimbursement frameworks. Each of these components can be empirically measured to reflect high validity and clear utility standards:

1. *Analytic validity* of a genetic test defines its ability to accurately and reliably measure the genotype of interest. Empirically, analytic validity includes *analytic sensitivity* (or the analytic detection rate, =number of polymorphism carriers detected by the test divided by the total number of polymorphism carriers in the population); *analytic specificity* (=number of normal genotype carriers detected by the test divided by the total number of individuals who do not carry the polymorphism); *laboratory quality control* (including measurements such as inter- and intra-assay variability, repeatability, and so on); and *assay robustness* (measuring how resistant the assay is to changes in pre-analytic and analytic variables, such as inter-laboratory variability, DNA extraction methodology, and so on).
2. *Clinical validity* of a genetic test defines its ability to detect or predict the associated phenotype. Empirically, clinical validity includes *clinical sensitivity* (or the clinical detection rate, =the proportion of individuals who have a well-defined clinical phenotype and whose test values are positive); *clinical specificity* (=the proportion of individuals who do not have the well-defined clinical phenotype and whose test results are negative); *prevalence of the specific phenotype* (=the proportion of individuals in the selected setting who have, or who will develop, the phenotype); *positive predictive value* (=the proportion of individuals with a positive test result who have, or will develop, an unwanted response when the drug is administered, PPV); *negative predictive value* (=the proportion of individuals with a negative test result who will not have, or will not develop, an unwanted response when the drug is administered, NPV); *penetrance* (the frequency, under given environmental conditions, with which a specific phenotype is expressed by those individuals with a positive test result); and *modifiers* (genetic or environmental).
3. *Clinical utility* of a genetic test defines the elements that need to be considered when evaluating the risks and benefits

associated with its introduction into routine practice. Empirically, clinical utility includes *administration settings recommendations* (such as optimal time line for administration of the test, either prior to treatment initiation, during the first phase of dosing adjustments, or else); *availability and effectiveness of counter-interventions*; *quality assurance* (procedures in place for controlling pre-analytic, analytic, and post-analytic factors that could influence the risks and benefits of testing); *pilot trials* (assessing the performance of testing under real-world conditions, where inclusion criteria allow for individuals to experience co-morbidities, take concomitant medications, and so on); *health risks* (adverse consequences of testing or interventions in individuals with either positive or negative test results); *economic evaluation* (financial costs and benefits of testing); *facilities* (assess the capacity of existing resources to manage all aspects of the service); *education* (quality and availability of informational materials and expertise for all aspects of a screening service); and *monitoring* (assess a program's ability to maintain surveillance over its activities and make adjustments).

4. *Ethical, legal, and social implications* surrounding a genetic test represent the safeguards and impediments that should be considered in the context of all the other components.

FDA Regulation Policy and Its Impact on PGx Testing Application in Real-Life Health Care

The FDA has recognized the importance of PGx and encourages its use in drug development. Recognizing that the clinical component of the overall cost of successful new drug development is ~58% of total costs (DiMasi et al., 2003) and that ~50% of Phase III trials fail, the FDA is promoting identification of PGx biomarkers at early drug development stages that would predict which drugs are likely to either fail or succeed in Phase III trials. This approach is reflected in the FDA's 2004 white paper (FDA, 2004a), which identifies PGx as a key opportunity for the Critical Path. In addition, the FDA believes that there is value in applying long-established PGx testing to older, marketed drugs in the post-marketing period to improve their risk/benefit ratio by optimizing or individualizing dosing (Lesko and Woodcock, 2004). While multiple enthusiastic statements supporting the incorporation of PGx analysis to both new drug development and monitoring of marketed drugs have been made by FDA officials (Frueh et al., 2005; Lesko and Woodcock, 2002, 2004), there has not been a proactive regulatory action enforcing genomic data submission. In fact, ever since the publication of *Guidance for Industry: Pharmacogenomic Data Submissions* in 2004 (FDA, 2004b), relatively few drug labels have been updated to include recommendation for PGx testing (Table 15.2). The current policy refrains from enforcement of submissions, while promoting Voluntary Genomic Data Submission (VGDS). It is remarkable to note that despite the non-compulsory nature of this guiding principle, VGDS submissions are rapidly amassing and were filed for over 30 different compounds in the first year, exhibiting a growth trend going forward (Frueh, 2006).

EXAMPLES OF PGx TESTS: PROMISING NEW DEVELOPMENTS AND MARKETED PRODUCTS

Warfarin

Perhaps the best example of a successful pharmacogenetic association, for which the clinical relevance is clear, features management of treatment with the oral anti-coagulant warfarin, Warfarin, a derivative of coumarin, is prescribed for the long-term treatment and prevention of thromboembolic events, with more than 21 million prescriptions annually in the United States alone (Rieder et al., 2005). However, due to the drug's narrow therapeutic index, a variety of complications are associated with its treatment, even after dose adjustment according to age, sex, weight, disease state, diet, and concomitant medications. Investigation of pharmacokinetic and pharmacodynamic drug properties indicated the additive involvement of two genes in determination of warfarin maintenance dose. The first gene to be identified encodes CYP2C9, which is responsible for most of the metabolic clearance (~80%) of the more pharmacologically potent *S*-enantiomer of warfarin (Rettie et al., 1992). Both *CYP2C9*2* and **3* cause a reduction in *S*-warfarin clearance, with 10-fold variation seen from the genotype linked with the highest (*CYP2C9*1/*1*) to lowest (*CYP2C9*3/*3*) activity. Numerous studies have associated these genotypes with initial dose sensitivity, delayed stabilization of maintenance dose, delays in hospital discharge, and increased bleeding complications (Gardiner and Begg, 2006). However, it is estimated that *CYP2C9* variants account for only 10–20% of the total variation in warfarin dose, with additional genetic and environmental factors playing larger roles in dose determination. The second gene identified as a predictor of dosing encodes the vitamin K epoxide reductase complex protein 1 (VKORC1), targeted by warfarin. Consideration of the *VKORC1* genotype or haplotype, together with the *CYP2C9* genotype, and factors such as age and body size are estimated to account for 35–60% of the variability in warfarin dosing requirements.

Despite the fact that these data have been reproduced by multiple independent groups, it remains to be tested in large, prospective, clinical studies whether initial dose may be tailored to patients by *CYP2C9* and *VKORC1* genotyping, coupled with known clinical variables. While these studies are currently under way, the clinical pharmacology advisory panel to the FDA acknowledged (November 2005) the importance and potential for genotyping of *CYP2C9* and *VKORC1* during the early phase of warfarin therapy, and the drug label was amended accordingly in August 2007. A month later, the FDA cleared for marketing the Nanosphere Verigene Warfarin Metabolism Nucleic Acid Test to aid physicians manage warfarin therapy.

Trastuzumab

Trastuzumab therapy for the treatment of breast cancer (involving a monoclonal antibody specifically targeting HER2/*neu*

Table 15.2 FDA valid pharmacogenetic biomarkers – associated dosage guidelines in drug labels and availability of approved genetic tests (Summer, 2007)

Drug [# drugs with similar label references]	Gene (*variant*)	Gene function/ significance	Dosage guidelines	Label section	PGx test?	FDA-approved test available?
Voriconazole [6]	CYP2C19 (*poor metabolizers*)	Major metabolizer	1. Contraindicated concomitantly with CYP450 inducers; 2. Increased maintenance dose regimen outlined for Phenytoin; 3. No adjustments[a] for oral contraceptives inhibiting CYP2C19 or HIV protease inhibitors that inhibit CYP3A4; 4. Reduce Cyclosporine and Omeprazole by half; 5. Reduce Tacrolimus to one third	Clinical pharmacology; Drug Interactions;	No	AmpliChip/Roche
Celecoxib [1]	CYP2C9 (*poor metabolizers*)	Major metabolizer	No	Clinical pharmacology; Drug interactions;	No	No
Atomoxetine [17]	CYP2D6 (*poor metabolizers*)	Major metabolizer	Specific regimen suggested when strong CYP2D6 inhibitor co-administered	Clinical pharmacology; Drug interactions; CYP2D6 metabolism; General dosing information	No	AmpliChip®/Roche®
Capecitabine [2]	DPD (*deficiency*)	Rate-limiting metabolizer	Contraindicated	Contraindications; Clinical pharmacology; Precautions	No	No
Rusburicase [3]	G6PD (*deficiency*)	Severe hemolysis caused by hydrogen peroxide by-product	Contraindicated	Boxed warning; Contraindications; Warnings	Recommended	No
Rifampin, isoniazid, and pyrazinamide [1]	NAT (*slow acetylator*)	Isoniazid major metabolizer	No	Clinical pharmacology; Adverse reactions; Drug interactions	No	No
Azathioprine [2]	TPMT (**2, *3A and *3C*)	Metabolizer	Recommended altered regimen, but not specified	Warnings; Lab tests; Dosage and administrations; Drug interactions	Recommended	No
Irinotecan	UGT1A1 (**28*)	Conjugation of active metabolite	Consider reduction of dose in homozygotes	Clinical pharmacology; Warnings; Dosage and administration	No	Invader assay/Third Wave

This table is adopted from Grossman (2007). Details in this table are derived from package inserts of the listed drugs and the table of valid genomic biomarkers in the context of approved drug labels http://www.fda.gov/cder/genomics/genomic_biomarkers_table.htm.

Note – A drug label change was introduced in August 2007 *VKORC1* (in addition to *CYP2C9*) genotyping in patients requiring warfarin therapy. An accompanying genetic test (Nanosphere Verigene Warfarin Metabolism Nucleic Acid Test) was FDA-approved in late September 2007.

[a] Most adjustments relate to CYP3A4 drug–drug interactions, although it is a minor metabolizer of Voriconazole.

over-expressing tumors) is not only an example of a protein therapeutic for which an obligatory biomarker assay has been issued, but it is also an example of the utilization of PGx research for drug rescue: trastuzumab is marketed solely for the subset of patients (~10–25%) who over-express HER2/*neu*, providing care to eligible patients and returning the investment to the developing manufacturer. Interestingly enough, although in practice the obligatory biomarker assay has been set to measure protein over-expression by immunohistochemistry, studies have shown stronger association when patient subsets were determined by fluorescence *in situ* hybridization, reflecting gene copy number (Krejsa et al., 2006).

6-Mercaptopurine

In 1953, the drug 6-mercaptopurine (6-MP) was marketed in the United States for the treatment of leukemia. Despite great expectations in the medical world, about 20 years ago fatal toxicity was discovered in 0.3% of treated children. A similar scenario was repeated for azathioprine marketed in 1968. It was later discovered that polymorphisms within the *TPMT* gene underlie the large inter-individual differences in the enzyme's activity, leading to a high risk of thiopurine-induced toxicity in homozygotes for the defective alleles, and inadequate therapeutic efficacy in patients with high activity TPMT. Tests for TPMT activity (genotype, enzyme activity, and metabolite screening) are available in the United States and throughout Europe; however, clinical implementation of these tests is very low. This is contrary to expectations based on cost-effectiveness analysis of TPMT testing in children with acute lymphocytic leukemia (ALL) showing high savings per life-year (Zika et al., 2006). Current clinical practice for the management of leukemia via thiopurine medications dictates careful monitoring of white blood cell counts and clinical outcomes in a fashion that negates the necessity of incorporating the genetic test for prevention of serious adverse events. It is expected, however, that as genetic tests become generally accepted for a variety of conditions, it will become progressively acceptable both by physicians and health systems, as well as by the general public, to use *TPMT* genetic tests as prognostic tools.

AmpliChip

The first microarray-based gene-chip, approved both in the United States and EU, was released to the market by F. Hoffmann-La Roche Ltd (Switzerland) in 2003 as the AmpliChip CYP450®. The product was designed to identify key genetic polymorphisms in two CYP450 genes, *CYP2D6* and *CYP2C19*, the products of which are cumulatively responsible for much of the first pass metabolism of many pharmaceutical compounds. As highlighted by insurance companies who refuse to cover the costs ($600–$1300), the test has not been clearly demonstrated to convey clinical utility and sufficient cost-effectiveness (Grossman et al., 2008; Thakur et al., 2007) and is deemed "experimental, investigational or unproven." The regulatory agencies cleared this test based solely on analytical performance and validity information, but indicated its utility specifically for clinical application. Thus clinicians, as well as patients, were misled to expect firm impact on clinical decision-making guidelines. Until unambiguous evidence proves the clinical use of this and other genetic tests, caution is advised in their interpretation and application in health care management.

FUTURE DEVELOPMENTS REQUIRED FOR THE FIELD TO FULLY MEET ITS EXPECTATIONS

Table 15.3 summarizes steps still needed for PGx testing to materialize into practical tools used in clinical decision-making. First, in order to capture the entirety of processes involved in, and affected by, any given drug, both *in vivo* and in interaction with environmental factors, open communications between researchers, practitioners, regulatory agencies, and pharmaceutical industries must be established. Launching PGx educational programs within the academic curriculum will facilitate acceptance of genetic testing both by health care providers and by the general public.

Next, facilitated by the latter, large prospective studies must be conducted in order to truly evaluate the utility of genetic tests for specific indications in "real-life" clinical setting. Tailoring the research to specifically target PGx-based end-points, rather than adopting and artificially manipulating retrospective measurements from treatment efficacy clinical trials, will assist in elucidating "true" PGx effects. In this fashion, employing placebo, control, and multiple dosing treatment groups may act as means for detection of specific drug-induced genetic mechanisms.

As research progresses and technology develops the need for investigation of gene–gene and gene–environment effects in a computationally reasonable manner will become more and more pressing. Such lines of investigation will require large, often combined, datasets, necessitating the standardization of phenotype definitions by the research community.

Lastly, incorporation of genetic testing in clinical practice will only become possible if regulatory and funding agencies acknowledge its potential and promote its development by changing their policy. PGx is a field uniquely attractive to both academic medical centers and the pharmaceutical and biotechnological industries. As such, governmental agencies, such as the US FDA and the European Medicines Agency (EMEA), regulate the development, marketing, and clinical applications associated with maturation of research advances. In order for PGx to wholly fulfill its potential and expectations a structured framework of incentives, priorities, and policies must be devised that would promote allocation of resources by each of the concerned parties. A potential solution to improve surveillance might be sampling and privacy-protected DNA banking from the first 250,000 patients treated with a newly marketed drug, or a similar risk-management system coordinated with regulatory agencies (Roses, 2004).

The most pronounced deficiency identified in PGx research relates to the need to improve drug safety and efficacy profiles of generic, off-patent drugs. In this aspect it is important to mention

TABLE 15.3 Steps required for PGx to fully meet its potential as "the right drug for the right patient"

Future challenges

- Integrative approach: Cross-talk between geneticists, statisticians, pharmacologists, molecular biologists, physicians, regulatory agencies, and pharmaceutical companies.
- Education programs should be incorporated into the academic curriculum.
- Prospective studies have to be conducted to evaluate the potential predictive power and clinical utility of PGx testing in clinical settings. Possibly employ placebo and control treatment groups in analysis to elucidate specific drug-induced genetic effects.
- Focus on gene–gene interactions and the net genetic effect.
- Integrate environmental effects and gene – environment interactions into the statistical models.
- Standardize unequivocal and reproducible phenotype definitions.
- Genetics will get more and more acceptable as a tool for clinical practice as policy will catch up with scientific developments.
- Surveillance of ADRs in the post-marketing phase through risk-management systems coordinated by both manufacturers and regulatory agencies.
- Establishment of incentive structure for the industry to improve drug safety and efficacy beyond the terms of current patent protection.

that withdrawn medications lose their patent protection, and thus commercially driven pharmaceutical companies lose interest in rescuing these formulations by conducting safety PGx experiments. To date, it has yet to be demonstrated whether PGx would be cost-effective in resurrecting failed marketed drugs (Shah, 2006), but thorough investigations have yet to be conducted. Only controlled incentives, calling for prolongation of patent protection terms and similar motivations, will drive the pharmaceutical industry to invest the considerable costs and resources required.

2009 UPDATE

The last year has witnessed much progress in the field of PGx, with many publications and several regulatory changes that promote the overall usefulness and significance of PGx to clinical management and novel drug discovery and development.

Most of the published work has been focused on marketed drugs as conducted by various independent groups, collectively contributing both to the elucidation of PGx pathways and biological etiology, as well as the demonstration of clinical utility and economic cost-effectiveness. Examples that have already impacted the usage of drugs by drug label updates include carbamazepine, for which serious dermatological reactions have been highly associated with carriage of *HLA-B*1502* in subjects with ancestry from South, East, and South-East Asia; abacavir, for which hypersensitivity reaction has been highly associated with carriage of *HLA-B*5701*; and warfarin, for which dosing adjustment has been associated with variants in both *CYP2C9* and *VKORC1*. Cleared genetic diagnostics are available for all of the above, but in the case of warfarin, health outcomes and prospective genotype-guided and cost-effectiveness studies are only now being carried out to define the exact dosing adjustment algorithms that should guide healthcare providers in the management of patients.

Another important advance in the field was the adoption of PGx testing as clinical guidelines for the management of colorectal cancer with erbitux and vectibix. The National Comprehensive Cancer Network updated its colorectal cancer guidelines last year to include *KRAS* testing. More recently, the American Society of Clinical Oncology (ASCO) also recommended that patients test their tumors for *KRAS* gene status before treatment with Erbitux or Vectibix. It is expected that the US FDA will relabel these drugs in accordance soon.

Other published data have yet to reach regulatory discussion and decision-making, but have generated significant evidence on clinical impact that would clearly affect future treatment regulation. The first such example relates to a whole-genome association study (WGAS) conducted in statin-treated patients that identified risk variants in the *SLCO1B1* gene and strong association with an increased risk of statin-induced myopathy (Link et al., 2008). The authors studied two sets of patients and control groups from large trials, where the training set was derived from about 1200 subject study (80 mg simvastatin once daily), and tested in patients derived from a 20,000 subject trial (40 mg simvastatin daily). This genome-wide study illustrated the power of hypothesis-free analyses in the context of rare, yet severe ADRs with strong genetic etiology. *SLCO1B1* encodes the organic anion-transporting polypeptide OATP1B1, which mediates the hepatic uptake of most statins and their acids. Candidate gene studies reported in the last few years have shown the involvement of genetic markers in this gene with various drug response phenotypes associated with both the safety and efficacy of several statins. A similar paper that speaks to the utility of WGAS studies in the context

of severe ADR investigations was published last year (Nelson et al., 2008). It is therefore clear that lessons learned from disease susceptibility WGAS would be applicable to implementations of PGx (McCarthy et al., 2008).

Another example that illustrates the potential usage of PGx testing for future guidance of treatment prescriptions, particularly as second and third in-class medications – which do not bear a genetically-associated safety issues – launch into the market, is clopidogrel. This pro-drug, used as an antiplatelet agent as one of the most frequently prescribed drugs in the world, exhibits substantial inter-individual variability in pharmacodynamic response linked to increased risk of cardiovascular events. Data published in recent months indicate that *CYP2C19* genetic variants are involved in pharmacokinetic exposure (Brandt et al., 2007; Kim et al., 2008), pharmacodynamic response (Hulot et al., 2006; Frere et al., 2008), and clinical outcomes (Simon et al., 2008; Trenk et al., 2008). However, as each study adopted differential study designs, target populations, end-point definitions, and genotypic models, the practical utility of the unequivocal role that *CYP2C19* polymorphism plays in the clinical management of patients and the precise treatment guidelines that should be adopted by healthcare providers and insurers awaits further investigation.

The pharmacoeconomics of PGx studies has gradually become a mainstay of PGx diagnostics evaluation, and several groups across various therapeutic areas, drug classes, and geographic locations have employed simulation and trialed datasets to provide recommendations for or against implementation of particular PGx tests in routine healthcare (Huang and Ratain, 2009; Matchar et al., 2007; McClain et al., 2008; Smits et al., 2007). A review systematically surveying pharmacoeconomic analyses of genetic screening strategies, including recommendations for good practice, regardless of country and analysis origin was published by Vegter et al. (2008).

Overall, the above examples illustrate the successful implementation of WGAS as well as large pathway candidate gene studies in the context of PGx analysis. In addition, it is clear that large, well-powered, and appropriately designed PGx replication studies and prospective genotype-guided versus standard-of-care studies are becoming important milestones in the adoption of treatment guidelines and future regulation (as is the case for warfarin [Teagarden and Mayor, 2008]). These case studies also illustrate the growing need for refinement of the paradigm of the development of diagnostics that accompany the prescription of new drugs. Novel models promoted by the FDA advocate co-development of genetic diagnostics in tandem with the development of the drugs for which they will be specifically indicated (Food and Drug Administration, 2005; Hinman et al., 2009; Simon, 2008). The rapidly growing market for direct-to-consumer (DTC) genetic tests highlights the changing environment and the need for adjusting the regulatory frameworks to support enhanced oversight on indicated usage while incentivizing research institutions and industry to utilize PGx diagnostics for improved and safer treatment prescription (Hogarth et al., 2008). Thus an inherent requirement to demonstrate clinical validity and utility in the target population of interest is advocated to become an integral part of a diagnostics' submission package. As a direct result, clear guidance on safe and effective usage would be issued immediately upon marketing, based on objective measurements derived from appropriately powered clinical trials.

2009 UPDATE REFERENCES

Brandt, J.T., Close, S.L., Iturria, S.J., Payne, C.D., Farid, N.A., Ernest, C.S., 2nd Lachno, D.R., Salazar, D. and Winters, K.J. (2007). Common polymorphisms of CYP2C19 and CYP2C9 affect the pharmacokinetic and pharmacodynamic response to clopidogrel but not prasugrel. *J Thromb Haemost* 5(12), 2429–2436.

Food and Drug Administration. (2005). Drug-Diagnostic co-development concept paper (draft – not for implementation).

Frere, C., Cuisset, T., Morange, P.E., Quilici, J., Camoin-Jau, L., Saut, N., Faille, D., Lambert, M., Juhan-Vague, I., Bonnet, J.L. et al. (2008). Effect of cytochrome p450 polymorphisms on platelet reactivity after treatment with clopidogrel in acute coronary syndrome. *Am J Cardiol* 101(8), 1088–1093.

Hinman, L., Spear, B., Tsuchihashi, Z., Kelly, J., Bross, P., Goodsaid, F. and Kalush, F. (2009). Drug-diagnostic codevelopment strategies: FDA and industry dialog at the 4th FDA/DIA/PhRMA/PWG/BIO Pharmacogenomics Workshop. *Pharmacogenomics* 10(1), 127–136.

Hogarth, S., Javitt, G. and Melzer, D. (2008). The current landscape for direct-to-consumer genetic testing: Legal, ethical, and policy issues. *Annu Rev Genomics Hum Genet* 9, 161–182.

Huang, R.S. and Ratain, M.J. (2009). Pharmacogenetics and pharmacogenomics of anticancer agents. *CA Cancer J Clin* 59(1), 42–55.

Hulot, J.S., Bura, A., Villard, E., Azizi, M., Remones, V., Goyenvalle, C., Aiach, M., Lechat, P. and Gaussem, P. (2006). Cytochrome P450 2C19 loss-of-function polymorphism is a major determinant of clopidogrel responsiveness in healthy subjects. *Blood* 108(7), 2244–2247.

Link, E., Parish, S., Armitage, J., Bowman, L., Heath, S., Matsuda, F., Gut, I., Lathrop, M. and Collins, R. (2008). *SLCO1B1* variants and statin-induced myopathy – A genome-wide study. *N Engl J Med* 359(8), 789–799.

Matchar, D.B., Thakur, M.E., Grossman, I., McCrory, D.C., Orlando, L.A., Steffens, D.C., Goldstein, D.B., Cline, K.E. and Gray, R.N. (2007). Testing for cytochrome P450 polymorphisms in adults with non-psychotic depression treated with selective serotonin reuptake inhibitors (SSRIs). *Evid Rep Technol Assess* (Full Rep) 146, 1–77.

McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P. and Hirschhorn, J.N. (2008). Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nat Rev Genet* 9(5), 356–369.

McClain, M.R., Palomaki, G.E., Piper, M and Haddow, J.E. (2008). A rapid-ACCE review of *CYP2C9* and *VKORC1* alleles testing to inform warfarin dosing in adults at elevated risk for thrombotic events to avoid serious bleeding. *Genet Med* 10(2), 89–98.

Nelson, M.R., Bryc, K., King, K.S., Indap, A., Boyko, A.R., Novembre, J., Briley, L.P., Maruyama, Y., Waterworth, D.M., Waeber, G. et al. Genome-wide approaches to identify pharmacogenetic contributions to adverse drug reactions. *Pharmacogenomics J* 2008; [Epub ahead of print].

Simon, R. (2008). The use of genomics in clinical trial design. *Clin Cancer Res* 14(19), 5984–5993.

Smits, K.M., Smits, L.J., Schouten, J.S., Peeters, F.P. and Prins, M.H. (2007). Does pretreatment testing for serotonin transporter polymorphisms lead to earlier effects of drug treatment in patients with major depression? A decision-analytic model. *Clin Ther* 29, 691–702.

Teagarden, J.R. and Mayor, T. (2008). Medco-Mayo warfarin study. American Health Information Community, Personalized healthcare workshop, *web-conference*.

Trenk, D., Hochholzer, W., Fromm, M.F., Chialda, L.E., Pahl, A., Valina, C.M., Stratz, C., Schmiebusch, P., Bestehorn, H.P., Buttner, H.J. et al. (2008). Cytochrome P450 2C19 681G>A polymorphism and high on-clopidogrel platelet reactivity associated with adverse 1-year clinical outcome of elective percutaneous coronary intervention with drug-eluting or bare-metal stents. *J Am Coll Cardiol* 51(20), 1925–1934.

Vegter, S., Boersma, C., Rozenbaum, M., Wilffert, B., Navis, G. and Postma, M.J. (2008). Pharmacoeconomic evaluations of pharmacogenetic and genomic screening programmes: A systematic review on content and adherence to guidelines. *Pharmacoeconomics* 26(7), 569–587.

REFERENCES

Aquilante, C.L., Langaee, T.Y., Lopez, L.M., Yarandi, H.N., Tromberg, J.S., Mohuczy, D., Gaston, K.L., Waddell, C.D., Chirico, M.J. and Johnson, J.A. (2006). Influence of coagulation factor, vitamin K epoxide reductase complex subunit 1, and cytochrome P450 2C9 gene polymorphisms on warfarin dose requirements. *Clin Pharmacol Ther* 79, 291–302.

Arnett, D.K., Claas, S.A. and Glasser, S.P. (2006). Pharmacogenetics of antihypertensive treatment. *Vascul Pharmacol* 44, 107–118.

CDC (2006). Evaluation of genetic testing. ACCE: A CDC-Sponsored Project carried out by the Foundation of Blood Research. http://www.cdc.gov/genomics/gtesting/ACCEhtm.

Chapman, J.M., Cooper, J.D., Todd, J.A. and Clayton, D.G. (2003). Detecting disease associations due to linkage disequilibrium using haplotype tags a class of tests and the determinants of statistical power. *Hum Hered* 56, 18–31.

Conrad, D.F., Andrews, T.D., Carter, N.P., Hurles, M.E. and Pritchard, J.K. (2006). A high-resolution survey of deletion polymorphism in the human genome. *Nat Genet* 38, 75–81.

de Bakker, P.I., Yelensky, R., Pe'er, I., Gabriel, S.B., Daly, M.J. and Altshuler, D. (2005). Efficiency and power in genetic association studies. *Nat Genet* 37, 1217–1223. http://www.broad.mit.edu/mpg/tagger/server.html.

de Bakker, P.I., Burtt, N.P., Graham, R.R., Guiducci, C., Yelensky, R., Drake, J.A., Bersaglieri, T., Penney, K.L., Butler, J., Young, S. et al. (2006). Transferability of tag SNPs in genetic association studies in multiple populations. *Nat Genet* 38, 1298–1303.

DiMasi, J.A., Hansen, R.W. and Grabowski, H.G. (2003). The price of innovation new estimates of drug development costs. *J Health Econ* 22, 151–185.

Endrenyi, L., Inaba, T. and Kalow, W. (1976). Genetic study of amobarbital elimination based on its kinetics in twins. *Clin Pharmacol Ther* 20, 701–714.

FDA (2004a). Challenge and opportunity on the critical path to new medical products. http://www.fda.gov/oc/initiatives/criticalpath/whitepaperhtml.

FDA (2004b). Guidance for Industry: Pharmacogenomic Data Submissions http://www.fda.gov/cber/gdlns/pharmdtasubhtm.

Flockhart, D.A. (2007) Drug interactions. http://medicine.iupui.edu/flockhart/.

Frueh, F.W. (2006). Impact of microarray data quality on genomic data submissions to the FDA. *Nat Biotechnol* 24, 1105–1107.

Frueh, F.W., Goodsaid, F., Rudman, A., Huang, S.M. and Lesko, L.J. (2005). The need for education in pharmacogenomics: A regulatory perspective. *Pharmacogenomics J* 5, 218–220.

Gardiner, S.J. and Begg, E.J. (2006). Pharmacogenetics, drug-metabolizing enzymes, and clinical practice. *Pharmacol Rev* 58, 521–590.

Gaszner, P. and Makkos, Z. (2004). Clozapine maintenance therapy in schizophrenia. *Prog Neuropsychopharmacol Biol Psychiatry* 28, 465–469.

Goldstein, D.B., Tate, S.K. and Sisodiya, S.M. (2003). Pharmacogenetics goes genomic. *Nat Rev Genet* 4, 937–947.

Grossman, I. (2007). Routine pharmacogenetic testing in clinical practice – dream or reality? *Pharmacogenomics* 8, 1449–1459.

Grossman, I., Avidanb, N., Singer, C., Goldstaub, D., Hayardeny, L., Eyal, E., Ben-Asher, E., Paperna, T., Pe'er, I., Lancet, D. et al. (2007). Pharmacogenetics of glatiramer acetate therapy for multiple sclerosis reveals drug-response markers. *Pharmacogenet Genomics* 17(8), 657–666.

Grossman, I., Sullivan, P.F., Walley, N., Liu, Y., Dawson, J.R., Gumbs, C., Gaedigk, A., Leeder, J.S., McEvoy, J.P., Weale, M.E., et al. (2008). Genetics determinants of vairable metabolism have little impact on the clinical use of leading antipsychotics in the CATIE study. Genet Med 10(10), 720–729.

Grossman, I., Liu, Y., Walley, N., et al. (Oct 2006). Pharmacogenetic analysis of antipsychotics: Comprehensive analysis of pharmacokinetic variants. Presented at: 56th Annual Meeting of the American Society of Human Genetics (ASHG). New-Orleans, LA, USA.

Heath, A.C. and Martin, N.G. (1994). Genetic influences on alcohol consumption patterns and problem drinking results from the Australian NH&MRC twin panel follow-up survey. *Ann NY Acad Sci* 708, 72–85.

Hillman, M.A., Wilke, R.A., Yale, S.H., Vidaillet, H.J., Caldwell, M.D., Glurich, I., Berg, R.L., Schmelzer, J. and Burmester, J.K. (2005). A prospective, randomized pilot trial of model-based warfarin dose initiation using CYP2C9 genotype and clinical data. *Clin Med Res* 3, 137–145.

Hughes, D.A., Vilar, F.J., Ward, C.C., Alfirevic, A., Park, B.K. and Pirmohamed, M. (2004). Cost-effectiveness analysis of HLA

B*5701 genotyping in preventing abacavir hypersensitivity. *Pharmacogenetics* 14(6), 335–342.

Hughes, S., Hughes, A., Brothers, C., Spreen, W. and Thorborn, D. (2007). PREDICT-1 (CNA106030) the first powered, prospective trial of pharmacogenetic screening to reduce drug adverse events. *Pharm Stat*.

Ioannidis, J.P., Ntzani, E.E., Trikalinos, T.A. and Contopoulos-Ioannidis, D.G. (2001). Replication validity of genetic association studies. *Nat Genet* 29, 306–309.

Jones, T.S., Yang, W., Evans, W.E. and Relling, M.V. (2007). Using HapMap tools in pharmacogenomic discovery: The thiopurine methyltransferase polymorphism. *Clin Pharmacol Ther* 81, 729–734.

Kelly, P.J., Stallard, N. and Whittaker, J.C. (2005). Statistical design and analysis of pharmacogenetic trials. *Stat Med* 24, 1495–1508.

Kendler, K.S., Martin, N.G., Heath, A.C., Handelsman, D. and Eaves, L.J. (1988). A twin study of the psychiatric side effects of oral contraceptives. *J Nerv Ment Dis* 176, 153–160.

Kim, K.A., Park, P.W., Hong, S.J. and Park, J.Y. (2008). The effect of CYP2C19 polymorphism on the pharmacokinetics and pharmacodynamics of clopidogrel: a possible mechanism for clopidogrel resistance. Clin Pharmacol Ther 84, 236-242.

Kindmark, A., Jawaid, A., Harborn, C.G., Barratt, B.J., Bengtsson, O.F., Andersson, T.B., Carlsson, S., Cederbrant, K.E. et al. (2007). Genome-wide pharmacogenetic investigation of a hepatic adverse event without clinical signs of immunopathology suggests an anderlying immune pathogenesis. *Pharmacogenomics J*, 1–10.

Kirchheiner, J., Henckel, H.B., Meineke, I., Roots, I. and Brockmoller, J. (2004). Impact of the CYP2D6 ultrarapid metabolizer genotype on mirtazapine pharmacokinetics and adverse events in healthy volunteers. *J Clin Psychopharmacol* 24, 647–652.

Krejsa, C., Rogge, M. and Sadee, W. (2006). Protein therapeutics new applications for pharmacogenetics. *Nat Rev Drug Discov* 5, 507–521.

Kurtzke, J.F. (1983). Rating neurologic impairment in multiple sclerosis an expanded disability status scale (EDSS). *Neurology* 33, 1444–1452.

Lesko, L.J. and Woodcock, J. (2002). Pharmacogenomic-guided drug development regulatory perspective. *Pharmacogenomics J* 2, 20–24.

Lesko, L.J. and Woodcock, J. (2004). Translation of pharmacogenomics and pharmacogenetics a regulatory perspective. *Nat Rev Drug Discov* 3, 763–769.

McLeod, H.L., Tan, B., Malyapa, R., Abbey, E., Picus, J., Myerson, R., Zehnbauer, B., Mutch, M., Dietz, D. and Fleshman, J. (2005). Genotype-guided neoadjuvant therapy for rectal cancer. *Proc Am Soc Clin Oncol* 23, 197.

Meadows, M. (2005). FDA approves heart drug for black patients. *FDA Consum* 39, 8–9.

Murphy, G.M., Jr., Kremer, C., Rodrigues, H.E. and Schatzberg, A.F. (2003). Pharmacogenetics of antidepressant medication intolerance. *Am J Psychiatry* 160, 1830–1835.

NCI-NHGRI Working Group on Replication in Association Studies (2007). Replicating genotype-phenotype associations. *Nature* 447, 655–660.

Nebert, D.W. and Dieter, M.Z. (2000). The evolution of drug metabolism. *Pharmacology* 61, 124–135.

Need, A.C., Motulsky, A.G. and Goldstein, D.B. (2005). Priorities and standards in pharmacogenetic research. *Nat Genet* 37, 671–681.

Nevins, J.R. and Potti, A. (2007). Mining gene expression profiles expression signatures as cancer phenotypes. *Nat Rev* 8, 601–609.

O'Reilly, R.L., Bogue, L. and Singh, S.M. (1994). Pharmacogenetic response to antidepressants in a multicase family with affective disorder. *Biol Psychiatry* 36, 467–471.

Pe'er, I., de Bakker, P.I., Maller, J., Yelensky, R., Altshuler, D. and Daly, M.J. (2006). Evaluating and improving power in whole-genome association studies using fixed marker sets. *Nat Genet* 38, 663–667.

Perry, P. and Hradek, K. (2004). Antipsychotic maintenance therapy. In Clinical Psychopharmacology Seminar.

PharmGKB Pathways. http://www.pharmgkborg/search/pathway/pathwayjsp.

Rauch, A., Nolan, D., Martin, A., McKinnon, E., Almeida, C. and Mallal, S. (2006). Prospective genetic screening decreases the incidence of abacavir hypersensitivity reactions in the Western Australian HIV cohort study. *Clin Infect Dis* 43, 99–102.

Rettie, A.E., Korzekwa, K.R., Kunze, K.L., Lawrence, R.F., Eddy, A.C., Aoyama, T., Gelboin, H.V., Gonzalez, F.J. and Trager, W.F. (1992). Hydroxylation of warfarin by human cDNA-expressed cytochrome P-450: a role for P-4502C9 in the etiology of (S)-warfarin-drug interactions. *Chem Res Toxicol* 5, 54–59.

Rieder, M.J., Reiner, A.P., Gage, B.F., Nickerson, D.A., Eby, C.S., McLeod, H.L., Blough, D.K., Thummel, K.E., Veenstra, D.L. and Rettie, A.E. (2005). Effect of VKORC1 haplotypes on transcriptional regulation and warfarin dose. *N Engl J Med* 352, 2285–2293.

Risner, M.E., Saunders, A.M., Altman, J.F., Ormandy, G.C., Craft, S., Foley, I.M., Zvartau-Hind, M.E., Hosford, D.A. and Roses, A.D. (2006). Efficacy of rosiglitazone in a genetically defined population with mild-to-moderate Alzheimer's disease. *Pharmacogenomics J* 6, 246–254.

Roses, A.D. (2004). Pharmacogenetics and drug development the path to safer and more effective drugs. *Nat Rev Genet* 5, 645–656.

Roses, A.D., Saunders, A.M., Huang, Y., Strum, J., Weisgraber, K.H. and Mahley, R.W. (Feb 2007). Complex disease-associated pharmacogenetics: Drug efficacy, drug safety, and confirmation of a pathogenetic hypothesis (Alzheimer's disease). *Pharmacogenomics J* 7(1): 10–28.

Singer, C., Grossman, I., Avidan, N., Beckmann, J.S. and Pe'er, I. (2005). Trick or treat the effect of placebo on the power of pharmacogenetic association studies. *Hum Genomics* 2, 28–38.

Shah, R.R. (2006). Can pharmacogenetics help rescue drugs withdrawn from the market? *Pharmacogenomics* 7(6), 889–908.

Skol, A.D., Scott, L.J., Abecasis, G.R. and Boehnke, M. (2006). Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 38, 209–213.

Spraggs, C.F., Pillai, S.G., Dow, D., Douglas, C., McCarthy, L., Manasco, P.K., Stubbins, M. and Roses, A.D. (2005). Pharmacogenetics and obesity common gene variants influence weight loss response of the norepinephrine/dopamine transporter inhibitor GW320659 in obese subjects. *Pharmacogenet Genomics* 15, 883–889.

Swan, G.E., Benowitz, N.L., Jacob, P., 3rd, Lessov, C.N., Tyndale, R.F., Wilhelmsen, K., Krasnow, R.E., McElroy, M.R., Moore, S.E. and Wambach, M. (2004). Pharmacogenetics of nicotine metabolism in twins methods and procedures. *Twin Res* 7, 435–448.

Tate, S.K., Depondt, C., Sisodiya, S.M., Cavalleri, G.L., Schorge, S., Soranzo, N., Thom, M., Sen, A., Shorvon, S.D., Sander, J.W. et al. (2005). Genetic predictors of the maximum doses patients receive during clinical use of the anti-epileptic drugs carbamazepine and phenytoin. *Proc Natl Acad Sci USA* 102, 5507–5512.

Telenti, A. and Goldstein, D.B. (2006). Genomics meets HIV-1. *Nat Rev Microbiol* 4, 865–873.

The International HapMap Project (2003). *Nature* 426, 789–796.

Thakur, M., Grossman, I., McCrory, D.C., Orlando, L.A., Steffens, D.C., Cline, K.E., Gray, R.N., Farmer, J., DeJesus, G., O'Brien, C.,

Samsa, G., Goldstein, D.B., Matchar, D.B. (2007). Review of evidence for genetic testing for CYP450 polymorphisms in management of patients with non psychotic depression with selective serotonin reuptake inhibitors (SSRIs). *Gen Med* 9, 826–835.

van den Akker-van Marle, M.E., Gurwitz, D., Detmar, S.B., Enzing, C.M., Hopkins, M.M., Gutierrez de Mesa, E. and Ibarreta, D. (2006). Cost-effectiveness of pharmacogenomics in clinical practice a case study of thiopurine methyltransferase genotyping in acute lymphoblastic leukemia in Europe. *Pharmacogenomics* 7, 783–792.

Warren, L.L., Hughes, A.R., Lai, E.H., Zaykin, D.V., Haneline, S.A., Bansal, A.T., Wooster, A.W., Spreen, W.R., Hernandez, J.E., Scott, T.R., Roses, A.D., Mosteller, M. (2007). Use of pairwise marker combination and recursive partitioning in a pharmacogenetic genome-wide scan. *Pharmacogenomics J* 7, 180–189.

Weiss, S.T., Litonjua, A.A., Lange, C., Lazarus, R., Liggett, S.B., Bleecker, E.R. and Tantisira, K.G. (2006). Overview of the pharmacogenetics of asthma treatment. *Pharmacogenomics J* 6, 311–326.

Wilke, R.A., Reif, D.M. and Moore, J.H. (2005). Combinatorial pharmacogenetics. *Nat Rev Drug Discov* 4, 911–918.

Wilkinson, G.R. (2005). Drug metabolism and variability among patients in drug response. *N Engl J Med* 352, 2211–2221.

Zika, E., Gurwitz, D. and Ibearreta, D. (2006). Pharmacogenetics and Pharmacogenomics: State-of-the-art and potential socio-economic impact in the EU. European Communities.

Zucman, D., Truchis, P., Majerholc, C., Stegman, S. and Caillat-Zucman, S. (2007). Prospective screening for human leukocyte antigen-B*5701 avoids abacavir hypersensitivity reaction in the ethnically mixed French HIV population. *J Acquir Immune Defic Syndr* 45, 1–3.

RECOMMENDED RESOURCES

Journals

Ioannidis, J.P., Ntzani, E.E., Trikalinos, T.A., and Contopoulos-Ioannidis, D.G. (2001). Replication validity of genetic association studies. *Nat Genet* 29, 306–309.

NCI-NHGRI Working Group on Replication in Association Studies (2007). Replicating genotype-phenotype associations, *Nature* 447, 655–660.

Roses, A.D. (2004). Pharmacogenetics and drug development: The path to safer and more effective drugs. *Nat Rev Genet* 5, 645–656.

Websites

http://www.hapmap.org/International Hap Map Project web site

http://www.cdc.gov/genomics/gtesting/ACCE.htm ACCE: A CDC-sponsored project for evaluation of genetic testing

http://www.fda.gov/cber/gdlns/pharmdtasub.htm US Food and Drug Administration site for pharmacogenetic data submissions

http://www.pharmgkb.org/search/pathway/pathway.jsp PharmGKB Pathways

CHAPTER

Clinical Implementation of Translational Genomics

Wendy K. Chung

INTRODUCTION

In the 55 years since the structure of DNA was described by Watson and Crick, there have been remarkable advances in our understanding of the genetic basis for human variation and disease. We have now defined the genetic basis for over 2130 monogenic human disorders, and clinical genetic tests are available for over 1410 disorders (http://www.genetests.org/). Despite the remarkable advances in our scientific understanding of genetics and genomics, the impact of such information on medical care for common conditions has to this point been modest. With the complete sequence of the human genome and many other model organisms as well as detailed characterization of human haplotype structure, we have now begun elucidating the complex genetic basis for common diseases such as macular degeneration, diabetes, obesity, inflammatory bowel disease, and breast cancer using genome-wide association studies. In some cases the same allele will increase risk for one disease and protect against another. The *ApoE4* allele is associated with a protective effect for age-related macular degeneration yet susceptibility to Alzheimer disease and hyperlipidemia. As additional scientific discoveries unfold, we will be challenged to clinically integrate this new information into routine patient care to improve health and quality of life in a cost-effective, socially acceptable manner. Early adopters of this new genetic information will provide invaluable experience to guide future implementation strategies. In this chapter, we will review the current and projected future use of genetics and genomics in clinical medicine and define the steps to successful integration of genomic medicine to improve the quality of health care. A schematic of the elements necessary to develop clinical genomic medicine is provided in Figure 16.1.

GENETIC STRATIFICATION WILL ALLOW MEDICAL CARE TO BE INDIVIDUALIZED AFTER A DIAGNOSIS IS MADE

How will germline genomic variation be clinically utilized? There are emerging data that the genetic stratification of phenotypically similar diseases has important therapeutic implications. Genetic characterization following an initial diagnosis may clarify prognosis and response to therapy. Three examples to illustrate this point.

1. *Genetic testing may provide information about future risk.* Women with breast cancer due to mutations in *Breast Cancer 1/Breast Cancer 2* (*BRCA1/BRCA2*) have an increased risk for a second primary cancer, usually breast or ovarian cancer, for which increased surveillance, chemoprevention, or prophylactic surgery are recommended (Verhoog et al., 1998). However, women with breast cancer without *BRCA1/BRCA2* mutations are not at increased risk for ovarian cancer (Kauff et al.,

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

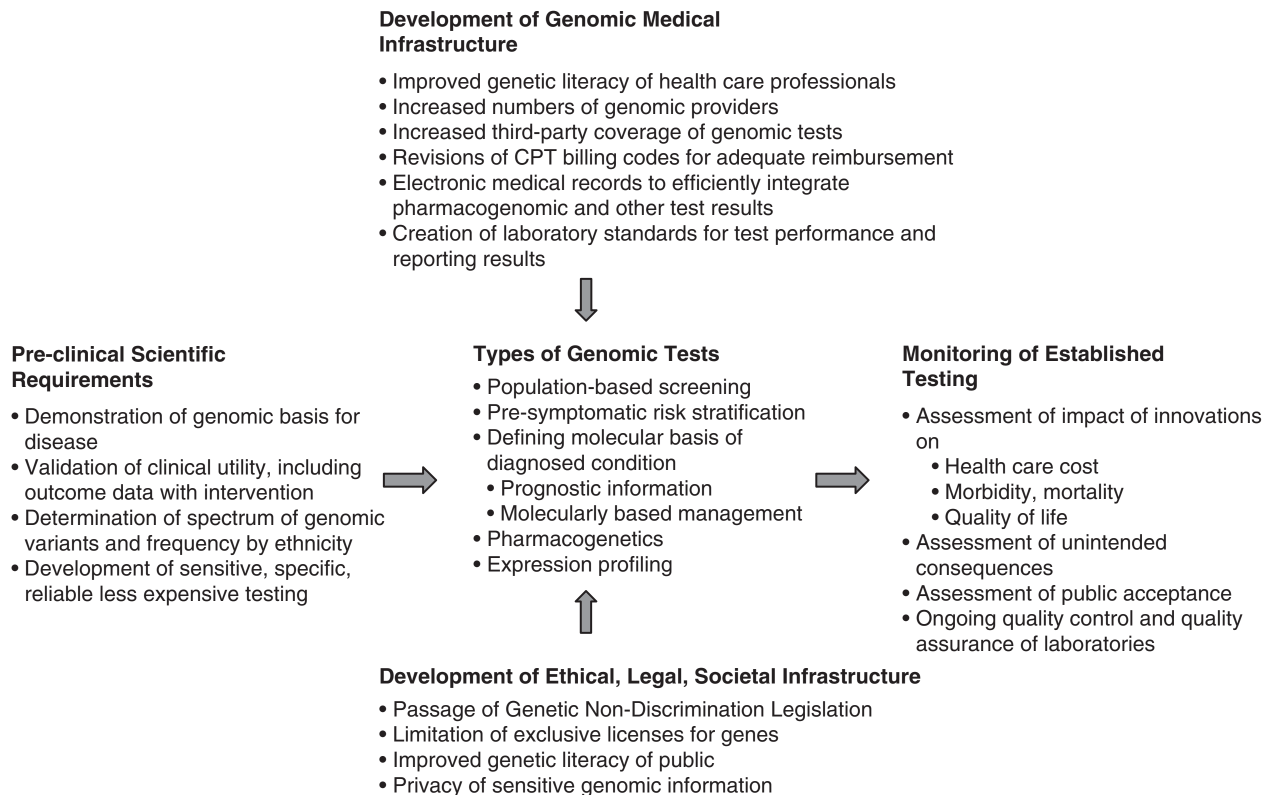


Figure 16.1 Development of clinical genomic medicine. The requirements to develop a health care system efficiently and effectively integrating genomic information are shown and include pre-clinical scientific data, improvements in medical infrastructure, changes in laws, policy, and public opinion, and monitoring of testing and its impact once established.

2005). Determining which women with breast cancer are at increased risk of ovarian cancer and should have prophylactic oophorectomies to reduce the risk of future ovarian cancer can be clarified with *BRCA1* and *BRCA2* testing.

2. *Genetic testing may provide information to optimize therapy*. Long QT syndrome (L\QTS) is an inherited predisposition to cardiac arrhythmias characterized by a prolonged QT interval on an electrocardiogram that can result in syncope and sudden cardiac death. There are currently nine different genetically identified causes of LQTS, all affecting cardiac ion channel conductance that are clinically difficult to distinguish. The three most common forms of LQTSs (LQT1, LQT2, and LQT3) have specific triggers for arrhythmias they can be avoided such as exercise-induced tachycardia in LQT1 and auditory stimuli during sleep in LQT2 (Schwartz et al., 2001). Specific molecularly based pharmacological intervention is also now available. For example, patients with LQT3, due to mutations in the cardiac sodium channel gene *SCN5A* that inappropriately activate and open the sodium channel, respond to sodium channel blockers such as flecainide (Moss et al., 2005). In contrast, beta-blockers are routinely prescribed for LQTS1 and LQTS2. Therefore, identifying the molecular subtype of LQTS allows for selection of the most appropriate medication and avoidance of specific triggers.

3. *Genetic testing may provide prognostic information to guide management*. An autosomal dominantly inherited form of diabetes called maturity onset diabetes of the young (MODY) is genetically heterogeneous and caused by mutations in six different genes. Prognosis, including diabetic complications, for patients with mutations in *Glucokinase* (MODY2) is much better than other forms of MODY. Molecular diagnosis can now provide reassurance that tight glycemic control is unnecessary to prevent diabetic complications in MODY2 patients who will remain only mildly hyperglycemic (Codner et al., 2006).

POPULATION-BASED GERMLINE GENOMIC SCREENING

Rather than waiting until a diagnosis is made, genetic and genomic variation can be used to prevent disease. Population-based genomic screening could provide the opportunity to improve health outcomes by disease prevention and increased surveillance to facilitate early diagnosis. It would allow population

stratification to identify individuals at increased risk who are most likely to benefit from preventive medications or interventions for which population-based therapy would not be appropriate based on cost or side effects of treatment. Data to support an individual's susceptibility to disease could also provide the necessary motivation to increase compliance with recommended health behaviors such as exercise and weight control for individuals at increased risk for diabetes and smoking cessation for lung cancer. Screening individuals at risk for adverse outcomes to specific environmental exposures would allow susceptible individuals to modify their work or home environment to minimize exposure. Screening individuals for adverse pharmacological reactions would provide a rational basis for drug selection and avoid harmful side effects. Such a strategy would ultimately increase the total number of drugs available for clinical use and decrease the cost of drug development by identifying and eliminating drug exposure to the small number of patients who would have adversely responded to the medication. Although there is great promise for population-based screening, it is important to ensure that the natural history and clinical utility (see section on clinical utility) have been adequately defined before introduction into clinical care. As an example, consider the case of hereditary hemachromatosis. Two common mutations with a single gene (*HFE*) account for the majority of autosomal recessively inherited hereditary hemachromatosis in Europeans (Jazwinska et al., 1996). Furthermore, because the complications of hereditary hemachromatosis are completely preventable by reducing iron stores through phlebotomy (which is inexpensive), presymptomatic screening for hereditary hemachromatosis could be clinically useful. With a frequency of 1 in 400 Caucasians carrying one or both of two common mutations in *HFE*, large-scale population-based screening is technically feasible. As many experts considered introducing hereditary hemachromatosis screening on a population-wide basis, it became increasingly apparent that the natural history of mutation carriers was incompletely understood and that the disease penetrance is less than 1% (Beutler et al., 2002). Although population-based screening for hereditary hemachromatosis still has clinical utility, the implications of a positive genetic test are less significant than had once been anticipated, dramatically altering the cost-benefit ratio. Criteria to use in considering further tests for population-based screening include the disease prevalence, disease penetrance, availability of sensitive, specific, and cost-effective screening tests which may depend upon the number of genes involved and the mutation spectrum, availability of acceptable means of prevention or early detection, and the infrastructure available to perform the screening and follow up the results with confirmatory testing.

NEWBORN SCREENING

Newborn screening is the largest population-based public health screening program currently in practice in the United States and throughout the world. Recent technological advances including tandem mass spectrometry and molecular genetic diagnostics have allowed many states in the United States to cost-effectively increase the number of disorders for which newborns are screened well beyond the classically screened diseases such as phenylketonuria, hypothyroidism, and galactosemia. Over 30 disorders are now screened in many US states. Most of recent expansion in newborn screening is for inborn errors of metabolism detected by metabolomic characterization of a dried spot of blood from a newborn heel stick. Additional molecular genetic diagnostic platforms have recently been added by many states as a second-tier screen to decrease the number of false-positives from metabolomic characterization. Two-tiered screening provides a complementary method to confirm inherited diseases with a small number of well-characterized mutations such as in medium-chain acyl dehydrogenase deficiency (MCAD), cystic fibrosis, and congenital adrenal hyperplasia. The use of tandem mass spectrometry and molecular genetic testing have now opened the possibility to screen for a vast number of inherited conditions. At the request of the General Accounting Office and Health Resources and Services Administration, the American College of Medical Genetics recently re-examined the criteria for determining which disorders to include in newborn screening (http://mchb.hrsa.gov/screening/). Unlike universally accepted screening for disorders such as phenylketonuria for which early detection provides the opportunity for complete prevention of permanent disabilities, some of the newly added disorders allow for early diagnosis and treatment initiation but do not provide a cure or effectively prevent all the long-term disabilities. There is now an ongoing, evolving dialogue about what disorders should be included in newborn screening and whether the criteria proposed by Wilson and Jungner (1968) must be satisfied. Cystic fibrosis is currently screened in newborns in 28 states. Krabbe disease, treatable only by bone marrow transplantation, is now performed in New York. Many other common genetic conditions such a spinal muscular atrophy, Fragile X, and muscular dystrophy have been considered even though only supportive care is currently available for these conditions. What treatment is efficacious enough to warrant population-based screening is debated. Clearly an important factor limiting screening for many conditions is availability of technology to allow for inexpensive, sensitive, and specific screening. Multiplexed DNA microarrays to screen for inherited disease susceptibility and infectious agents in newborns have been suggested (Green and Pass, 2005) and are being piloted. In general, pediatricians have endorsed the expansion of screening in high-risk infants (Acharya et al., 2005) since it leads to reduced time to diagnosis and initiation of treatment, decreased expense and invasiveness of diagnostic testing, and allows for informed reproductive planning for the family. However, as the number of disorders screened increases, it is equally necessary to improve test parameters to increase specificity and reduce the number of false-positive tests to spare the parents of unaffected children the anxiety of a potential diagnosis that will never be made (Waisbren et al., 2003). Not all genetic testing is appropriate for newborns. Genetic testing should be defined and limited to conditions for which there will be medical implications for minors, and no genetic testing should be performed for adult onset conditions for which there is no intervention during childhood.

SOMATIC GENOMIC VARIATION

In addition to testing for heritable germline genetic characteristic, we will need to characterize somatic variation over the lifetime of an individual, especially in characterizing oncological specimens. Cancer treatment is likely to be rapidly redefined with genomic characterization of tumors using a combination of comparative genomic hybridization, quantification of gene amplifications/deletions, identification of acquired genetic mutations, and gene-expression profiling. These methods are already clinically utilized and used on a small scale. Molecularly targeted medications with fewer side effects will be more frequently developed as we rationally develop drugs based upon specific molecular targets. Examples already include the tyrosine kinase inhibitor imatinib mesylate (Gleevec) for chronic myelogenous leukemia and gastrointestinal stromal tumors that over express the tyrosine kinases, trastuzumab (Herceptin) for *HER2* over expressing breast cancer, and gefitinib for activating mutations in *Epidermal Growth Factor Receptor* for non-small cell lung cancer (Lynch et al., 2004). These medications offer the possibility of increased efficacy and rely critically on accurate molecular tumor characterization to identify the subset of patients who will likely respond to these expensive therapies. Genomic Health Inc. offers another form of genomic tumor characterization, a limited expression profiling array (Oncotype Dx) of 21 genes for estrogen receptor positive, node negative, stage one or two breast cancer. The expression profile quantifies the likelihood of breast cancer recurrence in women with newly diagnosed, early stage breast cancer and assesses the benefit of chemotherapy with a numerical scoring system ranging from 1 to 100. The ability of this assay to analyze gene expression on paraffin embedded tissue offers an advantage over the requirement of freshly frozen tissue since new procedures for tissue handling and storage need not be developed to utilize this technology. The ability to simply, numerically quantify and integrate complex genomic information facilitates physician and patient communication and understanding to allow patients and physicians to make more rational decisions about therapy. Although not always covered by third-party payers, some patients are willing to personally bear the cost of testing to provide additional reassurance that chemotherapy will not benefit them and will allow them to avoid chemotherapy associated morbidity without increasing risk of recurrence. Third-party payer coverage will increase as clinical utility is demonstrated, patients and physicians requests for testing increases, and as diagnostic laboratories perform sufficient volumes of tests to negotiate carve out contracts with major health insurers and CMS to cover testing. There are currently multiple oncology clinical trials in progress to determine the precise algorithms by which expression profiles will be weighted and analyzed and to determine the clinical utility for prognosis and treatment efficacy (Buckhaults, 2006). Similar expression profiles from blood samples are currently being used to predict rejection for cardiac transplant recipients and obviate the need for an invasive cardiac biopsy if the expression profile indicates a low probability of rejection (Deng et al., 2006). It is likely that clinically relevant diagnostic tools will soon be available for multiple tumor types to assist with oncological management and multiple solid organ transplants to predict rejection.

NOVEL SOURCES OF GENOMIC VARIATION

In addition to the nuclear germline and somatic mutations resulting from alteration in one or a small number of nucleotides, there are additional sources of genomic variation, the clinical significance of which is not yet fully appreciated. Human genomic architecture and variation are now being characterized. It is likely that previously undetected variation in the copy number of genes will underlie a significant fraction of birth defects, mental retardation, and autism (de Vries et al., 2005; Schoumans et al., 2005; Sebat el al., 2007). It is also possible that copy number variation increases susceptibility to common psychiatric, neurological, and medical conditions (Sebat et al., 2004). Somatic copy number variation has long been associated with cancer and amplification or deletion of specific genes that may provide additional prognostic and therapeutic information (Pinkel and Albertson, 2005). Bacterial artificial chromosome and oligonucleotide arrays are currently being developed and are available on a research basis to provide a molecular karyotype with resolution of 10–100 kb and are recently clinically available with a resolution of 100–500 kb. It is likely that such oligonucleotide arrays will replace standard karyotypes as the first means of analyzing chromosomes for an overall assessment of genomic balance. However, to maximize its clinical utility, it will be necessary to characterize normal copy number variation and demonstrate clinically relevant associations of copy number variations/alterations with disease. As it has become less expensive to genomically characterize gene copy number on many of the same platforms used to genotype single nucleotide polymorphisms (SNPs), tests of association with common diseases will provide the data necessary to judge clinical validity and utility.

In addition to nuclear genetic variation, genetic variation also exists within the mitochondria which can vary from cell to cell and over the course of a lifetime due to mitochondrial heteroplasmy. Mitochips are currently available on a research and clinical basis to sequence the complete mitochondrial genome. As correlations between mitochondrial mutations are associated with diseases associated with degeneration and advancing age, it may become important to include methods to detect and quantify mitochondrial variation in blood as well as other tissues over the course of a patient's lifetime.

LABORATORY STANDARDS TO ENSURE ANALYTIC VALIDITY

As new methodologies and diagnostic tests such as those described above are developed, it will be necessary for diagnostic laboratories to maintain rigorous standards of quality control and quality assurance. New diagnostic assays will be developed in

Clinical Laboratory Improvement Amendment (CLIA) certified laboratories. Some assays will use reagents developed and utilized only by that laboratory ("laboratory developed tests") as has traditionally been done in molecular genetic diagnostic laboratories for rare and ultra-rare disorders. As molecular genetic testing is applied to more common disorders, it is more likely that analyte-specific reagents (ASRs) will be commercially prepared and packaged for wider distribution to a large number of laboratories. It will be necessary for novel diagnostic platforms to be FDA-approved before the platforms and ASRs can be widely distributed to commercial diagnostic laboratories. The rigorous review provided by the FDA for products such as the AmpliChip CYP450 provide assurance of the analytical performance (sensitivity, specificity, and reproducibility) of the diagnostic assay and interpretability of results by clinicians. Increasingly, review and oversight of novel diagnostic tests will be necessary to assure analytical quality; however, regulation by the FDA if not expeditious could impede the efficient transfer of diagnostic methods into clinical practice. Each laboratory will need to rigorously validate the analytical performance of new assays including splitting of samples within and between laboratories to compare results between analytical methods. Standards and guidelines for genetic testing have been developed and are regularly revised by the American College of Medical Genetics (http://www.acmg.net/resources/s-g/s-g-yes-no.asp). Regular proficiency testing including the use of common samples between laboratories will ensure ongoing test quality. It is anticipated that as new ASRs are developed, the level of technical expertise necessary to perform the assays and the cost of running the assays should decrease simultaneously as throughput increases with automation and multiplex assays. It is likely that in the future the majority of molecular genetic testing will move from the many small academic boutique laboratories currently performing most testing to large commercial laboratories and/or hospitals with larger capacity and potentially one day could be performed as point of care testing.

Communicating test results clearly and concisely is as important as the laboratory's analytical performance. Test reports should present sufficient detail about test methodologies to allow for comparison of methods between laboratories and allow the physician to determine the likelihood of a false-negative based upon the mutations included in the assay. As the scientific data and methods are likely to change over time, these reports will require periodic revision, especially with regard to clinical utility. Results should be succinctly reported in a final summary statement including sensitivity and specificity of the test, the possible implications of a positive and negative test result, and reference to or inclusion of necessary supporting data documenting clinical utility. Ideally, Gene Clinics (http://www.geneclinics.org/) or a similarly organized web-based format would provide periodically revised summaries of the genetic tests and their clinical utility. As data from these tests accumulate, de-identified centralized databases defined by gene or test should be maintained with phenotypic information and outcomes to allow for ongoing data collection and analysis to constantly improve test utility and interpretation, especially that of variants of unknown clinical significance.

CLINICAL VALIDATION AND CLINICAL UTILITY

Genomic testing is expected to simplify diagnostic work ups, provide prognostic information, improve and refine clinical management, identify individuals at increased risk, and decrease adverse outcomes. However, before this can be implemented, genomic testing will require clinical validation and demonstration of clinical utility before acceptance into health care. The data for such validation may initially come from studies of patients with the most severe disease manifestations. However, it will be important to analyze population-based samples as well to accurately determine association between genetic susceptibility and disease in an unbiased manner. Studies in multiple ethnic groups will allow characterization of allele frequencies in various populations. It is likely that large clinical trials will provide much of these data necessary for validation since many trials now routinely integrate genetic and genomic information to stratify response to therapy and adverse outcomes by genotype. Integration of genetic data into existing and developing clinical data collection systems such as birth registries, vital statistics, cancer registries, and biorepositories linked to electronic medical records will facilitate testing and validation of preliminary genetic associations and facilitate rapid, independent confirmation of results before clinical introduction.

Clinical utility is the ability in specific clinical circumstances of a genomic test to assist in clinical decision-making and improve health outcomes. Whenever possible, a genetic/genomic test should predict a defined clinical outcome with high positive and negative predictive value. Ideally, prospective, randomized, blinded controlled clinical trials genetically stratified prior to treatment would provide prospective data to validate clinical utility. To effectively evaluate the cost-effectiveness of a genetic test, the population frequency of the at-risk genotype, the age-related penetrance, morbidity and mortality of the disease, and the effects of interacting modifiable risk factors on genotypic expression must all be accurately known. Recommendations for clinical introduction of genetic testing should only be made when the diagnostic methodologies are reliable, patients have access to the clinical services necessary to make informed decisions and interpret the results of genetic testing, and when it becomes clinically apparent how to utilize the results of genetic testing. Once clinical testing is offered, it is also important to maintain ongoing data collection after testing is introduced to define any additional, unintended consequences.

As examples, we can compare the clinical utility of genetic testing for hereditary breast/ovarian cancer conferred by *BRCA1* and *BRCA2* and the risk for venous thromboembolism due to Factor V Leiden. The lifetime risks of breast cancer for carriers with *BRCA1* or *BRCA2* mutations are approximately 65% and 45%, respectively (Antoniou et al., 2003). Mutations carriers can pursue increased surveillance for breast cancer, risk reducing prophylactic mastectomy, or chemoprevention with tamoxifen. Arguments can even be made that with a mutation

frequency of 2.5% in the Ashkenazi population (Hartge et al., 1999), population-based screening for population-specific founder mutations in adult women should be considered. The clinical utility of *BRCA1/BRCA2* testing is derived in large part from the high penetrance of the mutations. On the other hand, thrombophilia susceptibility conferred by Factor V Leiden is the most commonly requested molecular genetic test, although the risk conferred is much more modest and ranges from two- to eightfold (Rosendaal, 1999). It has been proposed that women should be tested for the Factor V Leiden mutation prior to initiation of oral contraceptives due to the 30-fold increased risk of venous thrombosis for mutation carriers on oral contraceptives (Vandenbroucke et al., 1994). However, although the relative risk of thrombosis is significantly increased, the absolute risk is only 28/10,000 person years and the mortality is low in young women. Failure to use oral contraceptives by the numerous Factor V Leiden mutations carriers could have other unforeseen implications including unwanted pregnancies. Thus, no professional consensus has yet emerged about the utility of testing.

COST

A major impediment to integration of genetic/genomic information into health care has been the high cost of molecular genetic testing and the unwillingness of many third-party payers to cover this expense, many citing that it remains experimental. The cost of molecular genetic testing is determined largely by the costs of instrumentation, reagents, personnel, licensing fees, and professional liability insurance. As testing becomes increasingly automated and multiplexed using high-throughput assays on miniaturized scales, the cost of reagents and personnel will be reduced significantly. A goal of the National Human Genome Research Institute is to provide individual genome sequence for $1000 (Collins et al., 2003). Clearly if the technology can drive the cost of testing down to this extent, a major financial barrier to testing will be removed. Licensing fees have significantly increased the cost of testing for a limited number of conditions such as hereditary breast and ovarian cancer (*BRCA1* and *BRCA2*) to $3120 (http://www.myriadtests.com/) and LQTS to $5400 (http://www.pgxhealth.com/genetictests/familion/index.cfm). The genetic/genomic community is opposed to exclusive licenses on genes and is mixed in opinion on gene patenting, but it remains to be seen how much the gene patents and intellectual property regarding clinical utility of specific diagnostic testing will increase the average cost of testing.

In addition to the absolute cost of testing, the extent to which these costs are borne by third-party payers differs widely by laboratory, test, and payer. In the 12 years since *BRCA1/BRCA2* genetic testing became clinically available, the number of individuals tested has increased dramatically, in large part as insurance companies, Medicare, and recently Medicaid began to cover the cost of testing for patients meeting medical eligibility. While many patients might elect to pursue genetic testing for health maintenance and disease prevention, third-party payers are reluctant to expose themselves to this financial burden until the utility is proven. The majority of molecular genetic testing currently performed and paid for by third-party payers surrounds reproductive issues such as carrier screening for cystic fibrosis, recommended by the American College of Obstetrics and Gynecology for consideration in all women contemplating conception. Standard of clinical care guidelines endorsed by major professional medical organization are influential in getting coverage for new tests. Without third-party payer coverage for genomic testing, genomic testing will be utilized only by a small, motivated segment of society who can afford to self-pay.

A significant proportion of the financial benefit to genomic health care is attributable to preventive medicine and the opportunity to prevent chronic disease and cancer. Regional markets in which there are single, dominant insurers or organizations that are self-insured are more likely to initially support genomic health maintenance since they will see the financial benefits within a single organization. Markets with large numbers of competing insurers are reluctant to absorb large financial costs for which they will not personally see a return as consumers change between insurance carriers with regularity. However, as the cost incurred for genomic testing decreases and the utility becomes more apparent and as consumers demand this benefit, it is likely that additional major health insurance providers will add these services to both remain competitive and as a long-term cost-saving measure.

REIMBURSEMENT

Many diagnostic tests, especially those that are highly multiplexed, have been slow to be developed and made clinically available because the testing laboratories cannot charge third-party payers enough to cover their operating expenses. As new technologies for molecular genetic testing are developed, it will be necessary to revise and develop new Current Procedural Terminology (CPT) codes that accurately reflect the complexity of the testing. For many multiplex assays composed of hundreds to thousands of simultaneous assays, it will be necessary to raise the current ceiling for the maximum number of probes, hybridizations, and amplicons to reflect the complexity and cost of the new test methodologies. Once this ceiling is raised, laboratories may find it financially feasible to expand their testing platforms and test menus.

WHO WILL PROVIDE GENOMIC MEDICAL CARE?

Genetic medicine is currently provided in large part by board-certified medical geneticists and genetic counselors (see Box 1.1 in Chapter 1). Given the projected rate of growth and increasing specialization of genetic and genomic medicine and small number of 80 MD and/or PhD graduates per year in medical genetics (Korf et al., 2005), it will be necessary for all health care providers to integrate genomic information to some extent into their provision of health care. Genomics may follow other

diagnostic tests such as pathology, microbiology, and diagnostic imaging with specialized physicians performing and interpreting the tests, while practitioners become largely responsible for determining which patients require testing and how to utilize the test results in heath care. Medical geneticists will continue to play important roles in research, clinical development of the field, education, laboratory medicine, and possibly treatment, in addition to providing direct clinical care for patients with disorders with which they have clinical expertise.

Genetic counselors will continue to play vital roles in patient education and counseling. It will be important to recognize the role of genetic counselors by granting formal licensure and persuading insurers to reimburse for codes for genetic counseling to allow for reasonable reimbursement for genetic counselors' services. Genetic counselors may be increasingly employed by diagnostic laboratories to offer counseling to patients referred by their physicians for testing. The demand for genetic counselors will surpass supply and will necessitate expansion of training programs that currently graduate only 160 new genetic counselors annually. In the future, there may be a specialized track for *genomic* counselors who will provide patient education and laboratory support not for rare monogenic disorders, but for common, polygenic conditions. Other health care professionals and/or multimedia educational programs will also need to assist with patient education. Additional responsibility will be placed upon the patient to actively gather and share medical information with their family through the use of interactive, computer-based family history tools. However, the majority of medical professionals outside genetics are currently ill equipped to provide genomic medical care. When surveyed, 65% of obstetricians and gynecologists who routinely utilize genetics in their practice did not consider themselves sufficiently educated in genetics. They rarely referred patients for genetic counseling and 86% did not obtain informed consent for genetic testing (Wilkins-Haug et al., 2000). Genomic education on a massive, unprecedented scale will be necessary to implement genomic medicine (see Chapter 18).

GENOMIC LITERACY

Integration of genomic information into health care will require at least rudimentary understanding of genetics and genomics by health care providers and health care consumers. Increasingly, there will be medicolegal implications for physicians if they do not recognize specific heritable risks based upon the patient's personal or family history (Deftos, 1998). Additional time must be dedicated to genetics and genomics within medical training, beyond the average 29 hours currently allocated in 4 years of medical school (Friedman et al., 1998). Medical educators themselves will need to keep abreast of advances within genomics and rapidly revise curricula. Because genomic medicine is so rapidly changing, it will be necessary to also teach the standards by which genetic tests should be evaluated for scientific validity and clinical utility so practitioners will be equipped to judge new data as they emerge.

As genomics is integrated into medical, dental, nursing, and allied professional education, a new generation of health care providers will hopefully be capable of rapidly assimilating genomic information into the life cycle of their patients. The younger generation of health care providers is already anxious to adopt this new technology in their practices. Perhaps, they even place more confidence in this new scientific field than is yet warranted. A greater challenge remains for health care professionals educated prior to the expansion of genomics – perhaps graduating from medical school only 5 years ago. For previously trained health care professionals, it will be necessary to learn general underlying principles of genetics and genomics as well as specific tools relevant to their own area of clinical expertise. The National Coalition for Health Professional Education in Genetics (NCHPEG) has produced a set of core competencies in genetics essential for all health care professionals (http://www.nchpeg.org/eduresources/core/Corecomps2005.pdf). NCHPEG is currently developing a core curriculum for genetic education (http://www.nchpeg.org/eduresources/core/coreprinciples.pdf) and has produced a CD ROM "Genetics and Common Disorders: Implications for Primary Care and Public Health Providers" that should assist in educating health care professionals currently in practice. Geneticists will play critical roles as genomic medical educators and discern when information is clinically relevant for various subspecialty areas with clinical care recommendations supported by major medical professional organizations. Genomic educators will need to continuously update medical professionals with new discoveries and applications in simple terms that emphasize clinical utility. It will remain a scientific challenge to define clinical situations in which complex genomic information can be distilled into an appropriately simple interpretation that has unambiguous clinical implications that impact clinical outcomes.

Genomic literacy in the public must simultaneously improve and remove the mysticism and misconceptions some patients may have regarding hereditary information. We must enlighten those who believe genetics is deterministic who would otherwise fail to modify health behaviors, believing instead in an inevitable, unalterable fate. There is a wide and growing divide between patients utilizing genetic and genomic information in their health care, driven in part by education and socioeconomic status. Younger patients tend to utilize genetic information more aggressively and seek genetic information both out of curiosity and to make more informed life and reproductive decisions. Younger generations have also become accustomed to advances in assisted reproductive technology and increasingly push the ethical and social boundaries to use such genetic and reproductive technology for such trivial genetic factors as elective sex selection. Older patients tend to be less educated about emerging genomic technologies, and often the reason they cite for seeking genetic testing is to provide information for their children or grandchildren rather than for themselves.

As the public becomes more genetically literate, there will likely be a push from diagnostic laboratories to market genetic testing directly to consumers. This has already been done with

BRCA1/BRCA2 testing and is done with nutriceuticals. While such advertising campaigns may be effective in increasing-awareness and forcing health care professionals to learn new information about genetic testing, there will also be the danger that inappropriate tests will be ordered and potentially misused by patients and physicians who are not fully informed.

2009 UPDATE

There have been over 150 associations demonstrated between common SNPs and disease traits using Genome Wide Association Studies for diseases as diverse as age-related macular degeneration, type 1 and type 2 diabetes, obesity, inflammatory bowel disease, prostate cancer, breast cancer, colorectal cancer, rheumatoid arthritis, systemic lupus erythematosus, celiac disease, multiple sclerosis, atrial fibrillation, glaucoma, asthma, and restless leg syndrome. Because allele frequencies often vary between ethnic groups, it is likely that the relative importance of these common variants identified will depend upon ethnicity and/or geographic ancestry. To date most studies have been done in European Caucasians, and it will be extremely important to extend the studies to other ethnic groups.

Most of these associations have identified novel genes that were not previously known to be involved in the associated disease processes. The relative risk for most of these common variants has been modest. In only a few cases that have the common variants identified had effect sizes >2. These exceptions include Complement Factor H (CFH) in age-related macular degeneration (Edwards et al., 2005; Haines et al., 2005; Klein et al., 2005), *ApoE4* in Alzheimer's disease (Strittmatter and Roses, 1996), and *LOXL1* in exfoliative glaucoma (Thorleifsson et al., 2007). In general, these association signals have identified regions for further study but have not yet identified functional variants. Furthermore, it is possible that a single locus may contain multiple independent common risk variants, as has been suggested for CFH in age-related macular degeneration (Li et al., 2006; Maller et al., 2006) and at *IL23R* for Crohn's disease (Duerr et al., 2006). In rare cases, the same loci that have common variants conferring relatively small effect are also the same genes for which rare variants confer large effect in small numbers of patients. Such examples include the LDL receptor for familial hypercholesterolemia and *KIR 6.2*, *WFS1*, and *TCF2* for type 2 diabetes. It is yet to be determined whether this will be a more generalizable finding for other genes identified by association studies. There are now large efforts underway to try and identify mutations with large effects in the genes identified through the genome wide association studies. Clearly, there will be high clinical utility for these rare mutations conferring large phenotypic effects; however, it is unclear how frequently such rare mutations will be identified.

How the common variants with modest increased risk will be clinically utilized is unclear at this time. It will be important to validate the utility of individual as well as composite risks conferred by multiple SNPs for common disease susceptibility. It is most likely that this will initially be applied to diseases such as age related macular degeneration for which there is an extremely high relative risk based on a few common variants. Clinical testing for age related macular degeneration is already available through several clinical laboratories and genotypes are being incorporated into clinical trials. Clinical testing is available through several companies such as deCODEme, Navigenics, and 23andMe for many of the common polymorphisms identified through association studies. Initial studies in the area of type 2 diabetes suggest that the relative contribution of these additional common genetic variants will be minimal when the other clinical risk factors are fully utilized (Lyssenko et al., 2008; Meigs et al., 2008). Genetic testing for these common SNPs is not currently being covered by third-party payers and has fallen into the domain of direct-to-consumer marketing with out of pocket costs to consumers ranging in price from ~$1000 to $2500. How this information is being used by patients is not yet clear, but these early adopters will provide valuable insight into the future potential utility for using these common variants in disease risk stratification, how patients interpret their relative and absolute risks, and how such a risk interpretation influences health behaviors.

2009 UPDATE REFERENCES

Duerr, R.H., Taylor, K.D., Brant, S.R., Rioux, J.D., Silverberg, M.S., Daly, M.J., et al. (2006). A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* 314(5804), 1461–1463.

Edwards, A.O., Ritter, R., 3rd, Abel, K.J., Manning, A., Panhuysen, C. and Farrer, L.A. (2005). Complement factor H polymorphism and age-related macular degeneration. *Science* 308(5720), 421–424.

Haines, J.L., Hauser, M.A., Schmidt, S., Scott, W.K., Olson, L.M., Gallins, P. et al. (2005). Complement factor H variant increases the risk of age-related macular degeneration. *Science* 308(5720), 419–421.

Klein, R.J., Zeiss, C., Chew, E.Y., Tsai, J.Y., Sackler, R.S., Haynes, C. et al. (2005). Complement factor H polymorphism in age-related macular degeneration. *Science* 308(5720), 385–389.

Li, M., Atmaca-Sonmez, P., Othman, M., Branham, K.E., Khanna, R., Wade, M.S. et al. (2006) CFH haplotypes without the Y402H coding variant show strong association with susceptibility to age-related macular degeneration. *Nat Genet* 38(9), 1049–1054.

Lyssenko, V., Jonsson, A., Almgren, P., Pulizzi, N., Isomaa, B., Tuomi, T. et al. (2008). Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N Engl J Med* 359(21), 2220–2232.

Maller, J., George, S., Purcell, S., Fagerness, J., Altshuler, D., Daly, M.J. et al. (2006). Common variation in three genes, including a non-coding variant in CFH, strongly influences risk of age-related macular degeneration. *Nat Genet* 38(9), 1055–1059.

Meigs, J.B., Shrader, P., Sullivan, L.M., McAteer, J.B., Fox, C.S., Dupuis, J. et al. (2008). Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N Engl J Med* 359(21), 2208–2219.

Strittmatter, W.J. and Roses, A.D. (1996). Apolipoprotein E and Alzheimer's disease. *Annu Rev Neurosci* 19, 53–77.

Thorleifsson, G., Magnusson, K.P., Sulem, P., Walters, G.B., Gudbjartsson, D.F., Stefansson, H. et al. (2007). Common sequence variants in the LOXL1 gene confer susceptibility to exfoliation glaucoma. *Science* 317(5843), 1397–1400.

ACKNOWLEDGEMENTS

Roberto Almazan and Josue Martinez provided assistance with manuscript preparation.

REFERENCES

Acharya, K., Ackerman, P.D. and Ross, L.F. (2005). Pediatricians' attitudes toward expanding newborn screening. *Pediatrics* 116, E476–E484.

Antoniou, A., Pharoah, P.D., Narod, S., Risch, H.A., Eyfjord, J.E., Hopper, J.L., Loman, N., Olsson, H., Johannsson, O., Borg, A. et al. (2003). Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: A combined analysis of 22 studies. *Am J Hum Genet* 72, 1117–1130.

Beutler, E., Felitti, V.J., Koziol, J.A., Ho, N.J. and Gelbart, T. (2002). Penetrance of 845G→A (C282Y) HFE hereditary haemochromatosis mutation in the USA. *Lancet* 359, 211–218.

Buckhaults, P. (2006). Gene expression determinants of clinical outcome. *Curr Opin Oncol* 18, 57–61.

Codner, E., Deng, L., Perez-Bravo, F., Roman, R., Lanzano, P., Cassorla, F., Chung, W. (2006). Glucokinase mutations in young children with hyperglycemia. In *Diabetes/Metabolism Research and Reviews* 22(5), 348–355.

Collins, F.S., Green, E.D., Guttmacher, A.E. and Guyer, M.S. (2003). A vision for the future of genomics research. *Nature* 422, 835–847.

Deftos, L.J. (1998). The evolving duty to disclose the presence of genetic disease to relatives. *Acad Med* 73, 962–968.

Deng, M.C., Eisen, H.J., Mehra, M.R., Billingham, M., Marboe, C.C., Berry, G., Kobashigawa, J., Johnson, F.L., Starling, R.C., Murali, S. et al. (2006). Noninvasive discrimination of rejection in cardiac allograft recipients using gene expression profiling. *Am J Transplant* 6(1), 150–160.

de Vries, B.B., Pfundt, R., Leisink, M., Koolen, D.A., Vissers, L.E., Janssen, I.M., Reijmersdal, S., Nillesen, W.M., Huys, E.H., Leeuw, N. et al. (2005). Diagnostic genome profiling in mental retardation. *Am J Hum Genet* 77, 606–616.

Friedman, J.M., Blitzer, M., Elsas, L.J., 2nd., Francke, U. and Willard, H.F. (1998). Clinical objectives in medical genetics for undergraduate medical students. Association of Professors of Human Genetics, Clinical Objectives Task Force. *Genet Med* 1, 54–55.

Green, N.S. and Pass, K.A. (2005). Neonatal screening by DNA microarray: Spots and chips. *Nat Rev Genet* 6, 147–151.

Hartge, P., Struewing, J.P., Wacholder, S., Brody, L.C. and Tucker, M.A. (1999). The prevalence of common BRCA1 and BRCA2 mutations among Ashkenazi Jews. *Am J Hum Genet* 64, 963–970.

Jazwinska, E.C., Cullen, L.M., Busfield, F., Pyper, W.R., Webb, S.I., Powell, L.W., Morris, C.P. and Walsh, T.P. (1996). Haemochromatosis and HLA-H. *Nat Genet* 14, 249–251.

Kauff, N.D., Mitra, N., Robson, M.E., Hurley, K.E., Chuai, S., Goldfrank, D., Wadsworth, E., Lee, J., Cigler, T., Borgen, P.I. et al. (2005). Risk of ovarian cancer in BRCA1 and BRCA2 mutation-negative hereditary breast cancer families. *J Natl Cancer Inst* 97, 1382–1384.

Korf, B.R., Feldman, G. and Wiesner, G.L. (2005). Report of Banbury Summit meeting on training of physicians in medical genetics, October 20–22, 2004. *Genet Med* 7, 433–438.

Lynch, T.J., Bell, D.W., Sordella, R., Gurubhagavatula, S., Okimoto, R.A., Brannigan, B.W., Harris, P.L., Haserlat, S.M., Supko, J.G., Haluska, F.G. et al. (2004). Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 350, 2129–2139.

Moss, A.J., Windle, J.R., Hall, W.J., Zareba, W., Robinson, J.L., McNitt, S., Severski, P., Rosero, S., Daubert, J.P., Qi, M. et al. (2005). Safety and efficacy of flecainide in subjects with Long QT-3 syndrome (DeltaKPQ mutation): A randomized, double-blind, placebo-controlled clinical trial. *Ann Noninvasive Electrocardiol* 10, 59–66.

Pinkel, D. and Albertson, D.G. (2005). Comparative genomic hybridization. *Annu Rev Genomics Hum Genet* 6, 331–354.

Rosendaal, F.R. (1999). Venous thrombosis: A multicausal disease. *Lancet* 353, 1167–1173.

Schoumans, J., Ruivenkamp, C., Holmberg, E., Kyllerman, M., Anderlid, B.M. and Nordenskjold, M. (2005). Detection of chromosomal imbalances in children with idiopathic mental retardation by array based comparative genomic hybridisation (array-CGH). *J Med Genet* 42, 699–705.

Schwartz, P.J., Priori, S.G., Spazzolini, C., Moss, A.J., Vincent, G.M., Napolitano, C., Denjoy, I., Guicheney, P., Breithardt, G., Keating, M.T. et al. (2001). Genotype-phenotype correlation in the long-QT syndrome: gene-specific triggers for life-threatening arrhythmias. *Circulation* 103, 89–95.

Sebat, J., Lakshmi, B., Malhortra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J. et al. (2007). Strong association of de novo copy number mutations with autism. *Science* 316, 445–449.

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S., Massa, H., Walker, M., Chi, M. et al. (2004). Large-scale copy number polymorphism in the human genome. *Science* 305, 525–528.

Vandenbroucke, J.P., Koster, T., Briet, E., Reitsma, P.H., Bertina, R.M. and Rosendaal, F.R. (1994). Increased risk of venous thrombosis in oral-contraceptive users who are carriers of factor V Leiden mutation. *Lancet* 344, 1453–1457.

Verhoog, L.C., Brekelmans, C.T., Seynaeve, C., van den Bosch, L.M., Dahmen, G., van Geel, A.N., Tilanus-Linthorst, M.M., Bartels, C.C., Wagner, A., van den Ouweland, A. et al. (1998). Survival and tumour characteristics of breast-cancer patients with germline mutations of BRCA1. *Lancet* 351, 316–321.

Waisbren, S.E., Albers, S., Amato, S., Ampola, M., Brewster, T.G., Demmer, L., Eaton, R.B., Greenstein, R., Korson, M., Larson, C. et al. (2003). Effect of expanded newborn screening for biochemical genetic disorders on child outcomes and parental stress. *JAMA* 290, 2564–2572.

Wilkins-Haug, L., Erickson, K., Hill, L., Power, M., Holzman, G.B. and Schulkin, J. (2000). Obstetrician-gynecologists' opinions and attitudes on the role of genetics in women's health. *J Womens Health Gend Based Med* 9, 873–879.

Wilson, J.M. and Jungner, Y.G. (1968). *Principles and practice of screening for disease.* World Health Organization, Geneva.

RECOMMENDED RESOURCES

http://www.genetests.org
http://www.geneclinics.org/
http://mchb.hrsa.gov/screening
http://www.acmg.net/resources/s-g/s-g-yes-no.asp
http://www.nchpeg.org/eduresources/core/Corecomps2005.pdf
http://www.nchpeg.org/eduresources/core/coreprinciples.pdf

CHAPTER

The Role of Genomics in Enabling Prospective Health Care

Ralph Snyderman

INTRODUCTION

In the early 1900s, the emerging sciences of physiology, chemistry, immunology, microbiology, and radiology began to be introduced into the practice of medicine, which at that time, considered disease to be due to imbalances of bodily humors and transmitted through climacteric miasmas. For the first time in history, the application of science to medicine laid a solid foundation for understanding the pathophysiology of disease (Bynum, 2002; Snyderman, 2004; Snyderman and Williams, 2003). In particular, microbiology, through the identification of "causative" agents for numerous infectious diseases, had a profound impact on concepts concerning the cause and potential cure of illnesses. Given the increasing number of diseases determined to have specific "causes" along with the new found ability of chemistry to synthesize therapeutic "magic bullets," early 20th Century medicine logically pursued a scientific, reductionist approach to health care which continues today: that is, identify a disease's root cause and eliminate it. This approach has enabled medicine to reverse some diseases, prolong life, and at times, effect wondrous cures.

The reductionist approach to medicine, however, is limited in that the evolution of virtually all diseases is more complicated than a single root cause. The development of disease is based upon one's genetic susceptibilities, complex interactions between initiating factors, and health status. In the aggregate, these can lead to disease progression, or not (Figure 17.1). For example, while the tubercle bacillus is the causative agent of tuberculosis, the clinical manifestation of exposure, as a consequence of host susceptibility, can range from transient subclinical infection to rapidly fatal miliary disease. Even a well-defined genetic disease such as sickle cell anemia, known to be due to a single mutation, can have manifestation ranging from death in adolescence to far milder forms compatible with long-term survival. By focusing primarily on the reductionist model of disease development, our health care system has, unfortunately, not focused on the prevention or treatment of complex chronic diseases. At present, individuals with any of five chronic conditions account for roughly two-thirds of all health care expenses (Figure 17.2).

A more rational approach to health care is now possible. New emerging sciences, notably genomics, can transform medical practice now just as new applications of science and know-how transformed medicine 100 years ago (Figure 17.3). Genomic and related expanding fields of research can impact health care in numerous ways (Figure 17.4), but most profound may be their ability to predict risk, track disease progression and anticipate clinical events, thereby enabling truly personalized, predictive, preventative care. To do this, the strengths of today's disease-oriented reductionist approach must be integrated with an approach that focuses on prevention and minimization of disease with an emphasis on long-term strategies.

This personalized, predictive, and preventative approach to health care is known as "prospective health care" (Snyderman and

Copyright © 2010, Elsevier Inc.
All rights reserved.

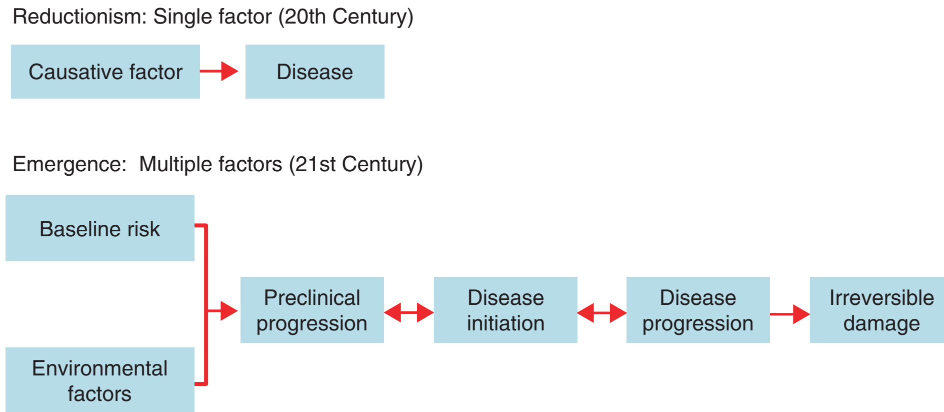
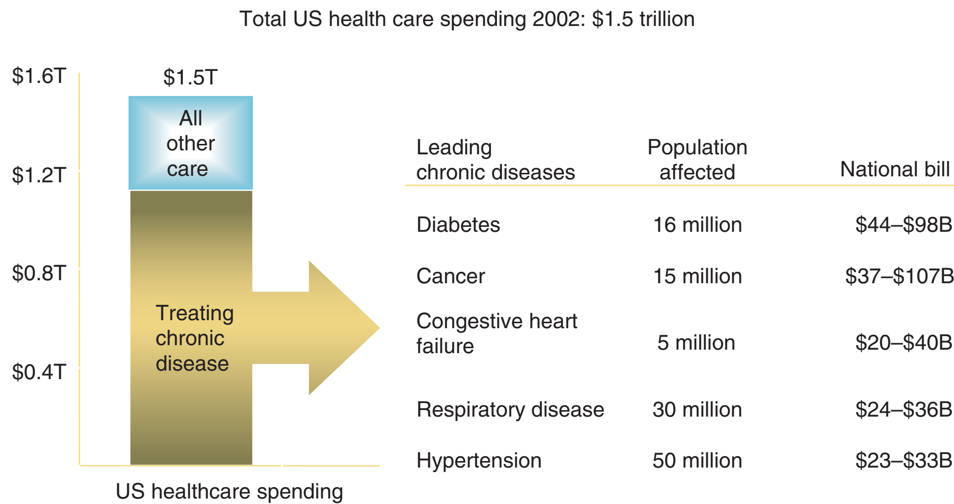


Figure 17.1 Concept of disease: reductionism versus emergence. The reductionist approach considers disease to be a consequence of a pathogenic factor; that is microbe. More accurately, disease development depends on the host's susceptibility to pathogenic factors and exposure to them (Snyderman and Yoediono, 2006).



Leading chronic diseases	Population affected	National bill
Diabetes	16 million	$44–$98B
Cancer	15 million	$37–$107B
Congestive heart failure	5 million	$20–$40B
Respiratory disease	30 million	$24–$36B
Hypertension	50 million	$23–$33B

Figure 17.2 Consequences of current approaches to health care. While little is spent on prevention, nearly three-quarters of health care expenditures fund the treatment of late-stage chronic disease. (Source: American Heart Association, American Cancer Society, American Lung Association, National Institute of Diabetes and Digestive and Kidney Disease.)

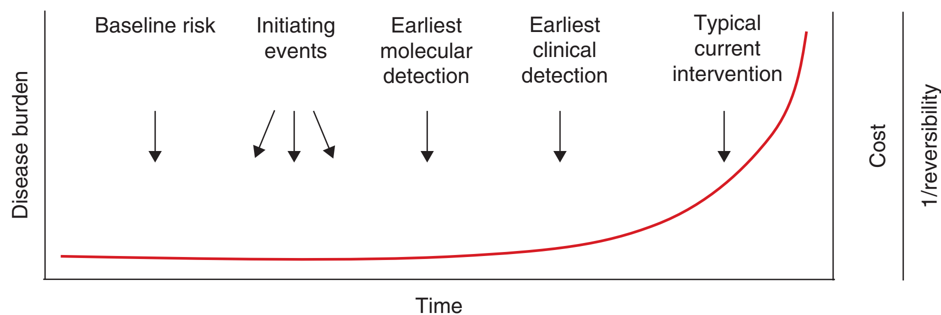


Figure 17.3 Disease progression. Diseases develop over time, with pathology increasing, reversibility decreasing and cost of care increasing (Snyderman and Yoediono, 2006).

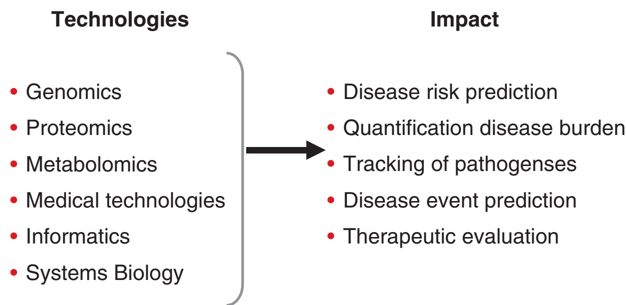


Figure 17.4 Impact of new technologies on prospective health care. Last century's science enabled reactive responses to disease while this century's science enables prospective approaches as well.

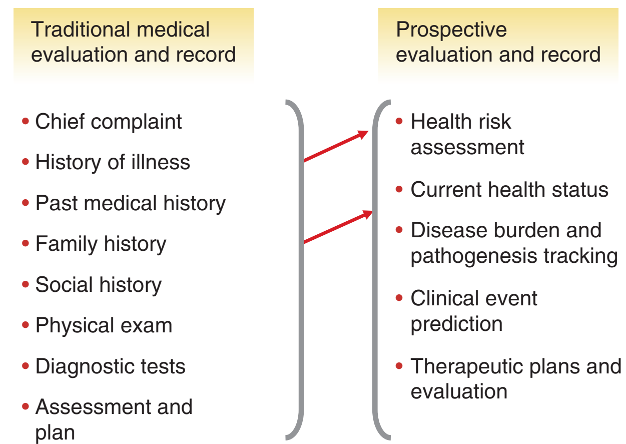


Figure 17.5 Current medical record versus prospective approach to health care. Physicians currently evaluate patients using a disease-focused approach. A more effective approach would include strategic health planning as indicated by the prospective approaches shown on the right (Snyderman and Yoediono, 2006).

Williams, 2003; Snyderman and Yoediono, 2006; West et al., 2006; Williams et al., 2003). Key features of prospective health care include personalized health risk assessment, tracking of pathogeneses, prediction of clinical events, and strategic planning to mitigate risks and evaluate therapeutic benefits. With a prospective health care model, individuals will be provided with personalized health plans which promote strategic health planning by incorporating individualized risk assessments for diseases, disease tracking, clinical event predictions and therapeutic planning (Figure 17.5). Genomic research will play an integral role in providing these capabilities given that the molecular processes underlying disease susceptibility, progression, and therapeutic response differ among individuals and can be measured, in part, through genomic analyses. Technological advances allowing the measurements at the DNA, mRNA, protein, and metabolic levels are already beginning to enable inferences regarding health risks or clinical outcomes (Snyderman and Langheier, 2006).

PREDICTIVE MODELS

At the core of prospective health care are personalized health risk assessment and predictive modeling. A predictive model uses statistical algorithms to identify factors that predict events (Anderson et al., 1991). The predictive model can be used to forecast events based on factors or biomarkers, which are most likely to correlate with or be causative of the future clinical event. The development of a predictive model uses various methods, such as logistical regression or neural network, to differentiate predictive factors from other variables which are not as useful for anticipating the clinical outcome of interest. The predictive model can then be applied to a current situation to determine the risk and timing of an event occurring. The better the model, the better the accuracy in predicting the occurrence and timing of the event. The details involved in developing a predictive model, as well as other related components such as individualized patient databases and risk model libraries, are beyond the scope of this chapter, but a simplified version is shown in Figure 17.6.

PREDICTIVE FACTORS

The predictive factors most commonly used for disease-related risk assessments encompass clinical, demographic, family history, and laboratory data. These types of data can provide insight into the likelihood of an individual developing a condition or event and are relatively cost-effective and collected routinely. However, generally they are limited in terms of being able to accurately predict *when* disease will occur. Furthermore, many false-positives and false-negatives are associated with these types of data, which are often epidemiologic in nature (i.e., Framingham Study) (Brindle et al., 2003; The International HapMap Consortium, 2003). Predictive factors, which serve as more accurate markers of disease susceptibility, likely will come in the form of biomarkers which are directly related to the cause of the clinical event. A biomarker is a characteristic (i.e., expressed gene, protein or metabolite) that is objectively measured and evaluated as an indicator of biological or pathogenic processes or pharmacological responses to therapeutic interventions. Genomic research will play an integral role in identifying predictive biomarkers which, in combination with other types of predictive factors, will enable more accurate risk analyses for baseline risk assessment, disease tracking, clinical event predictions and response to therapeutics. Biomarkers can be thought of as belonging to two general categories: (1) stable biomarkers that are inherited and/or change rarely over a lifetime, and (2) dynamic biomarkers which express ongoing biological activities. The latter include gene expression, protein expression and the measurement of metabolic factors. Measurement of genomic factors that lie in the causal pathway of a disease or a therapeutic response, or factors such as single-nucleotide polymorphisms (SNPs) that are highly associated with causal genes, will serve as better predictors of adverse outcomes than much of the demographic data now being collected. Stable DNA gene predictors will enhance baseline clinical risk assessment and primary prevention, while dynamic mRNA, protein and metabolic factors will reflect ongoing biological or pathological processes.

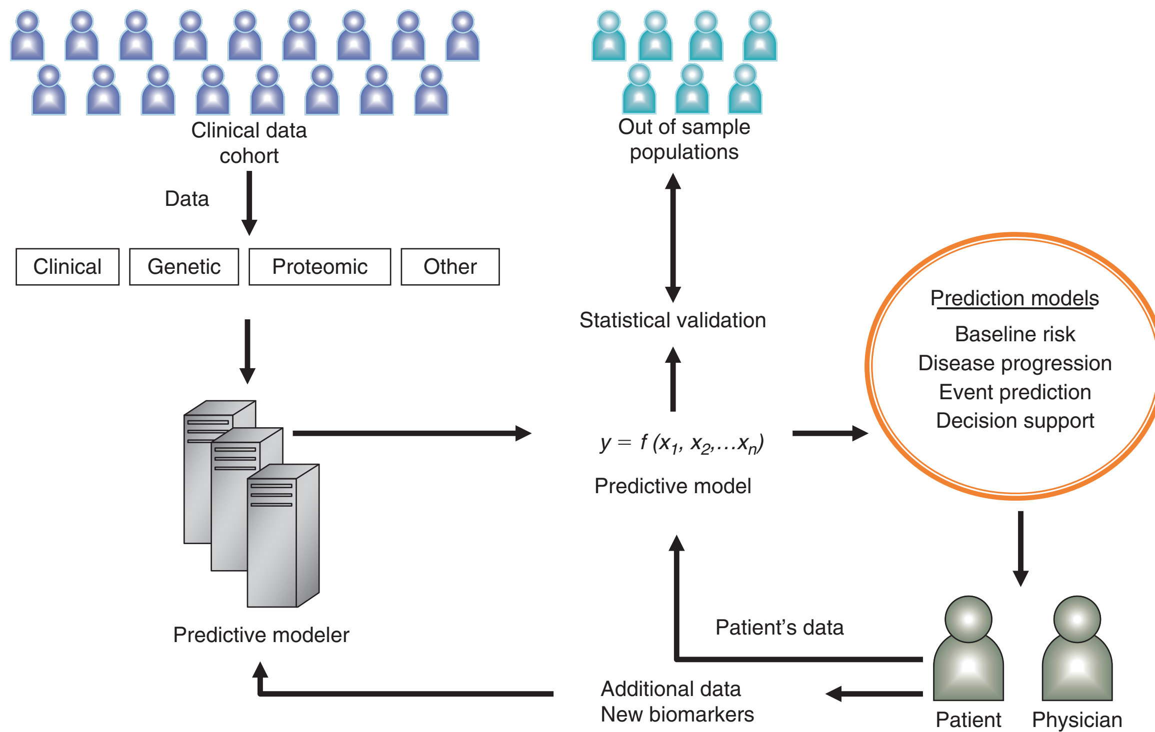


Figure 17.6 Methodology for creating clinical risk predictive models. Relevant clinical data and/or biomarkers are statistically analyzed to create validated risk models for particular clinical conditions or events (Snyderman and Yoediono, 2006).

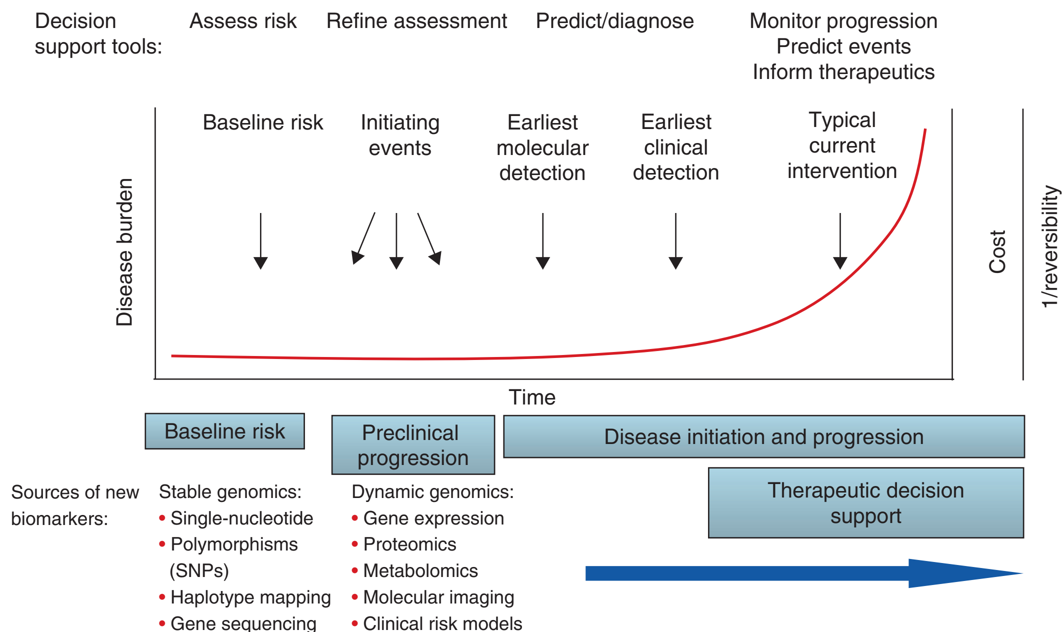


Figure 17.7 Contributions of new technologies to disease prediction and prediction of clinical events (Snyderman and Yoediono, 2006).

Analysis of these will enhance refined risk assessments to track disease progression, predict events, and guide therapeutic choices (Figure 17.7).

The advantage of genotypic data for baseline prediction is that they can be collected at birth or any time during one's life and, theoretically, needs to be collected only once. Baseline risk assessments using demographic data or static genomic information will likely have lower specificity (a higher number of false positives) than molecular measures that are dynamic and change as a function of time. Nonetheless, as more information regarding

inherited risks is gathered, baseline risk assessment will become increasingly useful. For secondary prevention (for example, predicting heart attack in an individual with diabetes), stable genomic data may be less valuable. In this case, relevant dynamic indicators will provide a more powerful predictor of the disease event. Over time, the cost of genotyping should plummet and the value of baseline risk analysis should increase markedly. Moreover, the identification of clinically valuable dynamic biomarkers is an area of fertile research and will enhance accurate ongoing risk assessment timing of disease events and prediction of therapeutic outcomes.

Identifying accurate predictors for baseline risk assessment will be facilitated greatly by clinical research and the HapMap project and the Genetic Association Information Network (GAIN) initiative, a public/private partnership between the NIH, foundation of NIH and industry. The GAIN study will fund $20 million of research to identify SNPs associated with common diseases. The International HapMap Consortium is characterizing common patterns of DNA sequence variation and the extent of linkage disequilibrium in the human genome. This will facilitate the characterization of genotypes and identification of key SNPs related to chronic disease; traditional and advanced association algorithms will allow the analysis of the HapMap (International HapMap Project, http://www.hapmap.org)(Niu, 2004; Liu et al., 2004; Evans et al., 2004). Online Mendelian Inheritance in Man (OMIM), a database of disease risk genes, is already revealing an increasing number of stable genomic factors that should be useful in risk assessment (Online Mendelian Inheritance in Man, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM). The role of an individual's gene variants in altering their metabolism and response to drugs is becoming important in drug development and in certain areas of clinical practice, particularly oncology (Lee et al., 2005).

For individuals whose genes, SNPs, family history, or clinical information identify them as high risk for a particular disease, surveillance will be needed to track possible disease progression and, when relevant, to evaluate therapeutic support. Such tracking will likely include the measurement of dynamic factors, including gene-expression, proteomics, and metabolomic assessments. Analyses to track baseline risk and disease development will hopefully be incorporated into personalized health plans in the future (Langheier and Snyderman, 2004). For example, children with a family history of type 1 diabetes could have a baseline risk assessment that evaluates various predictive SNPs. Those determined to be at high risk could undergo a surveillance protocol, tracking their levels of biomarkers associated with actual disease development and progression. (Barker et al., 2004; Eisenbarth, 2004). As effective preventative therapies are identified, these analyses could guide appropriate intervention before β-cells are destroyed.

Initial applications of genomic technologies are being applied to predict outcomes in defined clinical conditions. For example, gene-expression microarrays and proteomic techniques show promise for identifying the aggressiveness of cancer, allowing the creation of predictive models for likely survival time with and without treatment (Anderson and LaBaer, 2005; Berchuck et al., 2005; Pittman et al., 2004; Rich et al., 2005). Moreover, gene expression in circulating mononuclear cells is being used to predict organ rejection in patients with heart transplants, obviating the need for myocardial biopsy in some conditions (Deng et al., 2006). Many gene-expression tests are being developed to gauge appropriate chemo-therapeutic regimens for a given patient's cancer (Bild et al., 2006).

2009 UPDATE

Emerging Initiatives

In 2007, the US Department of Health and Human Services announced a new initiative called "Personalized Health Care," the objective of which was to combine genomic research breakthroughs with advances in health information technology to enable gene-based medicine, where one's susceptibilities to diseases would be quantified as early as possible so that preventive countermeasures could be taken (Personalized Health Care, http://www.hhs.gov/myhealthcare). An example of one of the many programs encompassed by the "Personalized Health Care" initiative is entitled "The Cancer Genome Atlas," a collaboration between the National Cancer Institute and the National Human Genome Research Institute. Researchers have been utilizing evolving technologies such as large-scale genome sequencing to better understand the molecular basis of cancer. Initial efforts have focused on cancers affecting the brain, lungs, and ovaries. This goal of this project is to improve capabilities for preventing, diagnosing, and treating cancer at a personalized level (The Cancer Genome Atlas, http://cancergenome.nih.gov).

The NIH has also collaborated with the Food and Drug Administration (FDA), the Centers for Medicare and Medicaid Services (CMS), as well as the Pharmaceutical Research and Manufacturers of America (PhRMA) and the Biotechnology Industry Organization (BIO), on initiatives such as "The Biomarkers Consortium." The goal of this consortium is to identify and validate new biomarkers to help expedite the

development and utilization of therapeutics and technologies for personalized prevention, early detection, and treatment of diseases (The Biomarkers Consortium, http:/www.biomarker-sconsortium.org).

A project in The Biomarkers Consortium under consideration by the National Institute of Diabetes and Digestive and Kidney Diseases focuses on the identification of biomarkers associated with pre-diabetes and type II diabetes and the creation of more reliable, less expensive, and faster predictive tests (Zerhouni et al., 2007).

Academic health centers such as the Lewis-Sigler Institute for Integrative Genomics at Princeton University, the Broad Institute of MIT and Harvard, and the Center for Genomic Medicine at the Duke Institute for Genome Sciences & Policy (IGSP) have been dedicated to identifying and validating predictive biomarkers. One project at the Broad Institute, entitled "Clinical Proteomic Technology Assessment for Cancer (CPTAC)," is seeking to establish standards for protein biomarker assessment (Clinical Proteomic Technology Assessment for Cancer Program, http://www.broad.mit.edu/proteomics/CPTAC), and at Duke's IGSP researchers recently outlined a genomic-based strategy for the personalized treatment of patients who have advanced-stage ovarian cancer (Dressman et al., 2007).

Centers within academic institutions, such as the IGSP Center for Systems Biology at Duke University and the Center for Cancer Systems Biology (a collaboration between the Dana-Farber Cancer Institute and Harvard Medical School), have dedicated vast resources to personalized medicine. Researchers from the Center for Cancer Systems Biology detailed an approach to identify genes associated with increased breast cancer risk. By utilizing this strategy, they determined that hyaluronan-mediated motility receptor (HMMR) was a potential breast cancer susceptibility gene (Pujana et al., 2007).

Although many of the recent research initiatives have been genome-based, genomic advances alone will not likely be sufficient to enable personalized medicine. Rather, multiple rapidly evolving predictive technologies involving other "omics," such as proteomics and metabolomics, will complement and enhance clinical know-how and capabilities to allow better predictive approaches. For example, studies on risk-predictive modeling for breast cancer recurrence have shown that the predictive accuracy of a combined clinical and genomic model is higher than a model utilizing just one type of data (Pittman et al., 2004). Furthermore, a risk predictive model for severe adverse outcomes from cancer chemotherapy was developed using only clinical and laboratory data (Lyman et al., 2006).

Consumer-Directed Products

Although many risk assessment tools are directed toward providers, other companies have developed personalized risk assessment tools for sale to consumers (Navigenics, Inc., http://www.navigenics.com; 23andMe, https://www.23andme.com). An assessment is available that analyzes 250 serum proteins to provide early diagnosis of clinically unapparent diseases (Biophysical Corporation, http://www.biophysicalcorp.com). In addition, consumers receive a medical history interview and a follow-up doctor consultation. Likewise, consumers may pursue individualized health and nutrition recommendations, which are based in part upon a genomic analysis (Sciona, http://www.sciona.com). Still other options are available from companies that perform genomic analysis to identify known risk-associated variants. Consumers are offered predictive information and options for updated genomic analysis as new susceptibility factors are identified. The direct-to-consumer marketing of diagnostics and predictive tools is a relative new phenomenon and its value and means of regulation have yet to be determined (see also Updates for Chapters 13 and 18-21).

With the rapid evolution of fields such as genomics, patients have become increasingly concerned about medical privacy, genetic discrimination by insurers and employers, and related issues. Indeed, there has been evidence that even healthy persons with a genetic susceptibility for a condition have faced discrimination by insurance companies and that those with genetic diseases have had a difficult time getting health insurance (Plantinga et al., 2003; AHRQ Launches New "Effective Health Care Program" to Compare Medical Treatments, http://www.ahrq.gov/news/press/pr2005/effectivepr.htm). This is an evolving area, which will necessitate careful policy development. These issues as well as others, such as assessing the reliability of data entered by patients or the accuracy of risk assessment tools, are legitimate concerns, which must be continuously monitored and addressed by the appropriate constituents.

2009 UPDATE REFERENCES

AHRQ Launches New "Effective Health Care Program" to Compare Medical Treatments. http://www.ahrq.gov/news/press/pr2005/effectivepr.htm

Biophysical Corporation. http://www.biophysicalcorp.com

Clinical Proteomic Technology Assessment for Cancer (CPTAC) Program. http://www.broad.mit.edu/proteomics/CPTAC

Dressman, H.K., Berchuck, A., Chan G. et al. (2007). An integrated genomic-based approach to individualized treatment of patients with advanced-stage ovarian cancer. *J Clin Oncol* 25, 517–525.

Lyman, G.H., Kuderer, N.M., Crawford, J. et al. (2006). Prospective validation of a risk model for first cycle neutropenic complications

in patients receiving cancer chemotherapy. ASCO Annual Meeting Proceedings Part I. *J Clin Oncol* 24(18S) (abstract).

Navigenics, Inc. http://www.navigenics.com

Personalized Health Care. http://www.hhs.gov/myhealthcare

Pittman, J., Huang, E., Dressman, H. et al. (2004). Integrated modeling of clinical and gene expression information for personalized prediction of disease outcomes. *Proc Natl Acad Sci USA* 101(22), 8431–8436.

Plantinga, L., Natowicz, M.R., Kass, N.E., Hull, S.C., Gostin, L.O. and Faden, R.R. (2003). Disclosure, confidentiality, and families: Experiences and attitudes of those with genetic versus nongenetic medical conditions. *Am J Med Genet C Semin Med Genet* 119(1), 51–59.

Pujana, M.A., Han, J.-D.J., Starita, L.M. et al. (2007). Network modeling links breast cancer susceptibility and centrosome dysfunction. *Nat Genet* 39(11), 1338–1349.

Sciona. http://www.sciona.com

The Biomarkers Consortium. http:/www.biomarkersconsortium.org

The Cancer Genome Atlas. http://cancergenome.nih.gov

Zerhouni, E.A., Sanders, C.A. and von Eschenbach, A.C. (2007). The Biomarkers Consortium: public and private sectors working in partnership to improve the public health. *Oncologist* 12(3), 250–252.

23andMe. https://www.23andme.com

ACKNOWLEDGEMENTS

The author would like to acknowledge the outstanding editorial assistance of Dr Ziggy Yoediono and Ms Cindy Mitchell.

REFERENCES

Anderson, K.M., Wilson, P.W., Odell, P.M. and Kannel, W.B. (1991). An updated coronary risk profile. A statement for health professionals. *Circulation* 83, 356–362.

Anderson, K.S. and LaBaer, J. (2005). The sentinel within: Exploiting the immune system for cancer biomarkers. *J Proteome Res* 4, 1123–1133.

Barker, J.M., Barriga, K.J., Yu, L., Miao, D., Erlich, H.A., Norris, J.M., Eisenbarth, G.S. and Rewers, M. (2004). Diabetes Autoimmunity Study in the Young. Prediction of autoantibody positivity and progression to type 1 diabetes: Diabetes Autoimmunity Study in the Young (DAISY). *J Clin Endocrinol Metab* 8, 3896–3902.

Berchuck, A., Iversen, E.S., Lancaster, J.M., Pittman, J., Luo, J., Lee, P., Murphy, S., Dressman, H.K., Febbo, P.G., West, M. et al. (2005). Patterns of gene expression that characterize long-term survival in advanced stage serous ovarian cancers. *Clin Cancer Res* 11, 3686–3696.

Bild, A.H., Yao, G., Chang, J.T., Wang, Q., Potti, A., Chasse, D., Joshi, M.B., Harpole, D., Lancaster, J.M., Berchuck, A. et al. (2006). Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 439(7074), 274–275.

Brindle, P., Emberson, J., Lampe, F., Walker, M., Whincup, P., Fahey, T. et al. (2003). Predictive accuracy of the Framingham coronary risk score in British men: Prospective cohort study. *BMJ* 327, 1267–1270.

Bynum, W. (2002). The evolution of germs and the evolution of disease: Some British debates, 1870–1900. *Hist Philos Life Sci* 24(1), 53–68.

Deng, M.C., Eisen, H.J., Mehra, M.R., Billingham, M., Marboe, C.C., Berry, G., Kobashigawa, J., Johnson, F.L., Starling, R.C., Murali, S. et al. (2006). Non-invasive discrimination of rejection in cardiac allograft recipients using gene expression profiling. *Am J Transplant* 6, 150–160.

Eisenbarth, G.S. (2004). Prediction of type 1 diabetes: The natural history of the prediabetic period. *Adv Exp Med Biol* 552, 268–290.

Evans, D.M., Cardon, L.R. and Morris, A.P. (2004). Genotype prediction using a dense map of SNPs. *Genet Epidemiol* 4, 375–384.

Langheier, J.M. and Snyderman, R. (2004). Prospective Medicine: The role for genomics in personalized health planning. *Pharmacogenomics* 5(1), 1–8.

Lee, W., Lockhart, A.C., Kim, R.B. and Rothenberg, M.L. (2005). Cancer pharmacogenomics: Powerful tools in cancer chemotherapy and drug development. *Oncologist* 2, 104–111.

Liu, T., Johnson, J.A., Casella, G. and Wu, R. (2004). Sequencing complex diseases with HapMap. *Genetics* 168, 503–511.

Niu, T. (2004). Algorithms for inferring haplotypes. *Genet Epidemiol* 27, 334–347.

Pittman, J., Huang, E., Dressman, H., Horng, C.F., Cheng, S.H., Tsou, M.H., Chen, C.M., Bild, A., Iversen, E.S., Huang, A.T. et al. (2004). Integrated modeling of clinical and gene expression information for personalized prediction of disease outcomes. *Proc Natl Acad Sci USA* 101, 8431–8436.

Rich, J.N., Hans, C., Jones, B., Iversen, E.S., McLendon, R.E., Rasheed, B.K., Dobra, A., Dressman, H.K., Bigner, D.D., Nevins, J.R. et al. (2005). Gene expression profiling and genetic markers in glioblastoma survival. *Cancer Res* 65, 4051–4058.

Snyderman, R. (2004). AAP Presidential address: The AAP and the transformation of medicine. *J. Clin Invest* 114, 1169–1173.

Snyderman, R. and Langheier, J. (2006). Prospective health care: The second transformation of medicine. *Genome Biol* 7, 104–127.

Snyderman, R. and Williams, R.S. (2003). Prospective medicine: The next health care transformation. *Acad Med* 78(11), 1079–1084.

Snyderman, R. and Yoediono, Z. (2006). Prospective care: A personalized, preventative approach to medicine. *Pharmocogenomics* 7(1), 5–9.

The International HapMap Consortium (2003). The International HapMap Project. *Nature* 426, 789–796.

West, M., Ginsburg, G., Huang, A. and Nevins, J. (2006). Embracing the complexity of genomic data for personalized medicine. *Genome Res* 16, 559–566.

Williams, R.S., Willard, H.F. and Snyderman, R. (2003). Personalized health planning. *Science* 300, 549.

RECOMMENDED RESOURCES

International HapMap Project [http://www.hapmap.org]

Medical Device Link [http://www.devicelink.com/ivdt/archive/03/04/002.html]

Online Mendelian Inheritance in Man [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM]

CHAPTER

Genome Policy Considerations for Genomic Medicine

Susanne B. Haga

INTRODUCTION

The success of the Human Genome Project will be measured largely with respect to advancements in biology and the realization of a new era of medicine based on the use of genomic data to predict, prevent, diagnose, and treat disease. However, the scientific challenge of sequencing the human genome will likely pale in comparison to the efforts required to translate genome sciences into personalized medicine. As we begin to apply the data and technologies arising from the Human Genome Project and subsequent projects including the HapMap Project (International HapMap Consortium, 2005) and the ENCODE Project (The ENCODE Project Consortium, 2007), the initial discovery, validation, and the development of a clinical test or therapy will be influenced by a range of science, health, and public policies. Although many of the policy issues are not unique to those faced by other new medical innovations, the genome revolution raises issues that span both the traditional science and health policy arenas (see Figure 18.1), some warranting resolutions specific to genomic medicine.

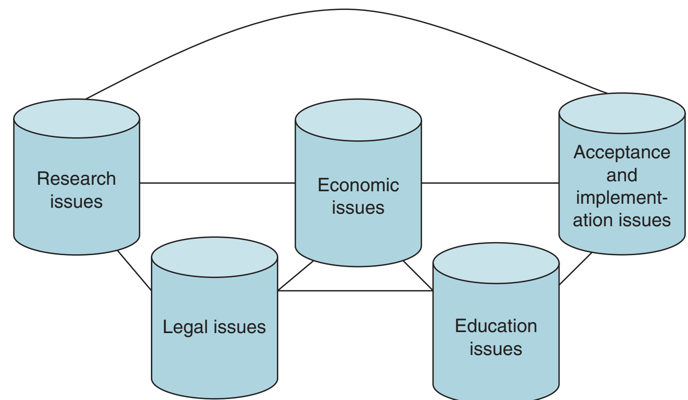


Figure 18.1 The spectrum and inter-connectedness of genome policy issues, based on (Haga and Willard, 2006).

This chapter will provide an overview of the major policy issues pertaining to the research, development, and translation of genomic medicine applications, including: research allocation and prioritization; the use and analyses of race in genome studies; the ethical issues linked to large-scale genome efforts such as biobanks, oversight, coverage, and reimbursement of new genomics applications; privacy and discrimination; and enhancing public and professional awareness. An in-depth analysis of several of the issues introduced here can be found in other chapters in this volume (see Chapters 13, 19 and 20).

Copyright © 2010, Elsevier Inc.
All rights reserved.

GENOME RESEARCH AFTER THE HUMAN GENOME PROJECT

The completion of the Human Genome Project marked the beginning of a new era, rather than the end of an era (Collins, 2003). Although a substantial investment in basic research, it is unlikely that anyone could have foreseen the widespread impact of not only the ultimate product – a reference sequence of the human genome – but also the parallel technology development that enabled the early completion of the sequence of the human genome and the genomes of countless other species. Its impact on the biotechnology sector and other fields as diverse as evolutionary biology, agriculture, biodefense, and medicine is still in the early stages but its scope has been staggering already.

Research Allocation and Prioritization

In 2003, after the completion of the Human Genome Project, a new vision for genomics research was created (Collins et al., 2003). The next 5 years of genomic research would concentrate on three major areas – society, health, and biology. Across these three areas, six crosscutting research tracts were identified, including technology development and ethical, legal, and social issues. This differs markedly from past strategic plans that focused predominantly on mapping and sequencing of the human genome and model organisms (Collins and Galas, 1993; Collins et al., 1998, 2003).

Despite the huge investment in basic biomedical research and the excitement generated by the Human Genome Project, concerns have been raised that the investment has not resulted in the anticipated development of new products, as the number of submissions and approval of new drugs and diagnostics has been on the decline (FDA, 2004). Because of this and political pressures to demonstrate return on investment from the doubling of the US NIH budget, the emphasis over the past several years has shifted toward translational and clinical science (Zerhouni, 2005). Recognition of the importance of translational research has led to changes in science policy such as the development of the NIH Roadmap (Zerhouni, 2003) and emphasis on the development of new tools and applications (FDA, 2006a). Translational genomics research will be an important part in each of these new efforts (see Chapter 14).

Over the past decade, Europe has also increased its supports for both basic and translational genomics research. In 2001, the European Commission allocated €100 million for an initiative called *Genome Research for Human Health*. Starting in 2002, the European Community Framework Program for Research, Technological Development and Demonstration (otherwise known as FP6) identified genomics as one of its seven major thematic areas (European Commission, 2002). The research budget for life sciences, genomics, and biotechnology was more than €2 billion. The program aimed to promote European genomics research through three major strategies: establishing a robust infrastructure; fostering interdisciplinary research with other fields such as physics, chemistry, mathematics, and computer science; and integrating genomics into medicine and biotechnology. In 2007, FP7 was revised with the primary goal of supporting Europe's global competitiveness. Genomics and biotechnology continue to have a prominent role in the development of new applications in medicine and food and agriculture. For the first time, investigator-driven projects will be funded in addition to the continued support of network initiatives (European Commission, 2002).

The field of genome sciences has already taken the next step forward toward applying this new knowledge and capability to medically relevant initiatives. For example, the NIH Chemical Genomics project involves the development of a public repository of small organic molecules to study cellular pathways in health and disease and hasten the identification of new drug targets and drug development (NIH Chemical Genomics Center, 2005). The Medical Sequencing Project involves the sequencing of well-phenotyped patients with autosomal and X-linked Mendelian diseases as well as common diseases to identify the underlying genetic cause of disease (National Human Genome Research Institute, 2007).

With respect to diagnostic screening and testing, the emphasis on translational research has provided opportunities to gather data about the clinical utility of new tools. Demonstrating clinical validity and utility is critical to the development of clinical guidelines and reimbursement policies that will support the appropriate use of these tests. For example, the National Cancer Institute is supporting a long-term prospective study called The *T*rial *A*ssigning *I*ndividua*L*ized *O*ptions for Treatment (*Rx*), or TAILORx (National Cancer Institute, 2006). This study will explore whether an expression profile of 16 genes associated with risk of recurrence for women with early stage breast cancer is useful in identifying those women who will most likely benefit from chemotherapy in addition to radiation and hormonal therapy. The study will follow 10,000 women for a period of 10 years, with follow-up of up to 20 years after initial therapies. While clinical trials represent the gold standard of evidence, these types of studies will not be feasible for every new genome application given the cost and time required, and other study approaches will need to be utilized. The continued support across translational genome research will be critical to the development of safe and effective applications in genomic medicine.

Race and Genomics

Population genetics research has spurred debate regarding the validity and use of racial categories as a variable in biological research. In particular, genomics research has provided insight into the genetic diversity (or lack thereof) of the human population, giving rise to questions about the use of race as a variable in biomedical research and whether genomics could provide improved measurements of variation in place of the concept of race (Burchard et al., 2003; Cooper et al., 2003). While a major conclusion of the Human Genome Project was the high degree of sequence identity between individuals, subsequent research has focused on the small differences between individuals and populations, particularly with respect to disease (Foster, 2005).

For instance, the 3-year US-led HapMap project generated a map of one million common genetic variants through analyses of more than 200 genomes of individuals from four different populations (International HapMap Consortium, 2005).

Similar to the HapMap project, the Mexican Genome Project, started in 2005, sought to determine the genetic variation of the heterogeneous *mestizo* population in Mexico (Mothelet and Herrera, 2005). Mestizo refers to individuals with mixed ancestry of European (primarily Spaniards) and Indian. The goal of the project was to determine whether any genetic differences between populations may be correlated to known health differences. Preliminary analysis of data from the project concluded that mestizo ancestry is a mixture of 35 ethnic groups (Wall, 2007). In general, 65% of Mexican ancestry can be traced to indigenous populations and 35% is due to non-indigenous groups (African, Asian, and European). Not surprisingly, regional differences were also found as numerous indigenous populations existed for some time prior to the arrival of the Europeans. The medical implications of these data remain to be determined. However, it is anticipated that the data will benefit other Latin American countries as well and will aid in the development of safe and effective drugs for these populations which are not typically studied in clinical trials conducted by United States and European drug makers (Mothelet and Herrera, 2005).

For genomic medicine, the main question is whether race can be used as a surrogate for biological variance to aid prevention, diagnosis and treatment of disease. Studies have detected not only differences in the prevalence of genetic variants (Hall, 1999; Monaghan et al., 1996), but also in the risks associated with the same genetic variant (Helgadottir et al., 2006), implying a role for additional genetic or environmental factors. While studies have shown that individuals from populations around the world can be clustered genetically into six groups (Rosenberg et al., 2002), other studies have demonstrated a range of mixed ancestry in African-Americans, Hispanics, and Mexicans (Parra et al., 1998; Shriver et al., 2003; Sinha et al., 2006).

Our new understanding of human variation and ancestry has yet to impact federal policies regarding the use of race in biomedical research and clinical trials. Current policies such as the National Institutes of Health Revitalization Act (1993), the US Food and Drug Administration (FDA) (1998, 2005a), and the European Medicines Agency (1998, 2005) require that study participants be identified by race and/or ethnicity to ensure a diverse study population and to allow subset analysis. In light of recent findings, however, the application of race and ethnicity categories to biomedical research and the requirement to subset and analyze clinical trials data seem to be outdated and inadequate measures of differences in treatment outcome or response (Haga and Venter, 2003). Although self-identified race has been shown to correspond well with genetic clustering (Tang et al., 2005), it appears to be more problematic clinically since it has not always been shown to be a consistent and reliable measure (Hahn et al., 1996; Johnson, 1974; Rankin and Bhopal, 1999). Moreover, the context in which the self-identification is made may influence an individual's decision, resulting in variable responses depending on who is asking the question and how the information might be used (Senior and Bhopal, 1994).

In using race as an analytical variable, it is important to consider the two major constructs of race: the use of race as a *social* measure to detect differences of potentially important factors such as access to healthcare, environmental exposures, and lifestyle; or the use of race as a *biological* variable to identify genetic (and therefore, biological) differences between populations underlying differences in disease prevalence, severity, and outcome. While there is little disagreement that genetic epidemiology studies would benefit from more diverse sample populations (Ioannidis et al., 2004; Tang, 2006), which may be achieved through using race as a surrogate, it is still unclear what the significance (social and/or biological) of a positive association is between a given phenotype and race. Thus, difficulty in defining the impact of different factors makes race a poor surrogate in general, and direct measurement of relevant factors (e.g., diet, environment) may reduce some of the challenges arising from the use of such a broad undefined variable such as race. Genetics and genomics present an alternative measure of human variation for which race has been used as a surrogate that will lead to more definitive and quantifiable data points (Shields et al., 2005).

Another major issue is the relationship between genetic and genomic data and health disparities. Differences in disease prevalence and outcome between groups, typically characterized by race, have been well-documented (Institute of Medicine, 2003). If individual differences in disease predisposition, prognosis or response to drugs may be accounted for by genetic variation, could group health differences also be accounted for by genetic variation between populations? Despite the small role that genetics is believed to play regarding disease prevalence and outcomes of common diseases, there is an inherent danger of extrapolating differences detected in individuals from one group to an entire race, potentially resulting in unvalidated support of race as a biological category (Wiegmann 2006). Indeed, the temptation to attribute health disparities to genetic causes may distract from other types of social and biomedical research that may lead to greater reduction in these observed differences (Sankar et al., 2004). Furthermore, race-based medicines such as BiDil, a drug approved by the FDA for the treatment of heart failure in self-identified black patients, have the added potential of reifying the biological concept of race and hinder efforts to reduce disparities (Haga and Ginsburg, 2006; Kahn, 2005).

POLICY ISSUES IN LARGE-SCALE GENETICS AND GENOMICS RESEARCH

The shift to study the genetic etiology of common, complex diseases and the development of high-throughput technologies such as microarray analysis and new sequencing techniques at increasingly cheaper costs have enabled the expansion of study populations in genomic studies. Whereas genetic studies

traditionally focused on small groups, particularly families, genomics studies are characterized by the collection of genetic data on hundreds or thousands of individuals, healthy and affected. The demand for DNA samples from individuals of various phenotypes has led to the creation of local and national biobanks or biorepositories worldwide. Given the enormity of national biobanks, several policy issues arise, particularly with respect to research policy. Among the many issues that have been debated about biobanks are the scientific need and merit of a national biobank, the cost and feasibility of successfully establishing and operating a biobank, the required infrastructure, accessibility to samples and data, informed consent, intellectual property (IP), privacy and confidentiality of data, and disclosure of research results to participants. While current policies may be applicable to some of these issues as they are not unique to biobanks, new policies may be required to address some of these issues such as data-sharing and data disclosure and to help reassure the public.

Biobanking

Several years prior to the completion of the Human Genome Project, a number of countries contemplated setting up a national research resource or "biobank" that would contain both clinical specimens and data for the purposes of large-scale genetic epidemiology studies. Samples and data would be collected from 60,000 to one million individuals and stored for five to 30 years (Godard et al., 2004).

Given that a national biobank depends on the public's support and participation, engaging the public is critical to its success (Haga and Willard, 2006; Haga and Beskow, 2008). Two approaches have been used to communicate with the public about national biobanks – a communication or partnership approach (Godard et al., 2004). In Iceland, a brief public consultation took place in the form of radio and television programs, town hall meetings, and public surveys. In 1998, the Act on a Health Sector Database was passed, authorizing the Ministry of Health and Social Security in Iceland to license a private firm (deCODE Genetics, Inc.) to establish and maintain the Health Sector Database, which would store the health records of Iceland's 270,000 citizens. Strong criticism from researchers and ethicists worldwide quickly followed, particularly with respect to the issue of "presumed consent." In response, an option to decline was made available and about 10% of the Icelandic population chose to opt out.

In contrast, the United Kingdom embarked on a massive public 3-year consultation campaign about its proposed national databank. The national diallog included town hall meetings, focus groups, interactive workshops, and published reports. The exchange of information helped inform policy-making to ensure that public concerns were addressed while at the same time raised public awareness about the project. In addition, public comments were requested on procedural and governance documents.

Other countries such as Japan, Canada, Estonia, United States, Tonga, and Latvia are considering or have already established national biobanks. Their approach to public consultation has fallen in between Iceland's communication approach and the UK's partnership approach. The different approaches utilized are reflective of different cultures, attitudes towards scientific research, history, and government and healthcare systems.

Data Disclosure

Research results can be disclosed in a variety of formats including individual or aggregate reports. Aggregate reports would be the only way to inform participants about study outcomes if the samples are collected anonymously or pooled. However, aggregate reports may dilute the value of results to individuals because participants will be left with uncertainty regarding the significance of the findings for themselves specifically (e.g., whether they were found to be at high risk versus low risk) (Shalowitz and Miller, 2005).

A number of studies have examined the attitudes and responses of research participants with respect to access to clinical research results, though not specifically to results from genetic studies. In general, studies have found that disclosure of research results to participants, even those of a serious nature, do not result in negative psychological impacts in the majority of research participants (Bunin et al., 1996; Schulz et al., 2003; Snowdon et al., 1998; Shalowitz and Miller, 2008). Participants indicated they would want access to research results (Wendler and Pentz, 2007), even if upsetting (Schulz et al., 2003) or if not considered clinically valid or useful (Wendler and Emanuel, 2002).

Despite current evidence that research participants would like the option of requesting research results even if of little clinical utility, the practice does appear to be common. An analysis of consent forms in the Children's Oncology Group found that only two of 202 studies offered the option to participants to receive research results and only 10 of 202 studies offered to provide new information after the study was completed (Fernandez et al., 2003a). Although almost 80% of clinical investigators surveyed agreed that clinical trial results should be offered to research participants, about 60% of investigators only offered results less than 20% of the time (Partridge et al., 2004). Moreover, more than 60% did not believe that the disclosure of research results would have a negative impact on participants (Partridge et al., 2004). The major reasons given for not returning results include potential harm of research participants and time constraints and cost related to development of research summaries or individual reports and re-contacting of participants (Fernandez et al., 2003b; Rigby and Fernandez, 2005).

While several policies on data disclosure have been developed, few exist at the federal level. Of the guidelines specific to genetic studies, the importance of clinical relevance and utility is emphasized. For example, the NHLBI guidelines identified the availability of a clinical intervention (treatment or preventative) as a key criterion if results were to be returned to study participants (Bookman et al., 2006). It has been suggested that investigators have an ethical responsibility to offer participants access to research results (MacNeil and Fernandez, 2006). Providing an option to disclose research data shows respect to the research

participant (Shalowitz and Miller, 2005). Those investigators who choose not to provide access to research data should have to justify this decision (Shalowitz and Miller, 2005). Given concerns about the potential harms of unvalidated or inconclusive data, it has been suggested that the option for access to research results should be provided but not necessarily encouraged and that the uncertain nature of the results may be emphasized (Shalowitz and Miller, 2005). Because of these potential risks, it has also been suggested that a second consent should be obtained prior to disclosure of research results (Fernandez et al., 2003c, d).

Data-sharing

Publicly accessible genomic databases employ a range of data access policies. Initially, the leaders of the Human Genome Project determined that all data generated from the project should be placed in a public repository and accessible to all interested users (The Wellcome Trust, 1996, 2003). Rapid deposit of sequence data was required from each of the sequencing centers. This policy decision was influenced by concerns that access to basic data would be impeded if traditional methods of scientific data disclosure were followed. For new publicly funded, large-scale international collaborations, similar policies may need to be developed in advance to address issues of data-sharing, intellectual property and use of specimens (if applicable) (Chokshi et al., 1999). The 2006 NIH proposal specifying policies on data-sharing for NIH-supported or conducted genome-wide association studies is a step in this direction (National Institutes of Health, 2006a).

Several studies have examined data-sharing and data-withholding practices of US biomedical scientists and, in particular, geneticists. According to the most recent study, geneticists were more likely to engage in data-withholding practices than other life scientists (Blumenthal et al., 2006). Among academic geneticists, at least one-third of requests for additional information, data, or materials and 10% of post-publication requests were denied (Campbell et al., 2002). Twenty-eight percent stated that data-withholding impacted their ability to confirm research findings (Campbell et al., 2002). Reasons for withholding included concerns about additional time and effort, protection of junior investigators, and protection of data intended for future publications (Campbell et al., 2002). Delay in publication may be caused by factors such as IP or desire to maintain a competitive advantage (Blumenthal et al., 1997). These three surveys combined have assessed data-sharing practices from the 1980's until today and are among the few sources of information about these scientific practices in genetics or, by extension, genomics.

Little research has been conducted to assess public attitudes regarding sharing of biomedical research data and specifically genomic research data. The privacy of genomic information, even if de-identified or stored anonymously, is tenuous, since as few as 30 single nucleotide polymorphisms (SNP) may uniquely identify an individual (Lin et al., 2004). Therefore, participants may need to be informed about the potential risks due to data-sharing practices.

INTEGRATING GENOMIC MEDICINE APPLICATIONS IN HEALTHCARE

While human genetics has focused traditionally on testing for Mendelian and rare diseases through established techniques such as karyotyping, fluorescence *in situ* hybridization, and restriction fragment length polymorphism analysis, genomic medicine builds on new technologies such as genome-wide analysis of SNP and copy number variation, metabolomics and whole genome sequencing. These new technologies have led to testing for a broader range of diseases including common, complex conditions such as Type II diabetes. However, the expanded scope of clinical genomic applications necessitates consideration of the adequacy of current policies to ensure the safe, effective, and appropriate use of the genomic applications. In particular, changes may be need to be made to the regulatory system, education of health professionals involved in the provision of these new applications, and patient protections.

Oversight

In the United States, it is difficult to determine the actual number of traditional genetic tests performed annually given the privatized health care system; the only published survey data estimated that over 175,000 tests were performed in 1996 (McGovern et al., 1999). In contrast, the Clinical Molecular Genetics Society conducts annual audits of the UK National Health Service laboratories. In 2005–2006, 78,600 post-natal genetic tests were performed, almost 20% more than in 2004–2005 (Clinical Molecular Genetics Society, 2004–2005, 2005–2006). More than 1500 prenatal genetic tests were also performed, representing a 12% increase from the previous year (Clinical Molecular Genetics Society, 2004–2005, 2005–2006). Of particular interest, the number of predictive and confirmatory tests almost doubled over the prior year. About 85 conditions were tested for overall, although the majority of tests were for familial cancers such as breast cancer (*BRCA1/2*) and colorectal cancer (HNPCC and FAP). In the European Union, it is estimated that more than 700,000 tests are performed annually (Ibarreta et al., 2003).

In the United States, the Clinical Laboratory Improvement Amendments (CLIA) program administered by the Centers for Medicare and Medicaid oversees laboratory operations including quality control, quality assurance and analytical test validity. Laboratories are subject to inspection every two years by a state CLIA inspector or an accredited organization such as the College of American Pathologists. In 2006, 198,000 laboratories were registered by the CLIA program. Tests that are packaged and marketed as kits are considered medical devices and require clearance or approval by the FDA. However, as the majority of genetic tests are offered as a clinical laboratory service, otherwise known as a "home-brew" test, they are not currently subject to FDA regulations. In addition to analytical validity, FDA also reviews the clinical validity of tests.

Over the past decade, two federal committees have explored the adequacy of current regulatory oversight mechanisms of genetic tests. In 1997, a NIH-Department of Energy Task Force indicated that an assessment of current review processes of the validity and utility of genetic tests may be warranted, but stopped short of providing specific suggestions (Holtzman and Watson, 1997). In 2000, the Secretary's Advisory Committee on Genetic Testing recommended that the FDA regulate all genetic tests and that oversight of genetic testing laboratories under the CLIA be strengthened (Secretary's Advisory Committee on Genetic Testing, 2000).

To date, these recommendations have not been implemented, although the creation of a genetic testing specialty under CLIA has been proposed and debated (Centers for Disease Control and Prevention, 2000). Several guidance documents have been issued by FDA, resulting in a gradual increase in oversight of new genomics applications. In 2005, a guidance document was issued encouraging sponsors to voluntarily submit pharmacogenomics data to help prepare reviewers on the type and complexity of genomics-based data (FDA, 2005b). Two relevant draft guidances were also released in 2005 – a draft guidance on nucleic acid-based *in vitro* diagnostic devices for detection of microbial pathogens was released (FDA, 2005f) and another on the co-development process of drug-device combination products as evident from their concept paper (FDA, 2005e). In addition, other guidance documents have been developed focusing on drug metabolizing enzymes genotyping system (FDA, 2005c), multiplex technology (FDA, 2005d), analyte specific reagents (FDA, 2006b), algorithm-based genetic tests (FDA, 2006c), genetic toxicology studies (FDA, 2006d), and pharmacogenetic tests (FDA, 2007).

Direct-to-Consumer Marketing

Genetic services and applications are often promoted directly to consumers, similar to drugs and other healthcare services. However, for many of these services and products, the involvement of a health care professional is not required, and consumers can purchase goods directly from the laboratory or manufacturer (an "over-the-counter" test). Laboratory services include testing for paternity, lifestyle, health, and ancestry. Products include DNA-tailored cosmetics and nutritional supplements. Consumers who purchase these services may decline to share the results with their health practitioner due to fear of genetic discrimination.

Direct-to-consumer (DTC) provision of genetic services and products, particularly those that are not health-related, represent an emergent industry. However, at this early stage of research and development, the complex genetic etiology and role of environmental factors for many phenotypes and traits are not clearly understood. Therefore, many of the early tests promoted directly to consumers likely lack robust validity and utility. As these tests are not subject to FDA oversight, the evaluation of the validity of genetic tests and products may be challenging for some consumers. Consumers may be vulnerable to ambiguous and exaggerated claims. In addition, the complex nature of genetic information related to health may warrant consultation with professionals trained in genetics or other specialist to ensure appropriate test interpretation and intervention.

General advertising of genetic tests accessible only through a physician has also increased. For example, Myriad Genetics launched an advertising campaign in two cities in 2002 for the *BRCA1/2* breast cancer tests. A study conducted the following year to assess the impact of the campaign on women's awareness of the tests found that consumers were substantially more likely to have heard of the test and seen an advertisement (Jacobellis et al., 2004). However, despite the increased awareness, the perceived knowledge and interest in testing did not differ between the pilot and control cities. Physician knowledge of breast cancer testing did not differ between the pilot and control cities, although physicians reported ordering a significantly higher number of tests in the pilot cities (Myers et al., 2006).

DTC genetic services and products have raised concerns among a number of groups. In 2002, the UK Health and Science Ministers requested an investigation by the Human Genetics Commission into DTC genetic testing. The commission identified two potential harms from DTC genetic testing: the impact of inaccurate or misinterpreted test results on consumer health and the absence of an adequate informed consent process (UK Human Genetics Commission, 2003). As a result, the Commission recommended that stricter controls be put in place for DTC genetic tests and that predictive genetic tests should be restricted from DTC sale.

In the United States, the American College of Medical Genetics (2004 and 2008) recommends "a knowledgeable healthcare professional should be involved in the process of ordering and interpreting a genetic test" to prevent potential harms such as test misinterpretation and inappropriate follow-up. In 2006, an investigation conducted by the General Accounting Office concluded that DTC nutrigenetic tests were medically unproven and ambiguous (Government Accountability Office, 2006). To educate consumers about home genetic test kits, the Federal Trade Commission, the Centers for Disease Control and Prevention, and the Food and Drug Administration developed a fact sheet (Federal Trade Commission, 2006). And, in 2007, the American Medical Association's House of Delegates passed Resolution 522, which requested further study of the practice of DTC advertising of genetic tests and, in particular, at the existing oversight of this field (American Medical Association, 2007).

Access/Reimbursement

Coverage of genomic applications is intimately linked to uptake of a new test and integration into clinical practice. The demonstration of the validity and utility of a genomic application as well as cost-effectiveness is essential to securing reimbursement. More than 150 economic analyses of genetic services have been conducted, many for adult conditions such as cancer (Carlson et al., 2005; Phillips and Van Bebber, 2004). While many tests have been shown to be cost-effective, others still have questionable utility despite demonstrated clinical validity.

For example, a commercially available molecular profiling test known as Allomap has been shown to be cost-effective to identify heart transplant patients at risk for rejection compared to endomyocardial biopsy to detect rejection, mainly due to the cost-savings gained from not having to perform a biopsy (Evans et al., 2005). But several tests developed to guide therapy decisions are considered by some to be of questionable value (Matchar et al., 2006).

The increased involvement of FDA review of genomic tests and more focused translational research has encouraged the coverage of these applications. For example, in 2005, the Blue Cross Blue Shield Association's Technology Evaluation Center (TEC) assessed four gene expression profiles for use in the management of breast cancer. Based on the available evidence at the time, the report concluded that none of the profiles met the TEC criteria (Blue Cross Blue Shield Association Technology Evaluation Center, 2005). The TEC criteria include approval by federal regulatory bodies and evidence demonstrating an improved outcome and comparable benefits to existing alternative treatments or applications. However, in 2007, a second review of the Oncotype DX test (one of the four tests evaluated in 2005) concluded that the test now met the TEC criteria and was considered to be useful specifically regarding adjuvant chemotherapy for women with estrogen receptor-positive, node-negative, tamoxifen-treated breast cancer (Blue Cross Blue Shield Association Technology Evaluation Center, 2007).

In addition to evaluating both clinical validity and utility, pharmacogenetic testing policies must also consider coverage of the drugs identified as being most likely to perform well or least likely to cause adverse side effects. For tests that are co-developed with drugs, the cost of the testing and drug could be determined by a single policy decision. But in the event that a test indicates a different drug, payors may be forced to revise their coverage policies for drugs that are not listed on their formularies.

Health Professional Education

As the applications of genetic and genomic information have expanded beyond the traditional specialties of medical genetics, pediatrics, and obstetrics into other fields such as oncology, cardiology, neurology, and psychiatry, the education of health professionals needs to evolve as well. While a number of surveys have documented the variable level of physician knowledge of genetics (Metcalfe et al., 2002), none has assessed knowledge of the newer field of genomics. The uptake and use of new applications will be stalled until health professionals gain some understanding about the appropriate use of these tests and the interpretation and application of test results. As a result, the continuum of health education from graduate or professional school to continuing education should aim to increase awareness and understanding of these new tools and therapies across virtually all medical specialties (Challen et al., 2005; Gurwitz et al., 2003). In addition, it will be important for health professionals to understand other issues related to the use of these technologies such as implications for family members and privacy and confidentiality issues.

Re-defining Roles of Health Professionals

Due to the expected widespread use of genomic applications, some re-organization of health professional roles may be required, given the limited number of genetics specialists and new applications such as pharmacogenetic testing (see Table 18.1) (Guttmacher et al., 2001). For example, the pharmacist may play a much greater role in the integration of pharmacogenetic testing by helping to ensure that the drug and dosage are safe based on the patient's genotype (Clemerson et al., 2006). Monitoring pharmacogenetic information to assure appropriate drug dosing would appear to be a natural extension of the role of pharmacists. To fulfill this expanded role, pharmacists would need greater access to patient medical records (Clemerson et al., 2006), although physicians and patients have expressed concerns regarding patient confidentiality if full access were permitted (Porteous et al., 2003).

In contrast, a board-certified genetics professional (physician or genetic counselor) may have a revised, lesser or even non-existent role in the provision of genomic testing such as a pharmacogenetic profile as these tests become routine. The need for a genetics specialist may be determined by several factors including complexity in determining appropriate candidates for testing and interpretation of test results, the extent of the informed consent process, reimbursement policies, familial implications and the risks and benefits of the test. However, given the relatively small size of the genetics workforce, any new requirements for their services may not be feasible. Consultation by a board-certified genetics specialist may need to be reserved for only the most complicated cases.

To increase the number of physicians with training in genetics or genomics, one possibility may be to increase the number of dual board-certified specialties with medical genetics. For example, an individual dual-certified in medical genetics and pharmacology would have expertise in genetics and pharmacology and be able to advise the treating physician about the best choice of treatment based on the patient's pharmacogenetic profile and other clinical information. However, the training requirements to fulfill the medical genetics certification should be substantially revised to provide direct knowledge relevant to the individual's primary profession with greater focus on test appropriateness, interpretation and test utility. In addition, genetics certification should not lengthen the overall training program.

Pilot development studies have found that new roles associated with genetic testing were mostly fulfilled by non-genetics specialists, in particular nurses (Bennett et al., 2007). For example, the cancer family nurse specialist is trained to identify patients with a family history of cancer that would place them at increased risk and indicate testing (Bennett et al., 2007). Indeed, genetic competencies have been developed and training programs have been accordingly revised to meet these new needs and to enable nurses to adequately provide genetics services and guidance (Jenkins and

Table 18.1 Changing roles of health professionals in genomic medicine

Health professional	New or revised role
Primary care Physician	Increased focus on family-history assessment, identification of appropriate patients for testing, discussion of risks and benefits of testing, interpretation of test results, revision of medical management based on test results, focus on preventative strategies
Nurse	Enhanced family-history taking and counseling skills; enhanced education in corresponding areas of genetics (e.g., cancer genetics for oncology nurses); awareness of ethical, social, and legal issues associated with testing
Genetic or Genomic Specialist (Genetic Counselor or Physician)	Consultation reserved for complex genetic cases; interpretation of whole-genome data, including sequence; increased role in professional education development
Pharmacist	Increased understanding of genetic factors in drug safety and efficacy; enhanced education of genetic etiology of drug response; consideration of pharmacogenetic test results prior to drug treatment; increased collaboration with prescribing physician; increased patient education with respect to pharmacogenetic testing; added consideration of pharmacogenetic test result to determine appropriate dose adjustment or drug selection

Calzone, 2007; Lewis et al., 2006). Studies have demonstrated the equivalence of care provided by nurses trained in genetics compared to board-certified genetic counselors (Torrance et al., 2006). Several groups have developed professional genetic competencies including the National Coalition for Health Professional Education in Genetics (2005), and the National Health Service's National Genetics Education and Development Centre and the Skills for Health Service (2007) have developed general competences for non-genetics practitioners.

Privacy and Confidentiality

The issue of medical privacy is a major concern in the United States. In 2005, the National Consumer Health Privacy Survey reported that 66% of Americans were very or somewhat concerned about medical privacy (California Health Care Foundation, 2005). Due to concerns that employers or other groups may have access to a medical record, 11% of survey respondents indicated paying for a cancer test or procedure out-of-pocket. The predictive nature and often deterministic views associated with genetic information have likely contributed to the heightened concerns linked to testing. In addition, as the number of health professionals with access to a patient's genomic test results increases, the ability to safeguard the information increases becomes more challenging.

Genetic Discrimination

Genetic discrimination involves the inappropriate use of genetic information, primarily with respect to health insurance and employment, but including other areas such as life insurance, long-term care insurance, adoption services, egg and sperm banks or educational admissions programs. While the fear of genetic discrimination has existed prior to the start of the Human Genome Project, the solutions have been slow to come by and, at best, provide a patchwork of protections in the United States and around the world. For example, about two-thirds of states have enacted legislation to prohibit or limit the use of genetic information for risk selection and risk classification for health insurance purposes. Passage of national legislation would provide universal coverage against health insurance and employment discrimination on the basis of genetic information for all. But despite all of the interest and support shown by federal lawmakers including the past two US presidents, national legislation has only just been enacted. Unbelievably, it has taken longer to pass genetic anti-discrimination legislation than it actually took to sequence the human genome.

Since the first bill on genetic discrimination in health insurance was introduced in the United States in 1995, at least 20 other bills have been introduced, but only one has passed providing partial protection against genetic discrimination for the group insurance market. In 2007, the Genetic Information Nondiscrimination Act was re-introduced to "prohibit discrimination on the basis of genetic information with respect to health insurance and employment." In contrast to previous years, it swiftly moved through the three House committees who have jurisdiction and passed by a floor vote of 420 to 3, thus moving further along than in any previous Congress. The bill was also passed by the Senate and signed into law by President Bush in May 2008.

Despite universal condemnation against genetic discrimination (UNESCO, 1997, 2003), the development and implementation of national protections has also been slow outside the United States. For example, although "genetic features" is included in the European Union's Charter of Fundamental Rights' section on non-discrimination, countries have had variable success in providing national protections (European Parliament, 2000). Austria, Belgium, France, the Netherlands, Luxembourg, Greece, and Italy have enacted legislation prohibiting access of genetic information without consent.

Various initiatives have been undertaken in the United Kingdom, Australia, and Canada to protect against genetic

discrimination by life insurers and employers. Similar to the United States, a patchwork of provincial legislative protections in Canada exist. A recent review of the Canadian Human Right Act by the Department of Justice recommended that the definition of disability include genetic predispositions (Department of Justice, Canada, 2005). Australia has been particularly active in investigating the protections of genetic information. In 2000, the Attorney-General of Australia and the Minister for Health and Aged Care called for an inquiry into genetic discrimination. An extensive public consultation was undertaken including widespread dissemination of two consultation papers (Australian Law Reform Commission, 2001, 2002), 15 public forums around the country, and 185 meetings with key stakeholders and interested parties. In 2003, a final report was released containing 144 recommendations directed at local, regional, and national governments as well as statutory authorities (Australian Law Reform Commission, 2003). And in the United Kingdom, a moratorium is in effect banning the use of genetic testing information for insurance underwriting purposes until 2014 (Association of British Insurers, 2008).

Intellectual Property

Intellectual property (IP) of genes and genetic material has been a major point of controversy since the early years of the Human Genome Project. Nearly 20% of the human genome is under US patents, comprising about 4400 genes (Jensen and Murray, 2005). More than 60% of these patents are owned by private entities (Jensen and Murray, 2005). In addition, almost a dozen complete genomes have been patented or are pending patent approval (O'Malley et al., 2005). A constellation of issues pertaining to ownership has garnered significant attention and concern from various stakeholders. From the single laboratory to national advisory bodies to international organizations, gene patenting has become a controversial topic, particularly with respect to the limitation of access to gene-related inventions and benefit-sharing (World Health Organization, 2005; Canadian Biotechnology Advisory Committee, 2006). The patenting of genomes has significantly broadened the scope of IP, and its impact remains to be seen.

Although there has been a significant amount of debate about the patenting of genes, concern also exists about the licensing practices of patent holders, as exemplified by *BRCA1/BRCA2* and Canavan's disease. To address some of these concerns, the NIH has also taken steps to promote open access to its large-genomic datasets and encourage fair licensing practices through IP user agreements and guidance documents (National Institutes of Health, 2005). In addition, an amendment to a Request for Applications for genome sequencing centers requires applicants to develop an "IP Management Plan" to ensure wide availability of sequencing data generated by awardees (National Institutes of Health, 2006b).

2009 UPDATE

While advances stemming from the genome sciences are certain to impact the practice of medicine and our understanding of human development, biology, biochemistry, and physiology, the safe and appropriate use of this new information and its associated technologies will warrant new policies and an improved level of understanding by patients, consumers, and providers. Over the last year, direct-to-consumer testing has substantially expanded with at least three companies providing whole-genome risk profiling, ranging in cost from $399 to $2500 for risk information for several dozen (Navigenics) to nearly a hundred (23andMe, Inc.) diseases and traits (ironically, the lower priced test reports risk information for the greatest number of conditions). Although these companies claim not to be offering medical information, some health departments, notably in California, have demanded that the laboratories be CLIA-certified and retain professional staff or cease marketing their products within these states. While many of these companies have acquiesced to these requests, they and other genetic testing providers will likely face additional demands for oversight, particularly test standardization and validation, and have begun working with other groups to define these standards (DHHS, 2008a; Kaye, 2008). Thus, although the commercial success of these companies remains to be seen, they have pushed to the forefront many of the issues that have been facing genetic, and now genomic, testing providers for some time, namely oversight and regulation. In addition, these companies have spurred physicians and public health experts to highlight the need for additional research to validate genetic associations (Khoury et al., 2008) and to increase physician training and awareness in preparation of increased testing outside of the healthcare system (Haga and Willard, 2008).

With passage of the genetic non-discrimination protections, federal agencies in the United States have begun to work on policies to implement the new law, which will become effective in 2009. Although the signing of this legislation took more than a decade to achieve, there are gaps in the legislation as it does not prohibit life insurers and long-term care insurers from using genetic information in their policy decisions (Hudson et al., 2008). Although many states have laws prohibiting genetic discrimination, some more stringent than the federal legislation, it will be of interest to assess the impact of passage of federal protections on patient concerns and whether it substantially alleviates this fear and removes this barrier from the decision-making process. Education of health professionals, patients and consumers about the new federal protections will be essential part of realizing the benefits.

Personalized medicine will likely remain a top health initiative in the new administration in the United States. In 2007, then-Senator Barack Obama introduced a bill to facilitate oversight and development of personalized medicine tools

(S. 976). In 2008, a companion bill (H.R.6498) was introduced in the House of Representatives by Congressman Patrick Kennedy. Other government priorities, such as family-history taking tools (DHHS, 2009) and development of decision support tools (DHHS, 2008b), are key components in initiatives on health information technology and personalized medicine, as they will help facilitate uptake of new genomic-based applications and the appropriate use of these applications.

2009 UPDATE REFERENCES

Department of Health and Human Services, U.S. (2008a). U.S. System of Oversight of Genetic Testing: A response to the charge of the secretary of health and human services. Report of the Secretary's Advisory Committee on Genetics, Health, and Society. April 2008. http://oba.od.nih.gov/oba/SACGHS/reports/SACGHS_oversight_report.pdf

Department of Health and Human Services, U.S. (2008b). Personalized health care: Pioneers, partnerships, and progress. http://www.hhs.gov/myhealthcare/news/phc_2008_report.pdf

Department of Health and Human Services, U.S. (2009). Surgeon General's new family health history tool is released, ready for "21st century medicine" (Press Release). http://www.hhs.gov/news/press/2009pres/01/20090113a.html

Haga, S.B. and Willard, H.F. (2008). Letting the genome out of the bottle. *N Engl J Med* 358(20), 2184.

Hudson, K.L., Holohan, M.K. and Collins, F.S. (2008). Keeping pace with the times – The Genetic Information Nondiscrimination Act of 2008. *N Engl J Med* 358(25), 2661–2663.

Kaye, J. (2008). The regulation of direct-to-consumer genetic tests. *Hum Mol Genet* 17(R2), R180–R183.

Khoury, M.J., Berg, A., Coates, R., Evans, J., Teutsch, S.M. and Bradley, L.A. (2008). The evidence dilemma in genomic medicine. *Health Aff (Millwood)* 27(6), 1600–1611.

REFERENCES

American College of Medical Genetics Board of Directors (2008). ACMG Statement on Direct-to-Consumer Genetic Testing. Available at http://www.acmg.net/StaticContent/StaticPages/DTC_Statement.pdf.

American College of Medical Genetics Board of Directors (2008). ACMG Statement on Direct-to-Consumer Genetic Testing. Available at http://www.acmg.net/AM/Template.cfm?Section=Policy_Statements&Template=/CM/ContentDisplay.cfm&ContentID=2975.

American Medical Association (2007). Resolution 522: Direct to Consumer Advertising and Provision of Genetic Testing. Available at http://www.ama-assn.org/ama1/pub/upload/mm/467/522.doc.

Association of British Insurers (2008). Insurance genetics moratorium extended until 2014. Available at http://www.abi.org.uk/Newsreleases/ViewNewsRelease.asp?nrid=16274.

Australian Law Reform Commission and Australian Health Ethics Committee of the National Health and Medical Research Council (2001). *Protection of Human Genetic Information* (IP 26). Available at http://www.austlii.edu.au/au/other/alrc/publications/issues/26/index.html.

Australian Law Reform Commission and Australian Health Ethics Committee of the National Health and Medical Research Council (2002). *Protection of Human Genetic Information* (Discussion paper 66). Available at http://www.austlii.edu.au/au/other/alrc/publications/dp/66/.

Australian Law Reform Commission and Australian Health Ethics Committee of the National Health and Medical Research Council (2003). *Essentially Yours: The Protection of Human Genetic Information in Australia* (ALRC 96). Available at http://www.austlii.edu.au/au/other/alrc/publications/reports/96/.

Bennett, C., Burton, H. and Fardon, P. (2007). Competences, education and support for new roles in cancer genetics services: outcomes from the cancer genetics pilot projects. *Fam Cancer* 6, 171–180.

Blue Cross Blue Shield Association. Technology Evaluation Center. Gene Expression Profiling for Managing Breast Cancer Treatment. Assessment Program 2005, 20 (3). Available at http://www.bcbs.com/tec/Vol20/20_03.pdf.

Blue Cross Blue Shield Association's Technology Evaluation Center (2007). Actions Taken by the Medical Advisory Panel. Available at http://www.bcbs.com/betterknowledge/tec/press/actions-taken-by-the-medical-1.html.

Blumenthal, D., Campbell, E.G., Anderson, M.S., Causino, N. and Louis, K.S. (1997). Withholding of research results in academic life science: Evidence from a national survey of faculty. *JAMA* 277, 1224–1228.

Blumenthal, D., Campbell, E.G., Gokhale, M., Yucel, R., Clarridge, B., Hilgartner, S. and Holtzman, N.A. (2006). Data withholding in genetics and the other life sciences: Prevalences and predictors. *Acad Med* 81(2), 137–145.

Bookman, E.B., Langehorne, A.A., Eckfeldt, J.H., Glass, K.C., Jarvik, G.P., Klag, M., Koski, G., Motulsky, A., Wilfond, B., Manolio, T.A. et al. (2006). Reporting genetic results in research studies: Summary and recommendations of an NHLBI working group. *Am J Med Genet A* 140A, 1033–1040.

Bunin, G.R., Kazak, A.E. and Mitelman, O. (1996). Informing subjects of epidemiologic study results. Children's Cancer Group. *Pediatrics* 97(4), 486–491.

Burchard, E.G., Ziv, E., Coyle, N., Gomez, S.L., Tang, H., Karter, A.J., Mountain, J.L., Perez-Stable, E.J., Sheppard, D. and Risch, N. (2003). The importance of race and ethnic background in biomedical research and clinical practice. *N Engl J Med* 348, 1170.

California Health Care Foundation and Forrester Research, Inc. (2005). National Consumer Health Privacy Survey. Available at http://www.chcf.org/documents/ihealth/ConsumerPrivacy2005Slides.pdf.

Campbell, E.G., Clarridge, B.R., Gokhale, M., Birenbaum, L., Hilgartner, S., Holtzman, N.A. and Blumenthal, D. (2002). Data-withholding in academic genetics: Evidence from a national survey. *JAMA* 287, 473–481.

Canadian Biotechnology Advisory Committee (2006). *Human Genetic Materials, Intellectual Property and the Health Sector.* Available at http://strategis.ic.gc.ca/epic/internet/incbac-cccb.nsf/en/ah00578e.html.

Carlson, J.J., Henrikson, N.B., Veenstra, D.L. and Ramsey, S.D. (2005). Economic analyses of human genetics services: A systematic review. *Genet Med* 7, 519–523.

Centers for Disease Control and Prevention. Notice of Intent: Genetic testing under the Clinical Laboratory Improvement Amendments. 65 Fed Reg 25928, May 4, 2000.

Challen, K., Harris, H.J., Julian-Reynier, C., Ten Kate, L.P., Kristoffersson, U., Nippert, I., Schmidtke, J., Benjamin, C. and Harris, R.GenEd Research Group (2005). Genetic education and non genetic health professionals: Educational providers and curricula in Europe. *Genet Med* 7(5), 302–310.

Chokshi, D.A., Parker, M. and Kwiatkowski, D.P. (1999). Data sharing and intellectual property in a genomic epidemiology network: Policies for large-scale research collaboration. *Bull World Health Organ* 84(5), 382–387.

Clemerson, J.P., Payne, K., Bissell, P. and Anderson, C. (2006). Pharmacogenetics, the next challenge for pharmacy?. *Pharm World Sci* 28(3), 126–130.

Clinical Molecular Genetics Society. Final 2004–2005 Audit Report. Available at http://www.cmgs.org/new_cmgs/Laboratory%20Information/FinalCMGSAudit0405.pdf.

Clinical Molecular Genetics Society. Final 2005–2006 Audit Report. Available at http://www.cmgs.org/new_cmgs/Laboratory%20Information/FinalCMGSAudit0506.pdf.

Collins, F.S. (May 22, 2003). Testimony Before the Subcommittee on Health Committee on Energy and Commerce, United States House of Representatives. Available at http://www.genome.gov/11007447.

Collins, F. and Galas, D. (1993). A new five-year plan for the U.S. Human Genome Project. *Science* 262, 43–46.

Collins, F.S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R. and Walters, L. (1998). New goals for the U.S. Human Genome Project: 1998–2003. *Science* 282, 682–689.

Collins, F.S., Green, E.D., Guttmacher, A.E. and Guyer, M.S. (2003). U.S. National Human Genome Research Institute. A vision for the future of genomics research. *Nature* 422, 835–847.

Cooper, R.S., Kaufman, J.S. and Ward, R. (2003). Race and genomics. *N Engl J Med* 348, 1166.

European Commission (2002). The Sixth Framework Programme in Brief. Available at http://ec.europa.eu/research/fp6/pdf/fp6-in-brief_en.pdf.

European Commission (2006). FP7 – Tomorrow's Answers Start Today. Available at http://ec.europa.eu/research/fp7/pdf/fp7-factsheets_ en.pdf.

Department of Justice, Canada (2005). Promoting Equality: A New Vision. [online] http://canada.justice.gc.ca/chra/en/toc.html.

European Medicines Agency (1998). ICH Topic E 5 (R1) Ethnic Factors in the Acceptability of Foreign Clinical Data. Available at http://www.emea.europa.eu/pdfs/human/ich/028995en.pdf.

European Medicines Agency (2005). Guideline on Risk Management Systems for Medicinal Products for Human Use. Available at http://www.emea.europa.eu/pdfs/human/euleg/9626805en.pdf.

European Parliament (2000). Charter of fundamental rights of the European Union. Article 21 (Non-discrimination). Available at http://www.europarl.europa.eu/charter/pdf/text_en.pdf.

Evans, R.W., Williams, G.E., Baron, H.M., Deng, M.C., Eisen, H.J., Hunt, S.A., Khan, M.M., Kobashigawa, J.A., Marton, E.N., Mehra, M.R. et al. (2005). The economic implications of noninvasive molecular testing for cardiac allograft rejection. *Am J Transplant* 5, 1553–1558.

Federal Trade Commission (2006). Facts for Consumers: At-Home Genetic Tests: A Healthy Dose of Skepticism May Be the Best Prescription. Available at http://www.ftc.gov/bcp/edu/pubs/consumer/health/hea02.htm.

Fernandez, C.V., Kodish, E., Taweel, S., Shurin, S. and Weijer, C. (2003a). Disclosure of the right of research participants to receive research results: An analysis of consent forms in the Children's Oncology Group. *Cancer* 97, 2904–2909.

Fernandez, C.V., Kodish, E., Shurin, S. and Weijer, C.Children's Oncology Group (2003b). Offering to return results to research participants: Attitudes and needs of principal investigators in the Children's Oncology Group. *J Pediatr Hematol Oncol* 25, 704–708.

Fernandez, C.V., Kodish, E. and Weijer, C. (2003c). *Informing study participants of research results: An ethical imperative. IRB* 25(3), 12–19.

Fernandez, C.V., Kodish, E. and Weijer, C. (2003d). Importance of informed consent in offering to return research results to research participants. *Med Pediatr Oncol* 41, 592–593.

Food and Drug Administration (FDA), "Investigational New Drug Applications and New Drug Applications, Final Rule," *Federal Register*, February 11, 1998; 63(8):6854-62 (codified at 21 CFR 312.33(a)(2) and 21 CFR 314.50(d)(5)).

Food and Drug Administration (FDA). Innovation or Stagnation? Challenge and Opportunity on the Critical Path to New Medical Products. Available at http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.html (2004).

Food and Drug Administration (FDA) (2005a). Guidance for Industry – Collection of Race and Ethnicity in Clinical Trials. Available at http://www.fda.gov/cber/gdlns/racethclin.pdf.

Food and Drug Administration (FDA) (March 2005b). Guidance for Industry – Pharmacogenomic Data Submissions. Available at http://www.fda.gov/cder/guidance/6400fnl.pdf.

Food and Drug Administration (FDA) (March 2005c). Guidance for Industry and FDA Staff. Class II Special Controls Guidance Document: Drug Metabolizing Enzyme Genotyping System . Available at http://www.fda.gov/cdrh/oivd/guidance/1551.pdf.

Food and Drug Administration (FDA) (March 2005d). Guidance for Industry and FDA Staff. Class II Special Controls Guidance Document: Instrumentation for Clinical Multiplex Test Systems. Available at http://www.fda.gov/cdrh/oivd/guidance/1546.pdf.

Food and Drug Administration (FDA) (April 2005e). Draft – Drug-Diagnostic Co-Development Concept Paper. Available at http://www.fda.gov/cder/genomics/pharmacoconceptfn.pdf.

Food and Drug Administration (FDA) (December 28, 2005f). Nucleic Acid Based *In Vitro* Diagnostic Devices for Detection of Microbial Pathogens-Draft Guidance for Industry and FDA Staff. Available at http://www.fda.gov/cdrh/oivd/guidance/1560.pdf.

Food and Drug Administration (FDA) (March 2006a). Innovation or Stagnation? Critical Path Opportunities List. Available at http://www.fda.gov/oc/initiatives/criticalpath/reports/opp_list.pdf.

Food and Drug Administration (FDA) (September 2006b). Draft Guidance for Industry and FDA Staff-Commercially Distributed Analyte Specific Reagents (ASRs): Frequently Asked Questions. Available at http://www.fda.gov/cdrh/oivd/guidance/1590.pdf.

Food and Drug Administration (FDA) (September 2006c). Draft Guidance for Industry, Clinical Laboratories, and FDA Staff-*In Vitro* Diagnostic Multivariate Index Assays. Available at http://www.fda.gov/cdrh/oivd/guidance/1610.pdf.

Food and Drug Administration (FDA) (June 2006d). Guidance for Industry and Review Staff. Recommended Approaches to Integration of Genetic Toxicology Study Results. Available at http://www.fda.gov/cder/guidance/6848fnl.pdf.

Food and Drug Administration (FDA) (June 2007). Guidance for Industry and FDA Staff. Pharmacogenetic Tests and Genetic Tests for Heritable Markers. Available at http://www.fda.gov/cdrh/oivd/guidance/1549.pdf.

Foster, M. (2005). Genetics and social identity after the HapMap. *Nat Rev Genet* 6, 879.

Godard, B., Marshall, J., Laberge, C. and Knoppers, B.M. (2004). Strategies for Consulting with the Community: The cases of four large-scale genetic databases. *Sci Eng Ethics* 10, 457–477.

Government Accountability Office (2006). Nutrigenetic Testing: Tests Purchased from Four Web Sites Mislead Consumers (GAO-06-977T). Available at http://www.gao.gov/new.items/d06977t.pdf.

Gurwitz, D., Weizman, A. and Rehavi, M. (2003). Education: Teaching pharmacogenomics to prepare future physicians and researchers for personalized medicine. *Trends Pharmacol Sci* 24, 122–125.

Guttmacher, A.E., Jenkins, J. and Uhlmann, W.R. (2001). Genomic medicine: Who will practice it? A call to open arms. *Am J Med Genet* 106(3), 216–222.

Haga, S.B. and Beskow, L.M. (2008). Ethical, legal, and social implications of biobanks for genetics research. *Adv Genet* 60, 505–544.

Haga, S.B. and Ginsburg, G.S. (2006). Prescribing BiDil: Is it black and white?. *J Am Coll Cardiol* 48(1), 12–14.

Haga, S.B. and Venter, J.C. (2003). FDA races in wrong direction. *Science* 301, 466.

Haga, S.B. and Willard, H.F. (2006). Defining the spectrum of genome policy. *Nat Rev Genet* 7(12), 966–972.

Hahn, R.A., Truman, B.I. and Barker, N.D. (1996). Identifying ancestry: The reliability of ancestral identification in the United States by self, proxy, interviewer, and funeral director. *Epidemiology* 7, 75–80.

Hall, D., Ybazeta, G., Destro-Bisel, G., Petzl-Erler, M.L. and Di Renzo, A. (1999). Variability at the uridine diphosphate glucuronosyltransferase 1A1 promoter in human populations and primates. *Pharmacogenetics* 9, 591–599.

Helgadottir, A., Manolescu, A., Helgason, A., Thorleifsson, G., Thorsteinsdottir, U., Gudbjartsson, D.F., Gretarsdottir, S., Magnusson, K.P., Gudmundsson, G., Hicks, A. et al. (2006). A variant of the gene encoding leukotriene A4 hydrolase confers ethnicity-specific risk of myocardial infarction. *Nat Genet* 38, 68–74.

Holtzman, N.A. and Watson, M.S. (1997). *Promoting Safe and Effective Genetic Testing in the United States: Final Report of the Task Force on Genetic Testing* (eds). Johns Hopkins University Press, Baltimore.

Ibarreta D, Bock, AK, Klein C, Rodriguez-Cerezo E. Towards quality assurance and harmonization of genetic testing services in the EU. Available at ftp://ftp.jrc.es/pub/EURdoc/eur20977en.pdf (2003).

Institute of Medicine (2003). *Unequal Treatment: Confronting Racial and Ethnic Disparities in Healthcare*. National Academy Press, Washington, DC.

International HapMap Consortium (2005). A haplotype map of the human genome. *Nature* 437, 1229–1320.

Ioannidis, J.P., Ntzani, E.E. and Trikalinos, T.A. (2004). Racial differences in genetic effects for complex diseases. *Nat Genet* 36, 1312–1318.

Jacobellis, J., Martin, L., Engel, J., VanEenwyk, J., Bradley, L.A., Kassim, S., Jorgensen, C., Litch, J.A. and Myers, M.F. (2004). Genetic testing for breast and ovarian cancer susceptibility: Evaluating direct-to-consumer marketing – Atlanta, Denver, Raleigh-Durham, and Seattle, 2003. *MMWR Morb Mortal Wkly Rep* 53, 603–606.

Jenkins, J. and Calzone, K.A. (2007). Establishing the essential nursing competencies for genetics and genomics. *J Nurs Scholarsh* 39(1), 10–16.

Jensen, K. and Murray, F. (2005). Intellectual property landscape of the human genome. *Science* 310, 239–240.

Johnson, C.E. (1974). U.S. Bureau of the Census, Technical Paper No. 31, Washington, D.C.

Kahn, J.D. (2005). From disparity to difference: How race-specific medicines may undermine policies to address inequalities in health care. *So Calif Interdiscipl Law J* 15, 105–129.

Lewis, J.A., Calzone, K.M. and Jenkins, J. (2006). Essential nursing competencies and curricula guidelines for genetics and genomics. *MCN Am J Matern Child Nurs* 31(3), 146–155.

Lin, Z., Owen, A.B. and Altman, R.B. (2004). Genomic research and human subject privacy. *Science* 305, 183.

MacNeil, S.D. and Fernandez, C.V. (2006). Offering results to research participants. *BMJ* 332, 188–189.

Matchar, D.B., Thakur, M.E., Grossman, I., McCrory, D.C., Orlando, L.A., Steffens, D.C., Goldstein, D.B., Cline, K.E., Gray, R.N. (2006). Testing for Cytochrome P450 Polymorphisms in Adults With Non-Psychotic Depression Treated With Selective Serotonin Reuptake Inhibitors (SSRIs). Evidence Report/Technology Assessment No. 146. (Prepared by the Duke Evidence-based Practice Center under Contract No. 290-02-0025.) AHRQ Publication No. 07-E002. Rockville, MD: Agency for Healthcare Research and Quality.

McGovern, M.M., Benach, M.O., Wallenstein, S., Desnick, R.J. and Keenlyside, R. (1999). Quality assurance in molecular genetic testing laboratories. *JAMA* 281, 835–840.

Metcalfe, S., Hurworth, R., Newstead, J. and Robins, R. (2002). Needs assessment study of genetics education for general practitioners in Australia. *Genet Med* 4, 71–77.

Monaghan, G., Ryan, M., Seddon, R., Hume, R. and Burchell, B. (1996). Genetic variation in bilirubin UPD-glucuronosyltransferase gene promoter and Gilbert's syndrome. *Lancet* 347, 578–581.

Mothelet, V.G. and Herrera, S. (2005). Mexico launches bold genome project. *Nat Biotech* 23, 1030.

Myers, M.F., Chang, M.H., Jorgensen, C., Whitworth, W., Kassim, S., Litch, J.A., Armstrong, L., Bernhardt, B., Faucett, W.A., Irwin, D. et al. (2006). Genetic testing for susceptibility to breast and ovarian cancer: Evaluating the impact of a direct-to-consumer marketing campaign on physicians' knowledge and practices. *Genet Med* 8, 361–370.

O'Malley, M.A., Bostanci, A. and Calvert, J. (2005). Whole-genome patenting. *Nat Rev Genet* 6, 502–506.

National Cancer Institute (2006). The TAILORx Breast Cancer Trial. Available at http://www.cancer.gov/clinicaltrials/digestpage/TAILORx.

National Coalition for Health Professional Education in Genetics (2005). Core Competencies in Genetics Essential for All Health-Care Professionals. Available at http://www.nchpeg.org/core/Corecomps2005.pdf.

National Health Service's National Genetics Education and Development Centre and the Skills for Health (2007). Competences for Genetics in Clinical Practice. Available at http://www.geneticseducation.nhs.uk/develop/index.asp?id=44.

National Human Genome Research Institute (2007). Medical Sequencing Program and Current Initiatives. Available at http://www.genome.gov/15014882#2.

National Institutes of Health Revitalization Act. Pub. L. No. 103-43 107 Stat. 122 (1993).

National Institutes of Health (2005). Best Practices for the Licensing of Genomic Inventions: Final Notice. *Fed Regist* 70, 18413–18415.

National Institutes of Health (2006a). Request for Information (RFI): Proposed Policy for Sharing of Data obtained in NIH supported or conducted Genome-Wide Association Studies (GWAS) (NOT-OD-06-094). Available at http://grants.nih.gov/grants/guide/notice-files/NOT-OD-06-094.html.

National Institutes of Health (2006b). Amendments and Corrections to RFA-HG-06-001 "Genome Sequencing Centers." Available at http://grants2.nih.gov/grants/guide/notice-files/NOT-HG-06-010.html.

NIH Chemical Genomics Center (2005). Available at http://www.ncgc.nih.gov/.

Parra, E.J., Marcini, A., Akey, J., Martinson, J., Batzer, M.A., Cooper, R., Forrester, T., Allison, D.B., Deka, R., Ferrell, R.E. et al. (1998). Estimating African American admixture proportions by use of population-specific alleles. *Am J Hum Genet* 63, 1839–1851.

Partridge, A.H., Hackett, N., Blood, E., Gelman, R., Joffe, S., Bauer-Wu, S., Knudsen, K., Emmons, K., Collyar, D., Schilsky, R.L. et al. (2004). Oncology physician and nurse practices and attitudes regarding offering clinical trial results to study participants. *J Natl Cancer Inst* 96(8), 629–632.

Phillips, K.A. and Van Bebber, S.L. (2004). A systematic review of cost-effectiveness analyses of pharmacogenomic interventions. *Pharmacogenomics* 5, 1139–1149.

Porteous, T., Bond, C., Robertson, R., Hannaford, P. and Reiter, E. (2003). Electronic transfer of prescription-related information: comparing views of patients, general practitioners, and pharmacists. *Br J Gen Pract* 53(488), 204–209.

Rankin, J. and Bhopal, R. (1999). Current census categories are not a good match for identity. *BMJ* 318, 1696.

Rigby, H. and Fernandez, C.V. (2005). Providing research results to study participants: Support versus practice of researchers at the American Society of Hematology annual meeting. *Blood* 106, 1199–1202.

Rosenberg, N.A., Pritchard, J.K., Weber, J.L., Cann, H.M., Kidd, K.K., Zhivotovsky, L.A. and Feldman, M.W. (2002). Genetic structure of human populations. *Science* 298, 2381–2385.

Sankar, P., Cho, M.K., Condit, C.M., Hunt, L.M., Koenig, B., Marshall, P., Lee, S.S. and Spicer, P. (2004). Genetic research and health disparities. *JAMA* 291, 2985–2989.

Schulz, C.J., Riddle, M.P., Valdimirsdottir, H.B., Abramson, D.H. and Sklar, C.A. (2003). Impact on survivors of retinoblastoma when informed of study results on risk of second cancers. *Med Pediatr Oncol* 41(1), 36–43.

Secretary's Advisory Committee on Genetic Testing (2000). Enhancing the Oversight of Genetic Testing: Recommendations of the SACGT. Available at http://www4.od.nih.gov/oba/sacgt/reports/oversight_report.htm.

Senior, P.A. and Bhopal, R. (1994). Ethnicity as a variable in epidemiological research. *BMJ* 309, 327–330.

Shalowitz, D.I. and Miller, F.G. (2005). Disclosing individual results of clinical research: Implications of respect for participants. *JAMA* 294, 737–740.

Shalowitz, D.I. and Miller, F.G. (2008). Communicating the results of clinical research to participants: attitudes, practices, and future directions. *PLoS Med* 5(5), e91.

Shields, A.E., Fortun, M., Hammonds, E.M., King, P.A., Lerman, C., Rapp, R. and Sullivan, P.F. (2005). The use of race variables in genetic studies of complex traits and the goal of reducing health disparities: A transdisciplinary perspective. *Am Psychol* 60, 77–103.

Shriver, M.D., Parra, E.J., Dios, S., Bonilla, C., Norton, H., Jovel, C., Pfaff, C., Jones, C., Massac, A., Cameron, N. et al. (2003). Skin pigmentation, biogeographical ancestry and admixture mapping. *Hum Genet* 112, 387–399.

Sinha, M., Larson, E.K., Elston, R.C. and Redline, S. (2006). Self-reported race and genetic admixture. *N Eng J Med* 354, 421–422.

Snowdon, C., Garcia, J. and Elbourne, D. (1998). Reactions of participants to the results of a randomised controlled trial: Exploratory study. *BMJ* 317(7150), 21–26.

Tang, H. (2006). Confronting ethnicity-specific disease risk. *Nat Genet* 38, 13–15.

Tang, H., Quertermous, T., Rodriguez, B., Kardia, S.L., Zhu, X., Brown, A., Pankow, J.S., Province, M.A., Hunt, S.C., Boerwinkle, E. et al. (2005). Genetic structure, self-identified race/ethnicity, and confounding in case-control association studies. *Am J Hum Genet* 76, 268–275.

Torrance, N., Mollison, J., Wordsworth, S., Gray, J., Miedzybrodzka, Z., Haites, N., Grant, A., Campbell, M., Watson, M.S., Clarke, A. et al. (2006). Genetic nurse counsellors can be an acceptable and cost-effective alternative to clinical geneticists for breast cancer risk genetic counselling. Evidence from two parallel randomised controlled equivalence trials. *Br J Cancer* 95(4), 435–444.

U.K. Human Genetics Commission (2003). Genes Direct: Ensuring the effective oversight of tests supplied directly to the public. Available at http://www.hgc.gov.uk/UploadDocs/Contents/Documents/Genes%20direct%20-%20FULL%20REPORT%20FINAL.pdf.

United Nations Educational, Scientific and Cultural Organization (UNESCO) (1997). *Universal Declaration on the Human Genome and Human Rights.* Available at http://portal.unesco.org/shs/en/ev.php-URL_ID=1881&URL_DO=DO_TOPIC&URL_SECTION=201.html.

United Nations Educational, Scientific and Cultural Organization (UNESCO) (2003). *A declaration on human genetic data.* Available at http://portal.unesco.org/shs/en/ev.php-RL_ID=1882&URL_DO=DO_TOPIC&URL_SECTION=201.html.

The ENCODE Project Consortium (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799–816.

Wall, A. (July 16, 2007). Races of Mexico and the Mexican Genome Project. Available at http://www.mexidata.info/id1442.html.

Weigmann, K. (2006). Racial medicine: Here to stay? The success of the International HapMap Project and other initiatives may help to overcome racial profiling in medicine, but old habits die hard. *EMBO Rep* 7(3), 246–249.

The Wellcome Trust (1996). Summary of principles agreed at the International Strategy Meeting on Human Genome Sequencing. Available at http://www.gene.ucl.ac.uk/hugo/bermuda.htm.

The Wellcome Trust (2003). Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility. Available at http://www.genome.gov/Pages/Research/WellcomeReport0303.pdf.

Wendler, D. and Emanuel, E. (2002). The debate over research on stored biological samples: What do sources think?. *Arch Intern Med* 162(13), 1457–1462.

Wendler, D. and Pentz, R. (2007). How does the collection of genetic test results affect research participants. *Am J Med Genet A*. DOI 10.1002/ajmg.a.31823.

World Health Organization (2005). Genetics, genomics and the patenting of DNA: Review of potential implications for health in developing countries. Available at http://www.who.int/genomics/FullReport.pdf.

Zerhouni, E. (2003). The NIH Roadmap. *Science* 302, 63–72.

Zerhouni, E.A. (2005). Translational and clinical science – time for a new vision. *N Engl J Med* 353, 1621–1623.

CHAPTER

Federal Regulation of Genomic Medicine

Janet Woodcock

INTRODUCTION

Regulation of the medical uses of genomic technologies is an evolving and controversial field. In the ideal case, medical product regulation strikes an appropriate balance between protecting patients and fostering innovation. Such balance can be difficult to achieve during the dynamic evolution of a new area of technology. Currently, the fields of genomic testing and pharmacogenomics are experiencing very active growth and corresponding regulatory and policy interest, and there is ongoing debate about the appropriate level of regulation of these technologies (Hudson, 2006; Javitt and Hudson, 2006). This chapter provides an introduction to both the regulatory framework and the ongoing policy challenges for genomic medicine.

Commercial introduction of genetic tests began in the late 1970s; however, the pace of introduction was slow at first and has rapidly accelerated over the last decade. It is estimated that over 1300 genetic tests are currently available in the United States (Gene Tests, www.genetests.org). Most of these are diagnostic tests for genetically based disorders. Recently, new types of *in vitro* diagnostics (IVDs) – for example, assays evaluating gene expression – are being introduced. Based on the current amount of activity in the field, an explosion of new assays can be anticipated over the next several years.

The landscape of genetic test regulation is somewhat confusing. The vast majority of available genetic tests are "laboratory-developed tests." Such assays are not FDA approved; rather, they are developed and conducted within a specific laboratory and are offered as a service by the laboratory. In contrast, all marketed test kits go through an FDA regulatory process of some type. Laboratories that develop an in-house test may submit an application for marketing clearance to the FDA; however, few have chosen to do so. This framework has resulted in a perception of an "uneven playing field" for the two assay categories. All laboratories reporting out human results require Clinical Laboratory Improvement Amendments (CLIA) certification. The CLIA program is administered by the Centers for Medicare and Medicaid Services (CMS). The respective roles of FDA and CMS have not been clear to all, especially newcomers, in the field and have resulted in considerable confusion. Recently FDA issued several "Guidance" documents intended to specifically clarify FDA regulation of newer diagnostic tests, including genomic tests. These will be discussed in more detail below.

Researchers interested in developing a specific assay for use in clinical medicine will need a basic understanding of the various regulatory regimes, as described further in this section. Exploratory assay development using human samples – for example, tissues or body fluids – can usually be conducted under the general human subject protection and privacy provisions implemented by academic institutions. Prior to undertaking analytical validation or clinical studies, however, it would be prudent for a developer to determine the potential regulatory

Copyright © 2010, Elsevier Inc.
All rights reserved.

status of the test, and the types of studies and regulatory clearances needed to bring the test to market successfully. Further evolution of the current regulatory framework for genomic tests is likely if the new discipline of genomic medicine is as successful as is currently anticipated.

In contrast to genetic diagnostic testing, the field of pharmacogenomics is in fairly early development. Genomic technologies have been used extensively in drug discovery for over a decade, but application to drug development is a more recent phenomenon. Pharmacogenetics can be defined as the study of the genetic contributions to drug responses (see Chapter 15). Drug metabolizing enzyme polymorphisms are the best understood examples of pharmacogenetics because correlations between phenotypes and drug concentrations have been well worked out for a number of polymorphisms. Despite decades of study, however, metabolizing enzyme polymorphisms are rarely taken into account in drug development or clinical practice, in part because convenient assays have not been available. Recently, genetic tests for common drug metabolizing enzyme polymorphisms have been introduced and several have been approved by the FDA; however, as of this writing, no drug label contains *specific requirements* for dose adjustment based on metabolizing enzyme test results. Routine use of such tests awaits demonstration of added value. Pharmaceutical developers generally seek to eliminate, prior to clinical testing, candidate drugs that are subject to polymorphic metabolism or that may be prone to drug–drug interactions based on metabolic factors. Traditional clinical drug development processes have not included individualized dose adjustments, and there are strong commercial and logistical disincentives to incorporating such individualization. The first pharmacogenetically directed, dose-adjusted drug development program will likely involve:

- a promising molecule that developers are unable to engineer to avoid an undesirable metabolic pathway;
- an indication for a very serious or life-threatening illness; and
- an expected narrow therapeutic index (e.g., cancer therapeutics).

Much of the current scientific interest in drug metabolism involves marketed drugs, many of which are metabolized by polymorphic enzymes. Clinical practitioners have long been accustomed to performing empirical dose adjustments based on clinical parameters, for example, INR for warfarin, leukocyte counts for 6-mercaptopurine, or side effects for many drugs. There is considerable skepticism within the clinical and reimbursement communities about the value of up-front genetic testing and dose adjustment. Resolution of these issues will require randomized clinical outcome trials that demonstrate added safety or effectiveness, or improved patient convenience/adherence and/or cost-effectiveness as a result of testing and dose adjustment. From a scientific perspective, it is unquestionable that the many-fold variation in drug exposure often introduced by polymorphic metabolism has important clinical consequences in both drug development and in patient care, particularly for drugs with narrow therapeutic indices. However, ingrained practices change slowly, and, in any case, there is a need for evidence-based dosing algorithms for practitioners in instances where dose adjustment is found to be important (see for example www.warfarin-dosing.org).

Although drug metabolism is the best understood example, the promise of pharmacogenomics for drug and vaccine development and use is much more expansive. Pharmacogenomic tests are being explored for selecting patient subgroups for treatment (based on probability of response, prognostic category, or disease subtype), identifying patients with a high probability of an adverse response to therapy, and for monitoring response. Additionally, the field of toxicogenomics, that evaluates the genomic responses to toxicants, is expected to contribute much more sensitive toxicity assays and greatly enhanced understanding of the mechanisms of drug toxicity. All these applications are in the early stages of development. The most extensive clinical use occurs in the field of antiretroviral therapy, where HIV drug resistance testing is routinely performed by assaying the viral genome, and therapy is directed by the results. Both FDA-approved and laboratory-developed HIV drug resistance assays are commercially available. With respect to the human genome, the most developed examples occur in cancer, where various genomic assays are employed to provide information about tumor aggressiveness and prognosis, or about likelihood of response to a targeted therapy (e.g., imatinib, trastuzumab). NIH, FDA and the pharmaceutical industry are also pursuing efforts to understand the genetic basis of drug-related adverse events. An historical example of a genetically based drug adverse event is drug-induced hemolysis in people with G6PD deficiency (metabolism based); a non-metabolic, recent example is the abacavir hypersensitivity reaction (Rauch et al., 2006). It is expected that some, but not all, drug-related adverse events will have a genetic contribution.

Because of the potential for pharmacogenomics to improve drug safety and effectiveness, the FDA has been vigorously involved in efforts to foster this new field. In November 2003, the Agency established the "voluntary genomic data submission process," a kind of safe harbor where pharmaceutical developers can share genomic experiments and data with the FDA in a non-regulatory context (www.fda.gov/cder/genomics/GDS.htm). Multiple data submissions have been shared with FDA and this process is expected to help ease the transition of pharmacogenomic technologies into mainstream preclinical and clinical drug development.

Nevertheless, numerous policy questions related to the use and regulation of pharmacogenomics in drug development remain to be answered. For example, what regulatory standards should apply during co-development of a pharmacogenomic test and drug to be used in combination? What data and studies are needed to qualify genomic biomarkers for regulatory use? "Regulation of Genomic Tests" explores the regulatory framework for the use of pharmacogenomic technologies in drug development and therapy, and discusses various unresolved policy areas.

REGULATION OF GENOMIC TESTS

Researchers intending to translate basic genomic discoveries into new diagnostic or prognostic assays will need to have a basic understanding of the regulatory framework governing such tests.

FDA regulations define IVDs products as "those reagents, instruments and systems intended for use in the diagnosis of disease or other conditions, including a determination of the state of health, in order to cure, mitigate or prevent disease or its sequelae." The regulations go on to say: "These products are devices as defined in Section 201(h) of the Federal Food Drug and Cosmetic Act …" (21 CFR 809.3) Therefore reagents, systems, and assays for human use are considered medical devices under the law and are subject to FDA regulation. However, medical device regulation has considerable flexibility, and this has been exercised by FDA in its oversight of IVDs. Additionally, the FDA Modernization Act passed by Congress in 1997 called upon FDA to use an approach to device regulation that is "least burdensome" on manufacturers while still achieving public health objectives, and this charge is taken into account when implementing regulatory standards.

Regulation of *In Vitro* Diagnostics as Medical Devices

Prior to embarking on a development program for a genomic test, researchers and developers should strongly consider consulting with the FDA staff about the development plan. Most genomic assays regulated as medical devices will be overseen within the Office of *In Vitro* Diagnostic Device Evaluation and Safety (OIVD) in the Center for Devices and Radiological Health (CDRH) at FDA. CDRH offers assistance to researchers, developers, and device manufacturers, who can request a "pre-IDE" meeting with OIVD to discuss the proposed development plan. Such consultations can save time and prevent wasted efforts. Developers can submit their planned protocols for evaluating the performance of the IVD and reach agreement with FDA staff on the extent and nature of the studies. Contacts can be reached through the CDRH web page (www.fda.gov/cdrh/index.html). FDA has also recently published a relevant draft Guidance entitled "Pharmacogenetic Tests and Genetic Tests for Heritable Markers" that contains advice on product development (www.fda.gov/cdrh/oivd/guidance/1549.html).

Regulatory Classification of *In Vitro* Diagnostics

Medical devices are classified into categories, based on risk, that govern the intensity of their regulation. For IVDs, the risk determination is primarily a function of the intended use of the device, based generally on the claims sought by the manufacturer. Many types of IVDs have been formally classified by the FDA in regulations. Class I devices are considered low risk and are subject to the fewest regulatory controls. These so-called "general" controls include, for example, facility registration, requirements for label information and format, manufacturing procedures, and reporting of adverse events. Many (non-genetic) IVDs have been assigned to Class I and are exempted from submitting applications for marketing to the FDA. Class II devices are moderate risk and require additional controls. Many genomic tests will be considered Class II IVDs. As a relevant example, drug metabolizing enzyme genotyping systems are Class II. In making this determination, FDA published the document "Class II Special Controls Guidance Document: Drug Metabolizing Enzyme Genotyping System," in 2005 (www.fda.gov/cdrh/oivd/guidance/1551.html). This document explains what data to submit to FDA in a Premarket Notification Submission, also known as a 510(k) (see below) for a diagnostic of this type. By submitting this information, and also complying with the "general" controls, a manufacturer can meet the requirements for marketing a drug metabolizing enzyme genotyping system. Special controls vary with the category of IVD. In some cases, clinical studies may need to be submitted as part of a 510(k) submission for a Class II IVD.

Class III IVDs are considered high risk and often require submission of a Premarket Approval Application (PMA). PMAs usually include data from clinical studies. If a device category has not been classified, it is considered Class III.

Applications for Marketing *In Vitro* Diagnostic Tests

The Premarket Notification Submission or 510(k)

When submitting a traditional 510(k) application, a manufacturer intends to demonstrate that its IVD is "substantially equivalent" (the legal terminology) to a previously legally marketed IVD, called the "predicate device." If a predicate IVD exists, clinical data in the 510(k) for a genomic test would include information on the analytic validation of the test, plus comparisons with the predicate IVD performance, using appropriate clinical samples (Figure 19.1).

If no predicate exists, the FDA can, when appropriate, use a process know as a *de novo* 510(k) to establish the special controls needed for a Class II IVD, as discussed for the drug metabolizing enzyme genotyping test systems mentioned above. In these instances, the medical literature or submitted data can be used by FDA to help determine the needed extent of special controls. The *de novo* 510(k) may contain clinical data consisting of protocol-based testing of clinical samples, may sometimes also include data gathered in prospective clinical trials, and in some cases may include data derived from the clinical literature and/or information found in recognized clinical practice standards.

In February 2007 FDA approved Agendia's MammaPrint® as a *de novo* 510(k). This test is intended to evaluate the likelihood of breast cancer recurrence within 5–10 years in patients with Stage I or Stage II disease. It is the first FDA-cleared molecular test using genetic profiling to predict breast cancer prognosis and is indicated as a prognostic marker only. Total FDA review time for this application was 30 days; the entire process (which necessitated the firm submitting a petition for down-classification) took less than 180 days. The special controls guidance document for these types of assays, "Class II Special Controls Expression Profiling Test System for Breast Cancer Prognosis," has been published (www.fda.gov/cdrh/oivd/guidance/1627.pdf). The clinical data for this application was based on analyses of stored samples from various studies (www.fda.gov/cdrh/reviews/K062694.pdf).

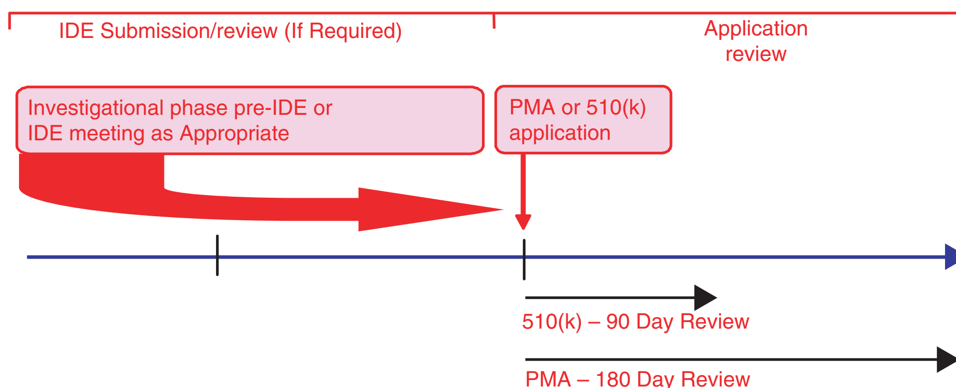


Figure 19.1 *In vitro* diagnostic development process. The requirement for an Investigational Device Exemption (IDE) filing is based on risk: most clinical studies of IVDs do not need to be conducted under an IDE. Sponsors may request a meeting with FDA at any time during the development process to obtain regulatory guidance on the development program or on how to submit an application: request a "pre-IDE" meeting if no IDE exists, or an "IDE" meeting if an IDE has been filed. The timeframe for FDA review depends on whether a 510(k) or PMA has been filed.

FDA has published extensive guidance on submission of 510(k)s for IVDs, including the documents "Format for Traditional and Abbreviated 510(k)s" (www.fda.gov/cdrh/ode/guidance/1567.pdf) and "Points to Consider for Collection of Data in Support of *in vitro* Device Submissions for 510(k) Clearance" (www.fda.gov/cdrh/ode/95.pdf) FDA has also recognized dozens of voluntary standards (most developed by the Clinical Laboratory Standards Institute or CLSI) for use in establishing test performance. Published guidances and FDA-recognized standards can be found on the CDRH web page (www.fda.gov/cdrh/index.html).

The Premarket Approval Application

The claims attached to certain genomic tests may cause them to be classified into Class III (high risk). An example of such claims could include prognostic information linked to therapeutic decision-making (e.g., more extensive radiation or chemotherapy for cancer). PMA submissions contain clinical data that substantiate these claims, usually derived from clinical trials. Developers often meet with the FDA staff to reach agreement on the design of these trials prior to their initiation.

Investigational *In Vitro* Diagnostics

Investigational device exemptions (IDEs) must be submitted to the FDA prior to beginning clinical studies of certain devices. IDEs are rarely needed for IVDs. In clinical studies where the results of the test will not be used in patient management, an IDE is not needed; however, the usual requirements for informed consent and IRB approval remain, where applicable. In those instances where the test results will determine management (e.g., patients are randomized to further interventions based on test results), an IDE should be submitted. OIVD staff may be consulted on the need for an IDE filing.

If investigational IVDs are shipped for research testing, for example, analytical validation, they should be labeled "For Research Use Only. Not for use in diagnostic procedures" (21 CFR 809.10 (c)(2)(i)). If they are shipped for use in comparative testing or as part of a clinical trial, they should be labeled "For Investigational Use Only. The performance characteristics of this product have not been established" (21 CFR 809.10 (c)(2)(ii)).

FDA recently issued a document entitled "Guidance on Informed consent for *In Vitro* Diagnostic Device Studies Using Leftover Human Specimens that are Not Individually Identifiable" that offers advice on the use of de-identified human specimens for testing (www.fda.gov/cdrh/oivd/1588.pdf).

Laboratory-Developed Tests

As stated above, the vast majority of currently marketed genetic tests are laboratory-developed tests that have not undergone the FDA review process. These tests are either fully developed by the laboratory or utilize a purchased "analyte-specific reagent" (ASR, see below) that is then configured into an assay by the laboratory, which carries out the test as a service. Such laboratory developed tests are not shipped for use outside the originating site. These tests are considered medical devices under the Food, Drug and Cosmetic Act (and the laboratories are considered device manufacturers), but FDA has not required marketing applications (using what is legally termed "enforcement discretion"), considering that the controls provided by the ASR regulations and the certifications for high complexity laboratories under CLIA (see below) to be sufficient. In September 2006 FDA issued a draft guidance document entitled "*In Vitro* Diagnostic Multivariate Index Assays" explaining that certain complex test systems, including certain genomic test systems, are Class II or Class III medical devices requiring 510(k) submissions or PMAs, respectively (www.fda.gov/cdrh/oivd/1610.pdf). The document defines the difference between *in vitro* diagnostic multivariate index assays (IVDMIAs) and the types of tests generally performed as laboratory-developed tests. This document and the related policy issues around laboratory-developed tests are generating considerable

discussion and debate, and the outcome of these debates will have consequences for developers of genomic tests.

Analyte-Specific Reagents

In 1997, FDA defined a group of reagents known as ASRs, and classified them (with some exceptions that are not germane to genomics) as Class I medical devices. This meant that ASRs were subject to "general controls" such as manufacturing and labeling requirements, but did not have to have applications cleared or approved by FDA to be put on the market. FDA accomplished this by publishing three regulations that defined and classified ASRs (21 CFR 864.4020), imposed restrictions on their sale, distribution and use (21 CFR 809.30) and established requirements for ASR labels (21 CFR 809.10(e)). In the regulations, ASRs were defined as "antibodies, both polyclonal and monoclonal, specific receptor proteins, ligands, nucleic acid sequences, and similar reagents which, through specific binding or chemical reactions with substances in a specimen, are intended for use in a diagnostic application for identification and quantification of an individual chemical substance or ligand in biological specimens" (21 CFR 864.4020). Thus, an ASR is a building block for an assay, but is not a test system. Subsequent to publication of these regulations, FDA observed a broadening of the implementation of ASRs beyond the stated parameters. Therefore, in September 2006, FDA issued a draft guidance entitled "Commercially Distributed Analyte Specific Reagents (ASRs): Frequently Asked Questions" (www.fda.gov/cdrh/oivd/guidance/1590.html). An important point in this document, for the purposes of this discussion, was that FDA did not consider "bundled" moieties (e.g., sets of primers), or extensively processed (e.g., arrayed on beads) reagents, or microarrays, to be ASRs. These distinctions are important to the genomic testing community and have elicited extensive comment. Undoubtedly the boundary between ASRs and test systems will continue to be clarified in the coming years.

Clinical Laboratory Improvement Amendments

Congress passed the CLIA in 1998. This law established quality standards for all laboratory testing to ensure the accuracy, reliability and timeliness of patient test results regardless of where the test is performed. The CLIA program is administered by the CMS (www.cms.hhs.gov/CLIA). The ASR regulations discussed above stipulated that the only clinical laboratories to which ASRs could be sold were those qualified under CLIA to perform high complexity testing (or alternatively, were regulated under the Veteran's Health Administration Directive 1106) (21CFR 809.30 (a)(2)). Currently, most genetic testing is performed by such laboratories. The oversight under CLIA relates to the quality of performance of laboratory testing. It does not extend to evaluation of the clinical utility of a given assay. There is ongoing controversy about the degree of oversight of genetic testing under CLIA (Hudson et al., 2006); specifically, whether there should be a specific genetic testing specialty area that would incorporate proficiency testing for genetic tests.

To summarize the current Federal regulation of genomic testing: most genetic tests are on the market as laboratory-developed tests. Such laboratories are subject to quality standards under CLIA. This is possible outside of FDA marketing clearance as a result of the ASR regulations and enforcement discretion on the part of FDA. FDA has approved a number of genetic tests, including tests for drug metabolizing enzyme polymorphisms, generally based on information available in the scientific literature. Recently, FDA issued draft guidances pointing out, among other things, that reagents such as micoarrays and sets of primers are not ASRs, and that IVDMIAs are not among the types of assays considered to be appropriate for laboratory-developed tests. These guidances have resulted in ongoing discussion and controversy, and the policy issues they raise have yet to be definitively resolved.

PHARMACOGENOMICS IN DRUG DEVELOPMENT AND CLINICAL MEDICINE: THE ROLE OF REGULATION

Researchers interested in translating genomic discoveries into new approaches to drug therapy will need to understand the challenges involved in incorporating these new scientific approaches into traditional drug development and use patterns. While genomic technologies are used extensively and routinely during drug discovery and in early drug development phases (e.g., lead optimization) routine utilization of genomics drops precipitously as candidates enter the animal safety and clinical testing phases. As a result, few investigational drugs now in the clinical pipeline are being studied in concert with pharmacogenomic assays (although most development programs include collection of human samples for potential genetic testing). Likewise, although extensive information exists about the contribution of drug metabolizing enzyme polymorphisms to variability in drug exposure, as of early 2007, no marketed drug label contains an explicit requirement (as opposed to recommendations) for dose modification based on genetic test results. (See "Genomics at FDA" (www.fda.gov/cder/genomics/genomic_biomarkers_table.htm) for an up-to-date list of pharmacogenomic tests utilized in drug labels.) These facts, when considered in light of the enormous potential for improvement in drug therapy that pharmacogenomics represents, demonstrate that the field is in its nascent stages.

Many barriers to rapid uptake of genomics in drug development and clinical medicine exist, beyond the purely scientific and technical challenges. Drug development is an enormously risky endeavor. Only about one in nine drug candidates entering clinical investigation reach the market, and every year promising candidates encounter catastrophic, highly public late-stage failures. Some compounds that reach the market fail to achieve commercial success. Additionally, drug development programs are extremely expensive. Furthermore, the clinical investigation and marketing application stages of development are subject to extensive regulatory scrutiny worldwide. These factors result in very conservative approaches on the part of pharmaceutical developers. Developing an investigational drug using a relatively untested pharmacogenomic-directed approach is perceived

as adding additional risk. The fact that drug development is a highly competitive activity also restricts the flow of information among developers, so that those who instigate new approaches may not share the successes or failures with others. Under these circumstances, the regulators may be the only group possessing a broad perspective on the state of the field.

Commercial considerations are also important in development. Developers normally pursue the broadest possible market for a new intervention. Many pharmacogenomic approaches (e.g., targeting therapy to a specific-identified population), would limit the indicated population. Complexity of use (e.g., dose adjustments, need for pre-testing) is also commercially undesirable and in fact can limit clinical uptake. For these reasons, along with the natural human optimism about the performance of a new discovery, pharmacogenomic tests have often been considered as a salvage approach when one-size-fits all development fails. Concerns about restrictions on markets are slowly being mitigated by the success of targeted therapies but still limit enthusiasm for pharmacogenomic approaches during drug development (Bernstein, 2006).

Once a drug is approved by the FDA, a manufacturer has a limited amount of time to recoup its investment before patent and exclusivity protections expire and generic competition occurs (generic copies are permitted for small molecule drugs but not for therapeutics approved as biological products). Therefore there is limited time to explore refinements of an approved therapy using pharmacogenomics. Once generic competition occurs there is no commercial incentive for further development.

The promise of pharmacogenomics and the barriers to its adoption led FDA to aggressively promote investigations of the new technology. Senior FDA staff published a 2002 paper calling for progress in the field and pointing out the Agency's interest (Lesko and Woodcock, 2002). In May 2002, a workshop between FDA and the pharmaceutical industry was held to explore barriers to information sharing (Lesko et al., 2003). At the workshop, industrial sponsors stated reservations about providing genomic information on compounds in development to the Agency because they feared uninformed questions and delays in product development; however, all agreed on the desirability of information sharing. FDA subsequently established a process for sharing data and results in a non-regulatory context, the "Voluntary Genomic Data Submission" (VGDS) (www.fda.gov/cder/genomics/GDS.htm). Subsequently, over thirty submissions of genomic data involving drugs in development, encompassing topics such as animal studies, preventive indications, identification of responsive subgroups, and genetic links to adverse events, have been discussed with FDA scientists. The goals of VGDS process include (1) encouraging the development and sharing of genomic data linked to therapeutic and preventive indications; (2) helping familiarize FDA scientific staff with the emerging uses of genomic information in drug development; (3) developing data standards and analytical methods suitable for regulatory use; and (4) enabling the transition of genomic studies from exploratory research to regulatory submissions. (See www.fda.gov/cder/genomics/presentations/Webinar.pdf for a detailed presentation of the VGDS process.) Subsequent workshops on drug-diagnostic co-development issues and on genomic data submission standards have been held (www.fda.gov/cder/genomics/biomarkers.htm, Hinman et al., 2006). The increased level of dialog between industry and the FDA has mitigated some of the concerns about regulatory barriers. The European Medicines Agency (EMEA) also instituted pharmacogenomics "Briefing Meetings" in 2003 (The European Agency for the Evaluation of Medicinal Products, 2003). In May 2005 FDA and EMEA held their first joint VGDS meeting, and additional joint meetings are occurring. Discussions of regulatory aspects of pharmacogenomics are now being held among regulators and industry worldwide under the auspices of the International Conference on Harmonization (ICH) of the Technical Requirements for Pharmaceuticals.

Translation from the laboratory into clinical development is only one of the translational barriers faced by new technologies. Achieving uptake in clinical practice and obtaining reimbursement are also major challenges: the utility and value of new technologies must be explicitly demonstrated. Ironically, the tight regulation of pharmaceuticals will tend to enhance acceptance of FDA-approved pharmacogenomic test-directed drugs. The regulatory standards for drug approval include the scientific demonstration of effectiveness and safety. When a new drug is approved in conjunction with a pharmacogenetic test, the likelihood of use as recommended in the drug label is very high – in part because of confidence in the approval standard, and in part because promotion (which is often extensive for a new therapy) is limited by law to the label uses. Conversely, "retrofitting" a pharmacogenetic test into established practice patterns can be very difficult. Both the clinical and reimbursement communities have made it very clear that, in the majority of cases, randomized outcome trials will be the standard that will govern acceptance – even for dose adjustment with drug metabolizing enzyme assays.

From a regulatory standpoint, the current issues in pharmacogenomics applied to drug development and clinical use include the following: development of an investigational drug incorporating a pharmacogenomic test; use of a pharmacogenomic test to enhance the performance of an approved drug; drug safety pharmacogenomics and toxicogenomics; and qualifying genomic biomarkers for regulatory use. These issues are discussed in more detail in the sections below.

Development of an Investigational Drug with a Pharmacogenetic Test to Select Patients for Therapy or for Monitoring

Use of diagnostics to discriminate among individuals who present with similar symptoms is routine clinical practice. Genomic tests are conceptually no different; however, these are new types of tests, and they will yield novel insights into the sources of human variability that are currently not utilized in decision-making. Pharmacogenomic tests may be used during drug development to identify patients with different disease subsets (that differ by prognosis, response to therapy, etc.), to identify individuals with disease responsive to an particular intervention (i.e., targeted therapies), to stratify individuals to different dose

regimens, to identify individuals with a high probability of an adverse reaction to treatment, or to monitor treatment response. In all these cases, the use of the diagnostic is intended to enhance drug safety or effectiveness. One major challenge in using pharmacogenomic tests during drug development is that often the assay itself must be developed in parallel. This raises the level of complexity significantly. A full discussion of the issues involved in a drug-diagnostic co-development program is beyond the scope of this chapter. The points below indicate the major areas that must be considered. FDA is planning to issue a guidance document covering this topic in detail (Woodcock, 2005).

Regulatory Considerations for Pharmacogenetic-Test-Directed Therapy Development

The types of data that need to be generated for an investigational drug paired with a pharmacogenomic test depend on the intended label claims for the drug and the test. If use of the drug requires results from the test, demonstration of the safety and effectiveness of the drug-test combination in the definitive trials will be required. If mention of the test in the drug label is strictly informational, a lesser standard of evidence applies. For example, data from retrospective analysis of samples might lead to a suggestion in the label that patients with result *x* on test *y* may have a lower probability of response. However, in this scenario, the drug would be indicated for the entire population, absent test *y*, and would need to be shown effective and safe in the entire population. Most pharmacogenomic information in current drug labels is of this latter type. Exceptions include trastuzumab and imatinib (Bernstein, 2006). Often, developers do not know, at the start of clinical trials, which label scenario they will end up with. In these cases, design of an adaptive program that could support either result may be desirable (see trial designs, below).

Drug and Diagnostic Development Considerations

Development of the Diagnostic Test

A reliable, reproducible test configuration should be achieved prior to use of the diagnostic in prospective drug trials. Stored samples are frequently used for this purpose, and methods for analytical validation of diagnostics are well worked out. A more significant challenge involves setting cut-off values for the test results. Genomic assays often contain multiple analytes, which are sometimes continuously variable, giving rise to a nearly infinite range of results. Use to direct drug therapy will require achieving unambiguous, often dichotomous (i.e., yes/no) results. These cut-offs often will be constructed via retrospective analysis of clinical samples from the investigational drug trials. This exercise will create an hypothesis: test *x* will segregate population *y* into distinct subgroups. This hypothesis can then be verified using additional stored samples or in prospective studies. Depending on the clinical implications of the subgroups, such data might suffice for a stand-alone claim for the diagnostic. However, a claim for use in combination with the drug will require trials demonstrating the clinical utility of the diagnostic as well as drug safety and effectiveness. If the results of such trials require retrospective analyses that readjust diagnostic cut-offs to achieve significant results, they will be regarded as hypothesis generating rather than confirmatory.

Trial Design Considerations for Investigational Drug Combinations

Major design issues arise in this context. These involve ways to integrate the developing information on drug and diagnostic performance when it is unclear whether the diagnostic adds value to the drug therapy (and it is still unclear if the drug has value). At the end of Phase 2 trials, developers may know that the diagnostic can discriminate subgroups but will not have information on how much this impacts performance of the drug. For example, suppose a genomic test appears to identify a patient subset more responsive to the therapy. One approach might be to perform an "enriched" trial, randomizing test-positive patients to investigational drug versus control. This trial answers the question: does the drug work in test-positive patients? It does not yield information on drug performance in test-negative patients (so it does not tell you if the test discriminates anything meaningful), or whether test-positive patients have generally better or worse prognoses. A more elaborate trial might enroll all patients with the condition and then stratify both treatment and control arms by genomic test result. This design can answer many questions but risks answering none definitively because of the multiple comparisons that arise. Simon (2005) and Temple (2005) have proposed various methodological approaches including adaptive designs to address these problems. A number of sequential designs could also be useful. The appropriate design in a given case will be influenced by both mechanistic (how plausible is the test-drug link?) and clinical considerations – for example, how much information on the test-negative subgroup is needed when a claim for effectiveness in the test-positive subgroup is sought? In cases where the mechanistic link is well understood – for example, HIV drug resistance – studying a test-defined subgroup alone is standard practice. In other cases, information on results from the test-negative group may be crucial. Developers should consult with FDA early in such development programs.

Dosing Adjustments

Incorporating pharmacogenetically driven dose adjustments into Phase 2 and 3 clinical trials is straightforward. Typically, dose finding is performed in the second phase of drug development. These trials evaluate the dose-response relationships for safety/tolerance and efficacy (which may be defined by a pharmacodynamic marker). "Dose-response" is more properly understood as "concentration response"; and there are excellent pharmacometric methods for analyzing the impacts of variables such as test results. When the effect of a polymorphism is large, the drug developer may then plan to adjust dosing in the efficacy trials for alleles with known effects on concentration. Questions will arise about default dosing for individuals with non-evaluable alleles, as well as how to extrapolate the results to populations with different allele frequencies. For life-threatening diseases, for example some cancer indications, where development programs may incorporate

a combined Phase 2–3 development program and may lack pharmacodynamic endpoints, opportunities to understand the impact of variation in drug metabolism are fewer, and evaluation should start as early as possible in the development program.

Pharmacogenomic Safety Tests

Drug toxicity is a major issue in development. Beyond toxicity attributable to relative overdosing from variable metabolism, genetically related adverse events can arise from drug target (or related pathway) polymorphisms, immunologic predisposition (e.g., abacavir hypersensitivity), genetically based disease variability, and genetically based organ vulnerabilities (White et al., 2006). Sorting out the mechanistic bases of drug adverse events will be straightforward in some cases and highly complicated in others. The size of most drug development programs will only allow elucidation of the most clear-cut genetic association. Regulatory issues do not complicate such investigations, except in the rare instance where the drug is not safe enough to be approved absent the test. In these cases, sponsors should discuss with FDA the amount of evidence needed to demonstrate that use of the test confers an adequate level of safety.

Pharmacogenomic Tests for Approved Drugs

It is likely that successful implementation of pharmacogenomic testing would significantly improve therapeutic outcomes of drug therapy. In particular, drug toxicity exacts a severe human and economic toll in the United States, being estimated as the 4th–6th leading cause of death, causing tens of thousands of hospitalizations, and costing billions of dollars (Budnitz et al., 2006; Lazerou et al., 1998). Surveys suggest that drugs with polymorphic metabolism contribute a disproportionate share of harm (Phillips et al., 2001). Nevertheless, as detailed above, there are significant barriers to even initiating the work that needs to be done. Once opportunities are identified out of the basic science, funding for controlled clinical outcome trials needs to be obtained. Pharmaceutical companies may sponsor trials for marketed drugs that are not off-patent. Trials of pharmacogenomic tests for off-patent drugs may require public funding or sponsorship by consortia of consisting of government, health care systems, the pharmaceutical industry, etc. For example, currently a number of groups are considering or initiating trials of pharmacogenetic-directed dosing of warfarin. Issues beyond funding include how to actively involve in consortia diagnostic companies and researchers who have developed tests.

FDA has been actively seeking, as an initial step, to incorporate established information on drug metabolizing enzymes into drug labels without stipulating testing and dose adjustment. Examples include 6MP (6-mercaptopurine) (the enzyme TPMT), atomoxitine (Strattera) (CYP2D6) and irinotecan (UGT1A1) (www.fda.gov/cder/genomics/genomic_biomarkers_table.htm) It is possible that in the future, pharmacogenetic-directed dosing could be explored during studies of these agents that are being conducted for other purposes. For a good discussion of the issues involved in developing pharmacogenetic information on a marketed drug (see Ratain, 2006).

Many rare, serious adverse drug reactions occur across multiple drug classes and may have a genetic association (e.g., Stevens Johnson syndrome). These reactions occur so infrequently, however, that it is likely that international consortia will be required to assemble adequate cases and controls to perform genetic analysis. Other serious drug side effects occur more commonly, within a drug class (e.g., tardive dyskinesia). Public private partnerships between government and industry are being explored as ways to gather cases and carry out the needed analyses.

Toxicogenomics

When the toxicant is a drug, the areas of toxicogenomics and pharmacogenomics overlap. Many drugs share mechanisms of toxicity with non-pharmaceutical chemicals. Mechanistic understanding of drug toxicity will greatly improve the science of drug safety. The current protocols for preclinical animal toxicology testing during drug development are highly empirical and offer few ways to evaluate the relevance of a given toxic finding to humans. Toxicogenomics may offer methods to discriminate by comparing cellular responses in various animal species and in human cells. In some cases, a toxic finding in an animal species may not be mechanistically relevant in humans. Toxicogenomics is also anticipated to be a source of more sensitive toxicity assays than currently those currently used in traditional animal and human testing. Under the auspices of the C-Path Institute, a nonprofit organization in Tucson, Arizona, a consortium of twelve pharmaceutical companies are pooling, evaluating and validating nontraditional toxicity assays (some of them genomic), on an organ system basis, with the intent of submitting the best performing assays to FDA for regulatory assessment (www.c-path.org/programs/SafePath/tabid/61/Default.aspx). It is expected that toxicogenomic tests will be generally relevant to drug development, rather than linked to a particular drug.

2009 UPDATE

The debate over the role and scope of federal regulation of genetic tests in the United States intensified in 2008. In April, the HHS Secretary's Advisory Committee on Genetics, Health, and Society (SACGHS) issued a report entitled "US System of Oversight of Genetic Testing: A Response to the Charge of the Secretary of Health and Human Services" (http://www4.od.nih.gov/oba/SACGHS/reports/SACGHS_oversight_report.pdf). This report contains an extensive set of recommendations for increased oversight by various federal agencies in the United States, including broader requirements for premarket clearance on the part of Food and Drug Administration (FDA). However, any new policy initiatives in this area will be the purview of the

new administration. It is widely recognized that any changes in regulatory policy should be accompanied by increases in funding for translational research as well as modifications of reimbursement policies (Hudson, 2008; Khoury et al., 2008; Woodcock, 2008). Such changes would help to provide adequate incentives for clinical development in the face of increased evidentiary requirements.

Pharmacogenomic testing could be described as having entered mainstream medicine in 2008. The FDA added genetic information to multiple drug labels, including abacavir, carbamazepine, and warfarin, with strong recommendations for pretesting in the case of abacavir and carbamazepine. The debate over the use of genetic algorithms for initial dosing of warfarin intensified. FDA is evaluating multiple additional pharmacogenetic tests for incorporation in drug labels. In December 2008, the FDAs Oncology Drugs Advisory Committee met to consider the proposal of the manufacturers of cetuximab (Erbitux) and panitumumab (Vectibix) that administration of these drugs be limited to patients whose tumors contain wild-type K-RAS. (http://www.fda.gov/ohrms/dockets/ac/08/briefing/2008-440961-00-FDA-index.htm.) This issue is currently under consideration at FDA. Additional diagnostic genetic tests were also approved by the FDA in 2008.

2009 UPDATE REFERENCES

Hudson, K. (2008). The health benefits of genomics: Out with the old, in with the new. *Health Aff* 27, 1612–1615.

Khoury, M.J., Berg, A., Coates, R., Evans, J., Teutsch, S.M. and Bradley, L.A. (2008). The evidence dilemma in genomic medicine. *Health Aff* 27, 1600–1611.

Woodcock, J. (2008). The human genome and translational research: How much is enough? *Health Aff* 27, 1616–1618.

REFERENCES

Bernstein, K. (2006). Targeted Narrow. *BioCentury* 14, 1–9.

Budnitz, D.S., Pollock, D.A., Weidenbach, K.N., Schroeder, T.J. and Annest, J.L. (2006). National surveillance of emergency department visits for outpatient adverse drug events. *JAMA* 296, 1858–1866.

Hinman, L.M., Huang, S-M., Hackett, J., Koch, W.H., Love, P.Y., Pennello, G., Torres-Cabassa., A. and Webster, C. (2006). The drug diagnostic co-development paper. Commentary from the 3rd FDA-DIA-PWG-PhRMA-BIO pharmacogenomics workshop. *Pharmacogenomics J* 6, 375–380.

Hudson, K.L. (2006). Genetic testing oversight. *Science* 313, 1853.

Hudson, K.L., Murphy, J.A., Kaufman, D.J., Javitt, G.H., Katsanis, S.H. and Scott, J. (2006). Oversight of US genetic testing laboratories. *Nat Biotech* 24, 1083–1090.

Javitt, G.H. and Hudson, K.L. (2006). Federal neglect. Federal regulation of genetic testing. *Issues SciTechnol* XXII, 59–66.

Lazerou, J., Pomeranz, B. and Corey, P. (1998). Incidence of adverse drug reactions in hospitalized patients: A meta-analysis of prospective studies. *JAMA* 279, 1200–1205.

Lesko, L.J., Salerno, R.A., Spear, B.B., Anderson, D.C., Anderson, T., Brazell, C., Collins, J., Dorner, A., Essayan, D., Gomez-Mancilla, B. et al. (2003). Pharmacogenetics and pharmacogenomics in drug development and regulatory decision making: Report of the first FDA-PhRMA-DruSAFE-PWG workshop. *J Clin Pharmacol* 43, 1–17.

Lesko, L.J. and Woodcock, J. (2002). Pharmacogenomic-guided drug development: Regulatory perspective. *Pharmacogenomics J* 2, 20–24.

Phillips, K.A., Veenstra, D.L., Oren, E., Lee, J.K. and Sadee, W. (2001). Potential role of pharmacogenomics in reducing adverse drug reactions. A systematic review. *JAMA* 286, 2270–2279.

Ratain, J.M. (2006). From bedside to bench to bedside to clinical practice: An odyssey with irinotecan. *Clin Cancer Res* 12, 1658–1660.

Rauch, A., Nolan, D., Martin, A. et al. (2006). Prospective genetic screening decreases the incidence of abacavir hypersensitivity reactions in the Western Australia HIV Cohort Study. *Clin Infect Dis* 43, 99–102.

Simon, R. (2005). Developing and validating genomic classifiers. http://www.fda.gov/cder/genomics/biomarkers.htm

Temple, R. (2005). Use of genomic biomarkers in a regulatory environment. http://www.fda.gov/cder/genomics.htm

The European Agency for the Evaluation of Medicinal Products, London, 23 January, 2003. (CPMP/4445/03).

White, T.J., Clark, A.G. and Broder, S. (2006). Genome-based biomarkers for adverse drug effects, patient enrichment and prediction of drug response, and their incorporation into clinical trial design. *Pers Med* 3, 177–185.

Woodcock, J. (2005). Importance of genomic biomarker validation in the context of pharmacogenomic initiatives at the FDA. http://www.fda.gov/cder/genomics/biomarkers.htm

21 CFR 809.10 (c)(2)(i)

21 CFR 809.10 (c)(2)(ii)

21 CFR 809.10(e)

21 CFR 809.30

21 CFR 809.30 (a)(2)

21 CFR 809.3

21 CFR 864.4020

RECOMMENDED RESOURCES

See presentations from FDA-DIA-PhRMA-BIO-PWG Workshop on "Application and Validation of Genomic Biomarkers for Use in Drug Development and Regulatory Decision Making", Bethesda, MD, October 6–9, 2005. http://www.fda.gov/cder/genomics/biomarkers.htm

Gene Tests. http://www.genetests.org/

CHAPTER

Economic Issues and Genomic Medicine

David L. Veenstra, Louis P. Garrison and Scott D. Ramsey

INTRODUCTION

The era of genomic medicine offers significant promise for the development of novel health care technologies that ultimately improve patients' quality of life and life expectancy. As with any new health care technology, questions arise as to their potential budgetary impact and cost-effectiveness. But because genomic technologies inherently involve diagnostic or prognostic testing, and because of the complexities of incomplete gene penetrance and multiple gene and environmental interactions, their assessment can be more challenging. In addition, perhaps more than in any other area of medicine, questions have arisen in regard to the economic incentives to develop these technologies. Formal health economics frameworks can be used to gain insights into these issues, and provide guidance for resource investment, technology appraisal, and policy development.

In this chapter, we provide a brief introduction to the principles of health economics and discuss the unique aspects of the application of cost-effectiveness analysis to genomic medicine. We review drivers of the cost-effectiveness of genomic technologies and assess several recent state-of-the-art examples. We then examine the particular challenges in pharmacogenomics of economic incentives for developing genetic tests linked to therapeutics. Lastly, we propose an approach to help establish value-based reimbursement of genomic testing technologies. The content of this chapter will help readers understand the multidisciplinary nature of the factors that will influence the value of genomic medicine, and provides a framework for assessing these factors in a systematic, quantitative fashion.

ECONOMIC EVALUATION AND COST-EFFECTIVENESS ANALYSIS

Health Economics and Foundations of Economic Analysis

Drawing on methods and concepts from economics, clinical epidemiology, psychology, and the decision sciences, the field of "cost-effectiveness" research has synthesized a set of tools and a theoretical framework for evaluating the complex issues in health care (Garber and Phelps, 1997; Weinstein et al. 1996). At a broad level, effectiveness estimates are generally derived from randomized clinical trials, disease progression estimates from long-term follow-up epidemiological studies, quality of life values from health state preferences studies, and costs from a wide variety of sources including payment schedules (i.e., Medicare) and attributable- and micro-costing studies. These factors, which are often considered either explicitly or implicitly in decision-making, can be synthesized using decision analytic techniques to inform formal economic analyses as discussed below (Detsky et al., 1997).

Copyright © 2010, Elsevier Inc.
All rights reserved.

Methodological Approaches to Economic Evaluation in Health Care

Several specific yet related methodological approaches are used in the economic evaluation of health care technologies: (1) cost-minimization, (2) cost-consequences analysis, (3) cost-benefit analysis, (4) cost-effectiveness analysis, and (5) cost-utility analysis (Flowers and Veenstra, 2004). These methods vary primarily in the way intervention effectiveness is valued. For example, for cost-minimization analysis, it is assumed there is no difference in effectiveness or side effects. In both cost-effectiveness and cost-consequences analysis, effectiveness is measured in natural, clinical units such as heart attacks or infections avoided. In cost-benefit analysis, a monetary value is assigned to effectiveness (e.g., a heart attack avoided might be "valued" at $100,000). And in cost-utility analysis, effectiveness is measured in quality-adjusted life years (QALYs), which account for improvements in both life expectancy and quality of life. Another characteristic that differentiates these methods is that cost-minimization and cost-consequences analyses do not involve the calculation of a ratio; for cost-minimization, only costs are presented, and in cost-consequences both costs and effectiveness are calculated, but not combined in a ratio. Of note, although "cost-effectiveness analysis" is a specific type of economic evaluation, the term is used generally to refer to all types of economic evaluation in health care.

In some cases, cost-effectiveness studies can be based on a single randomized clinical trial. However, because clinical trials are usually in a controlled setting, the costs incurred are not representative of utilization in a real-world setting. The "efficacy" observed in a controlled setting can also be distinguished from the "effectiveness" that could be expected in practice. Finally, the timeframe of clinical trials for chronic conditions are generally not sufficient to evaluate long-term outcomes. Because of these reasons, as mentioned above, modeling techniques such as decision analysis are often used to extrapolate the results from clinical studies using primarily epidemiologic and economic data from other sources.

In any economic evaluation, it is important that the technology being evaluated is compared to current medical practice. Weinstein and Stason, in 1977, defined the incremental cost-effectiveness ratio (ICER) as

$$\mathrm{ICER} = (C_2 - C_1)/(E_2 - E_1)$$

where C_2 and E_2 are the cost and effectiveness of the new intervention being evaluated, and C_1 and E_1 are the cost and effectiveness of the standard therapy (Weinstein and Stason, 1977). The costs and effects that are included in the equation depend on the perspective of the analysis. From a societal perspective, indirect costs and effects such as patient time away from work, downstream medical care costs years or decades after the intervention, and the quality of life of the patient and even their family need to be considered. Because of the all-encompassing nature of the societal perspective, it is generally better suited for national health care plans. More relevant to health care plans or providers in the United States is the payer perspective, which addresses primarily direct medical care costs incurred by the payer (e.g., drug cost, professional fees, hospital stay).

Resource Allocation Decision-Making

How can cost-effectiveness information guide health policy decisions? The favored approach, from an academic perspective, is to utilize cost-utility analysis because it allows for comparisons across interventions and diseases, accounts for impact on life expectancy and quality of life, and has theoretical foundations in welfare economics. Medical interventions are considered to be cost-effective when they produce health benefits at a cost comparable to other commonly accepted treatments. A general guide is that interventions that produce one QALY (equivalent to 1 year of perfect health) for under $50,000 are considered cost-effective, those between $50,000 and $100,000 per QALY are of intermediate cost-effectiveness, and above $100,000 per QALY generally is not considered cost-effective. The cutoff of $50,000 per QALY was derived loosely from the cost of providing dialysis for a patient for one year – a service paid for by Medicare for any US citizen. However, these criteria are somewhat arbitrary, and can vary across disease and indication-specific therapeutic areas, particularly in the United States (Neumann, 2007).

The application of cost-effectiveness analysis has increased dramatically in the past decade as a result of increasing health care costs and the desire to deliver the greatest health value for the money. The formal application of cost-effectiveness analysis to drug coverage decisions has its origins in countries with single-payer health care systems (e.g., government sponsored). Recently, multiple countries and health care systems have begun to adopt requirements for such pharmacoeconomic information. These requirements formalize an otherwise implicit demand for health care technologies that are cost-effective, and will influence, to a certain extent, "go/no-go" decisions in drug research and development.

The United Kingdom, Canada, and Australia all have formal requirements in place for cost-effectiveness information and programs in place for evaluating cost-effectiveness data – the National Institute for Health and Clinical Excellence (NICE, United Kingdom), the Canadian Agency for Drugs and Technologies in Health (CADTH, Canada), and the Pharmaceutical Benefits Advisory Committee (PBAC, Australia) (CADTH 2007; NICE 2007; PBAC, 2007), as well as several European countries such as the Netherlands (Drummond et al., 1999). In the United States, cost-effectiveness information is most often used in support of drug formulary listing in managed care settings. Many managed care organizations and pharmacy benefits managers utilize guidelines that require outcomes and economic information for formulary evaluation. In addition, the Academy of Managed Care Pharmacy (AMCP) has adopted guidelines for the submission of information, including outcomes and cost-effectiveness data, to support formulary consideration (Fry et al., 2003; Neumann, 2004; Spooner et al., 2007).

EVALUATING GENOMIC TECHNOLOGIES

Unique Aspects of Genomic Medicine

The tools of health economics and cost-effectiveness analysis can be used to assess any health care technology, from immunization strategies, to new drug therapies, to educational programs. Genomic medicine is not so unique (i.e., exceptional) that fundamentally new methods in economic evaluation are needed. However, genomics can add layers of complexity to health care interventions such that particular care should be taken in their evaluation.

Disease-Risk Testing

The validity of a prognostic genomic test will have a significant impact on its potential cost-effectiveness. The evaluation of validity will be similar to other non-genomic prognostic tests, yet potentially more complex, due to challenges in validating genetic markers with small contributions from multiple genes and environmental interactions. Thus, when undertaking an economic study, care should be taken that the performance of the test has been evaluated by peer review, and perhaps more importantly, reproduced by independent investigators in different patient populations. An issue that will be more unique, to a degree, to genetic testing for disease risk is whether to test family members. The potential costs and clinical and patient outcomes should be considered. If potentially substantial, this strategy should be included in the analysis (Ramsey et al., 2001).

Pharmacogenomics

Pharmacogenomics tests face a similar issue with validation of a test's ability to predict drug response and differentiating it from prognostic ability – that is, predicting disease prognosis rather than the effect of treatment on the disease. Furthermore, as discussed later in the chapter, the pairing of pharmacogenomic tests with drugs may lead to uncertainty about which – the test or the drug – provides the incremental value, and this can have implications for pricing or reimbursement decisions. Lastly, it is possible that genetic information obtained for one purpose may also provide information, for example, about disease risk. The potential impact of such ancillary information should be considered in a broad sense; for example, disease-risk information that is available after the test is conducted versus the use of CYP2D6 drug-metabolizing enzyme polymorphism data to help guide therapy for drugs other than the original one for which the test was obtained. Thus, there could be potential benefits but also harms (Henrikson et al., 2007).

Cost-effectiveness studies of both disease-risk and pharmacogenomic tests in some cases may benefit from more complex analytical frameworks and simulation methods. However, it generally is not possible, due to statistical power limitations, to identify specific risks for individual patients. Rather, in genetic association studies, patients tend to be categorized into larger subgroups to improve statistical power. For example, the breast cancer disease recurrence risk score (RS) derived from a 21-gene profile with the Oncotype Dx test could be utilized in a similar fashion to the Framingham models that predict the risk of a cardiovascular event based on non-genetic factors. In general, when more complex modeling approaches are employed, the added value should be explained in addition to justification of data sufficiency.

Cost-Effectiveness Framework and Drivers

We have previously developed a set of criteria, based on a formal cost-effectiveness framework, for evaluating the potential clinical and economic benefits of genomic tests (Table 20.1) (Flowers and Veenstra, 2004). The key aspects of these criteria are highlighted in the following paragraphs, and a more in-depth evaluation of the epidemiologic considerations in bringing pharmacogenomics to clinical practice is presented.

Assessing the incremental cost-effectiveness of a pharmacogenomic strategy involves evaluating factors that are common to the evaluation of all screening strategies, as well as factors that are specific to genetic testing. These include factors associated with the genetic variation of interest, the genetic test, the disease state, and the treatment options. Questions to consider in assessing the cost-effectiveness of a pharmacogenomic treatment strategy include:

- What is the frequency of the genetic variation?
- How closely is the variation linked to a consistent phenotypic drug response?
- Are there other significant influences on drug response such as diet, disease, or drug interactions?
- What are the sensitivity and specificity of the genomic test?
- How prevalent is the disease of interest?
- What are the characteristic outcomes associated with the disease with and without treatment?
- How does the pharmacogenomic strategy alter these outcomes?
- What alternative treatment options are available?
- How effective are current monitoring strategies for preventing severe adverse drug reactions and predicting drug response?

Pharmacogenomics is more likely to be cost-effective when: (1) the polymorphism under consideration is prevalent in the population and has a high degree of penetrance; (2) genetic testing is highly sensitive and specific, and less costly alternative tests that could be used to individualize therapy are not readily available; (3) the disease state involves outcomes with significant morbidity or mortality without treatment; and (4) the treatment involves significant outcomes and/or costs that can be impacted by genotype-individualized therapy.

Case Study: Gene-Expression Profiling and Breast Cancer Treatment

Breast cancer is the leading incident cancer among women of all major ethnicities in the United States and is the second highest

TABLE 20.1 Factors that influence the cost-effectiveness of genomic testing strategies

	Factors to assess	Features that favor cost-effectiveness
Gene	Prevalence Penetrance	• Variant allele is relatively common • Gene penetrance is high
Test	Sensitivity, specificity, cost	• High specificity and sensitivity • A rapid and relatively inexpensive assay is available
Disease	Prevalence	• High disease prevalence in the population
	Outcomes and economic impacts	• High untreated mortality • Significant impact on quality of life • High costs of disease management using conventional methods
Treatment	Outcomes and economic impacts	• Reduction in adverse effects that significantly impact quality of life or survival • Significant improvement in quality of life or survival due to differential treatment effects • Monitoring of drug response is currently not practiced or difficult • No or limited incremental cost of treatment with pharmacogenomic strategy

Adapted from Flowers and Veenstra (2004).

source of cancer mortality. Adjuvant chemotherapy has been shown to increase recurrence-free and overall survival, but also may produce significant toxicity such as alopecia, nausea/vomiting, and myelosuppresion, and may lead to longer-term complications such as permanent ovarian failure in pre-menopausal patients (Shapiro and Recht, 1994, 2001). Current NIH clinical guidelines (Eifel et al., 2001) recommend adjuvant chemotherapy for women with tumors larger than 1 cm or lymph node involvement. Additionally, tumor markers such as HER2 and histologic grade are used for risk assessment (Bast et al., 2001; NCCN, 1999). Despite widespread use, these criteria are imprecise predictors of distant recurrence (Sauter and Simon, 2002).

Gene-expression profiling (GEP) utilizing DNA microarrays (van de Vijver et al., 2002) or RT-PCR (Paik et al., 2004) has been proposed as an alternative approach to identify patients for adjuvant chemotherapy, (King and Sinha, 2001) potentially sparing low-risk patients from this treatment. There are two gene-expression profiles currently marketed for clinical use in breast cancer. One of these profiles, MammaPrint® (Agendia), was developed by van't Veer and colleagues at the Netherlands Cancer Institute (van de Vijver et al., 2002). The other test, Oncotype Dx™, was developed by Genomic Health, Inc. in the United States (Paik et al., 2004). The assay marketed by Agendia utilizes a 70-gene microarray-based profile performed on fresh frozen tissue and is intended for patients younger than 55 years with Stage I invasive breast cancer or Stage II node-negative invasive breast cancer. In contrast, the test marketed by Genomic Health employs a 21-gene profile utilizing RT-PCR for expression analysis on paraffin embedded tissue and is intended for patients with node-negative, estrogen receptor positive (ER+) disease. Although gene-expression profiling has been proposed as an alternative to clinical guidelines to identify appropriate patients for adjuvant chemotherapy, the potential long-term clinical and economic outcomes associated with gene-expression profiling are not clear. In other words, "Do the quality of life benefits and cost savings of avoiding chemotherapy outweigh the potential increased risk of recurrence in women not given chemotherapy and cost of the test?"

To address this question, we developed a decision analytic model to evaluate the incremental cost and QALYs of gene-expression profiling using MammaPrint versus NIH clinical guidelines in a hypothetical cohort of pre-menopausal early-stage breast cancer patients (Oestreicher et al., 2005). We assumed patients and doctors would follow test or guideline recommendations, and that chemotherapy response would be similar regardless of risk group, as no data to the contrary were available. Our findings suggested that use of MammaPrint gene-expression profiling could result in an absolute 5% decrease in the proportion of cases of distant recurrence prevented, 0.21 *fewer* QALYs, but a cost savings of $2800 (Table 20.2). Regardless of the test cutoff used to identify a poor prognosis tumor, the MammaPrint gene-expression assay did not produce equal or greater QALYs than NIH guidelines.

These findings are fairly striking, and importantly, what do they imply? First, test performance is very important when a pharmacogenomic test is being used to recommend against treatment that is standard of care. In this case, because of test sensitivity of 84% for MammaPrint versus 98% for NIH criteria, some women who will progress will be categorized as low risk and not recommended for chemotherapy. Secondly, the efficacy of treatment in women categorized as low risk versus high risk can have a significant impact. In other words, prognostic ability

Table 20.2 Performance, costs and outcomes of gene expression profiling versus NIH guidelines

Outcome	Gene Expression Profiling	NIH Guidelines	Difference
Sensitivity	84%	98%	−14%
Specificity	51%	5%	+46%
Proportion of women treated with chemotherapy	61%	96%	−35%
Proportion of distant recurrences prevented	29%	34%	−5%
Costs	$29,754	$32,636	−$2882
Quality-adjusted life years	9.86	10.08	−0.21

Adapted from Oestreicher et al. (2005).

to evaluate disease risk is important, but so is predictive ability for treatment response. In an evaluation of the clinical and economic utility of a test, these two aspects of test performance, disease-risk and treatment response, should be clearly delineated. In some cases, as illustrated here, prognostic (disease-risk) information alone may not be sufficient to justify utilization of a genomic test.

The Oncotype Dx assay has been evaluated for its ability to predict chemotherapy outcomes in addition to disease recurrence risk. Paik et al. reported in a retrospective study that women with a low RS are also unlikely to respond to chemotherapy (relative risk, 1.31; 95% CI, 0.46–3.78) (Paik et al., 2006). The impact of this effect was incorporated in a cost-effectiveness analysis conducted by Hornberger and colleagues (2005). Using a similar methodological approach to Oestreicher et al., the author developed a Markov model and forecast overall survival, costs, and cost-effectiveness of using the Oncotype Dx RS in patients classified as having low or high risk of distant recurrence based on National Comprehensive Cancer Network (NCCN) clinical guidelines. In the analysis, 8% of patients were classified as having low risk of distant recurrence by NCCN guidelines, and the RS reclassified 28% of these patients to an intermediate/high-risk group. The remaining 92% of patients were classified at high risk of distant recurrence by NCCN guidelines and the RS reclassified 49% of these to a low-risk group. Overall, use of the Oncotype Dx test was predicted to *increase* quality-adjusted survival by 0.09 years and reduce overall costs by $2000. The cost-effectiveness was most influenced by the propensity to administer chemotherapy based on test results, and by the proportion of patients at low risk as defined by NCCN guidelines. The author concluded that if applied appropriately, the test is predicted to increase quality-adjusted survival and save costs.

These cost-effectiveness studies served several valuable purposes. Most importantly, they clarified that without the ability to predict response to chemotherapy, these gene-expression profiles may not provide justifiable improvements (if any) in overall patient outcomes in a cost-effective manner. Thus, important areas for subsequent validation work have been highlighted. Furthermore, such analysis can help inform reimbursement decisions (Watkins et al., 2007). Thus, for example, with the Oncotype Dx test, assuming the prognostic and particularly the predictive validity of the test hold, the test may offer not only a cost-effective intervention, but potentially a cost-saving one. Another issue that remains to be addressed, preferably in "real-world" (non-controlled) settings, is the concordance of test results with actual treatments received. For example, a recent study indicated that a majority of sampled physicians and patients found the test results useful, but there was an absolute decrease in the proportion of patients choosing chemotherapy of only 2% (Lo et al., 2007). These treatment patterns may change as clinicians and patients gain experience in the use of genomic test results in decision-making.

Of note, in an effort to validate these retrospective findings, the first US-based phase III randomized controlled clinical trial – The Trial Assigning IndividuaLized Options for Treatment (Rx), or TAILORx, comparing the efficacy of Oncotype Dx for choosing women receiving adjuvant chemotherapy – was launched on May 23, 2006, to examine whether genes that are frequently associated with risk of recurrence for women with early-stage breast cancer can be used to assign patients to the most appropriate and effective treatment [http://www.cancer.gov/clinicaltrials/digestpage/TAILORx, accessed 11/9/06]. This study will provide valuable prospective validation of previous retrospective studies, as well as the modeled outcomes from cost-effectiveness studies.

ECONOMIC INCENTIVES AND THE FUTURE OF GENOMIC MEDICINE

Economic Incentives in Drug and Test Development

Eight years have elapsed since the initial sequencing of the human genome, and the number of new genetic tests commonly used in clinical practice is quite small. For example, a recent report from the Royal Society cautions: "Pharmacogenetics is unlikely to revolutionize or personalize medical practice in the immediate future" (The Royal Society, 2005). Robert Califf has argued that achieving this promise will require a major overhaul of the US clinical research enterprise as well as substantial educational efforts (Califf, 2004).

In two recently published papers, Garrison and Austin explored the lack of appropriate economic incentives that may in part be contributing to the current pace, in addition to the inherent challenges in genomics research (Garrison and Austin, 2006, 2007). The translation of the basic science of pharmacogenetics

and other "-omics" biomarkers as applied to drug development and clinical care is occurring in a complex legal, regulatory, and reimbursement environment. Understanding and appropriately shaping this environment is vital for encouraging biomarker research and personalized health care.

Suppliers of new medical technologies – both therapeutics and diagnostics – face a complex and heavily regulated commercial environment. Furthermore, the manner in which diagnostics are reimbursed in most developed countries may provide limited incentives for the development of new genomics-based tests. Both the proponents of genetic tests and the skeptics sometimes fail to appreciate this complexity. Proponents sometimes argue: "How could this not be a good thing?" We would be able to target drugs to the subgroups who respond most favorably and limit the delivery to those most likely to suffer side effects. Skeptics sometimes argue that no pharmaceutical company would have incentive to restrict its market to a smaller subgroup: "Why would they ever want to develop a test to do this?"

Both sides have a valid point to some extent, but neither represents the full complexity of the incentives involved. Garrison and Austin analyzed the incentives in terms of "value creation and capture." Innovative therapies can create aggregate economic value in at least five alternative ways that are not mutually exclusive. First, they may reduce mortality and morbidity compared to current treatments. Second, they may save on average on the costs of treating side effects or the underlying disease. Third, they may result in greater utilization by those afflicted. Fourth, they may encourage better compliance, providing individual patients with more net benefit. Fifth – and this is less recognized and is important for new diagnostic tests – they can make people better off by reducing uncertainty.

The authors developed and analyzed five different scenarios of combining a new genomics-based diagnostic with a therapeutic treatment. They argued that who captures the value created by the combination depends on a number of factors, and that it is the potential for value capture that generates the incentives for genomic test development and marketing. In thinking about incentives in this marketplace, it is important to keep a few key features in mind. First, intellectual property rights are essential to provide a basic incentive for the invention and development for new diagnostic tests and new therapeutics, particularly pharmaceuticals. It is widely known and appreciated that new innovative drugs receive temporary (20-year) patents that limit the ability of competitors to enter since they can't just copy the innovative compound and would have to develop and fully test a new molecule. It is less appreciated that effective patent life is 8–12 years, as it takes years to perform the trials that regulators review for licensing. A second key feature is that once a product is on the market and has a price, it can be difficult to change that price. Indeed, in some countries, "administered pricing" is used to set the price based on reference products or prices in other countries. The price granted in those situations may only relate in a limited way to the value created, and they may not be allowed to change over time. This is particularly true in Europe, whereas there is some ability to change pharmaceutical prices over time in the United States.

Diagnostics differ from drugs in that their prices are even more heavily controlled in most markets, as initial price is often pegged to perceived costs of provision rather than additional value created. This can be called "cost-based" pricing and reimbursement as opposed to "value-based" pricing and reimbursement. In effect, suppliers of new genomics-based diagnostics are likely to capture less of the incremental value they create.

Given these marketplace realities and constraints, Garrison and Austin argue that the incentives for value creation through a linked genomics-based diagnostic and therapeutic will depend on these factors as well as whether the therapy is already on the market. For their hypothetical exercise, they posited that a new diagnostic is developed in a market where only 20% of patients truly benefit from a treatment, but heretofore they could not be identified. Eliminating the 80% of users who don't benefit would be a great benefit to payers but would be damaging to the pharmaceutical company if they could not raise price for the 20% of responders. On the other hand, it should clear that if the test were available before the drug was launched, the company might have been able to charge a price nearly five times as high and still capture the same amount of value.

In principle, if the innovative diagnostic enters the market after that drug has a price, then the diagnostic manufacturer might be able to capture this value, creating a powerful incentive for developing the diagnostic. However, in a regulated, cost-based pricing and reimbursement environment for diagnostics, there will not be this strong incentive. Hence, if the drug manufacturer can't easily increase price for the 20% who respond, and if the potential diagnostic manufacturer can't capture the cost savings because of price controls, there will be a limited incentive to develop tests for already marketed products.

The authors point out that if the diagnostic and therapeutic could be developed in tandem, even more total economic value would be created due to the additional reduction in uncertainty. The drug manufacturer would be in a better position to capture this value, given that they operate in a relatively value-based pricing environment. However, drug development is a challenging and uncertain scientific enterprise in and of itself, and adding the complexity of companion diagnostic may not be that appealing unless the scientific rationale is already well developed.

In addition to the scientific and economic barriers highlighted above, two other factors deserve mention that might be inhibiting the development of pharmacogenetics-based test-drug combinations. One factor is the high cost of the basic research that is needed to validate genetic markers. Financing this remains a question: What should fall to the public sector, how much should the private sector contribute, and what is best done in partnership are under active debate. Second, approval and reimbursement for drugs in the United States and EU require greater levels of evidence than is customary for new diagnostic tests. Some would argue that the lower evidentiary requirements for regulatory approval of tests have discouraged the development of better clinical data and that payers practice "cost-based" reimbursement in part because of this lack of evidence on clinical and economic value. In contrast, both the AMCP Format for

Formulary Submissions guidelines in the United States and the requirements for submissions to the National Center for Clinical and Health Excellence (NICE) in the United Kingdom request economic models that synthesize the clinical and cost evidence to assess value added of new drugs. Similar standards and mechanisms do not exist for diagnostics, although the discussion is beginning. But asking for more evidence means raising the costs of developing – and particularly validating – new diagnostics. What are the incentives for diagnostic manufacturers if no additional rewards are forthcoming through the reimbursement system? Below we discuss a potential approach for incorporating evidence assessment into reimbursement decisions for genomic tests.

ESTABLISHING VALUE-BASED REIMBURSEMENT FOR GENOMIC TECHNOLOGIES

Rationale

Laboratory tests are an important yet often overlooked segment of health care and the health economy. For example, it has been estimated that as much as $56 billion was spent on laboratory diagnostic services in 2005 (Blue Cross Blue Shield Association, 2005). More importantly, laboratory tests initiate a cascade of decisions regarding further testing, prevention or treatment-decisions that ultimately determine health outcomes and costs of care. A recent report estimated that although diagnostics account for 1.6% of the Medicare total costs, they influence 60–70% of downstream treatment-decisions (The Lewin Group, 2005). Genomic testing for mutations, polymorphisms, and haplotypes expands the laboratory diagnostics market in many areas, including pregnancy and neonatal testing, predictive testing for disease susceptibility, and pharmacogenomic testing. As noted above, gene-expression profiling of tumors is a burgeoning field that includes both predictive and pharmacogenomic characteristics.

Historically, the evidence supporting the validity and clinical utility of laboratory tests has been limited (Feinstein 2002; Reid et al., 1995; Weinstein et al., 2005). The pathway to the marketing of tests does not require the same data as therapeutics, and technology assessments of laboratory tests have not undergone the revolutionary changes in evidence review that has occurred for drugs (Neumann, 2004). Instead, payment for tests is linked to an antiquated coding system, as noted above (Raab and Logue, 2001). Recently, we presented a rationale and outline for re-structuring the way laboratory tests are evaluated and reimbursed (Ramsey et al., 2006). We summarize those concepts below.

Process

We propose that coverage and reimbursement for laboratory tests, including genomic tests, should move toward an evidence- and value-based approach, using some of the tools that have been adopted for pharmaceutical assessment by many US health care payers.

Developing a Language for Describing Benefit

Test manufacturers, laboratory service providers, and health insurance plans will benefit from standardizing the way evidence supporting new laboratory tests is presented. Methodological standards for the evaluation of diagnostic tests have been published (Reid et al., 1995), and a useful framework has been proposed to evaluate diagnostic technologies (Table 20.3) (Fryback and Thornbury, 1991). While many agree about the value of using these domains to evaluate tests, including genomic tests, there is less agreement on how much evidence is necessary for an insurance coverage decision.

Considering the domains from Table 20.1 when applied to genomic tests, it is important to evaluate the incremental impact of a test; that is, the improvement that the new test provides over-current (non-genomic) strategies (Garrison and Austin, 2006). Since prospective trials directly comparing new laboratory tests with established diagnostic strategies are uncommon (particularly those evaluating their impact on patient outcomes), decision-analytic modeling techniques can aid evaluations (Buxton et al., 1997). Models help frame questions, provide transparent mechanisms for stating hypotheses about cause and effect, highlight deficiencies in clinical data, and force decision-makers to make explicit judgments about values for data that are used to inform the model. Published standards for models are readily accessible to decision-makers who wish to assess their quality (Weinstein et al., 2003).

TABLE 20.3 Hierarchy of diagnostic evaluation

Level	Characteristic	Description
1	Technical feasibility and optimization	Ability to produce consistent results
2	Diagnostic accuracy	Sensitivity, specificity, positive predictive value, negative predictive value
3	Diagnostic thinking impact	Percentage of times physicians' estimated probability of a diagnosis changes following the test result
4	Therapeutic choice impact	Proportion of times planned therapeutic strategy changes following test results
5	Patient outcome impact	Percentage of patients who improve with the test versus those who improve without the test
6	Societal impact	Cost-effectiveness

Adapted from Fryback and Thornbury (1991).

A Definition of Value for Genomic Tests

We define value for genomic tests as it is defined for other health technologies: the intervention provides an overall benefit to the patient at an acceptable cost; or in other words is cost-effective. There are four well-recognized criteria for identifying an intervention cost-effective: (1) Less costly and at least as effective; (2) More effective and more costly, with the added benefit worth the added cost; (3) Less effective and less costly, with the added benefit of the alternative not worth the added cost; and (4) Cost saving with an outcome equal to or better than that of the alternative. Assessing value for tests can be difficult because tests are intermediate steps in the treatment pathway. The advantage of a cost-effectiveness framework is that it allows flexibility, because value for genomic tests can be defined narrowly (e.g., least expensive way to make a diagnosis) or broadly (improves survival at an acceptable added cost).

A Format for Dialogue Among Genomic Test Manufacturers, Providers, and Payers

When health insurers and test representatives come together to discuss genomic tests, the process should be transparent, and adopting a standard format would help clarify expectations and improve the decision-making process. A useful template for tests has been developed by the AMCP for the evidence-based evaluation of drugs. More than 50 public and private health insurers covering over 100 million lives have adopted the AMCP format (Fry et al., 2003; Neumann, 2004). A template for manufacturer's reporting of clinical and economic information regarding laboratory tests is shown in Figure 20.1. Since the clinical utility of genomic tests is most often similar to other laboratory tests, the template accommodates differences in evidence that are typically available for new genomic tests.

Implementation

Some health plans might have a designated organizational unit that evaluates genomic tests, providing a structure for soliciting and reviewing manufacturer's products. In the case of pharmaceutical products, that unit is the Pharmacy and Therapeutics (P&T) committee. Health plan staff working with P&T committees usually are receptive to receiving information from manufacturers or service laboratories about new products. Indeed, the "unsolicited request" process pioneered by the AMCP was designed to create a structure for dialogue between manufacturers and payers. A similar scheme would improve transparency and the flow of information for new genomic tests.

P&T committee members are often not health plan employees. Having such a quasi-independent group evaluate novel laboratory tests may improve the credibility of the decision process in the eyes of manufacturers, clinical laboratories, physicians, and patients. Still, maintaining committees is costly, and may not be justified given the relatively low volume of genomic tests that are introduced annually. One option is to fold the test evaluation process into the existing P&T structure. Alternatively, payers could hire consultants to evaluate select genomic tests and make recommendations regarding coverage and reimbursement. Payers should be timely, in both coverage decisions and setting reimbursement levels, and when decisions are made, they should be supported with a rationale. In cases where requests for coverage are denied, such information allows manufacturers to design studies or collect other data that address concerns regarding the quality or content of the information supporting the product.

Genomic tests are developed under a different regulatory structure than pharmaceuticals and have a unique and complex function in medical care. Novel genomic tests often come to market with little information supporting their role in clinical decision-making or evidence regarding their impact on patient outcomes. We posit that patient care and outcomes will improve when there is a structured process for evidence generation and discussion between payers, manufacturers, and marketers. While it is unlikely and perhaps not necessary that the evaluation process for genomic tests will equal what is required for drugs, we can move much further toward a system that supports better gathering and sharing of high-quality evidence. Test manufacturers, clinical laboratories, payers, and clinicians all must play a role in this process.

1. Product Information
- 1.1. Product description.
- 1.1.1. Place of the product in therapy.
- 1.1.2. Disease description. The disease description should include the disease and characteristics of the patients in the target population.

2. Supporting Clinical and Economic Information
- 2.1. Evidence-table spreadsheets of all published and unpublished clinical studies.
- 2.2. Outcomes studies and economic evaluation supporting data
- 2.2.1. Evidence-table spreadsheets of all published and unpublished outcomes studies.

3. Cost-effectiveness Modeling Report
- 3.1. Model overview. We recommend that producers and users of modeling studies subscribe to the sound guidance provided by the ISPOR Good Practice Modeling Principles.

4. Supporting Information
- 4.1. References contained in dossiers.
- 4.2. References for economic models.

Figure 20.1 Evidence and transparency standard to support coverage and reimbursement for diagnostic, therapeutic, and genetic testing (from Ramsey et al., 2006).

ECONOMIC CHALLENGES

The primary economic challenges facing genomic medicine can be surmised as the following: (1) providing evidence of the value of genomics and (2) providing economic rewards/incentives commensurate with the value added. More specifically, as outlined in a recent report developed on the implications of pharmacogenomics for the pharmaceutical and biotechnology industries, there are several key challenges for genomic medicine (Garrison et al., 2007):

- Regulatory pathways have not yet been optimized to encourage the co-development of diagnostics and therapeutics.

- Current economic incentives – as reflected in our intellectual property and reimbursement systems for diagnostics and drugs – are generally not structured to reward appropriately and consistently innovative value creation for drugs, diagnostics, and pharmacogenomics-based targeted regimens.
- The integration of pharmacogenomic diagnostic development with pharmaceutical development is difficult because of differences in the underlying business and translational science models of the two sectors.
- Genomics technologies are perceived to raise ethical, legal, and social issues to such a degree that a special National Institutes of Health program was established to address them; although specific genomic applications may not always involve such issues, this suggests the broad range of stakeholders that will be involved in the public debates.
- Stakeholder literacy about genomics is limited, and positions on public policy issues are not yet clearly defined.

What can be done to help achieve these goals? We suggest several policy areas for focus (Garrison et al., 2007):

- Pharmaceutical and biotechnology companies will need to add a systematic evaluation of potential pharmacogenomic and other biomarkers as part of their due diligence research and development processes.
- Substantial federal government support for basic research generally will be critical before translational private research activities are viable.
- The pharmaceutical and biotechnology industries will be major beneficiaries of this basic research, and should participate actively in the public discussion of priorities.
- Companies will need to provide a rationale to regulators why they have or have not included genomics or other biomarkers in their clinical trial development programs.
- The pharmaceutical, biotechnology, and diagnostic industries have not taken a unified or proactive position on appropriate regulatory processes and initiatives, and it may not be possible to reach a consensus. Still, it may be a good time to begin a policy review and discussion.
- For pharmacogenomics-targeted pharmaceuticals to have greater commercial viability, the pharmaceutical, biotechnology, and diagnostic industries must engage in the public policy debate on national coverage and reimbursement issues for such drugs and tests.

Will cost-effectiveness analysis by itself speed the development of genomic medicine? This is not inherently likely. The fundamental challenges facing genomic medicine at this point are (1) the identification and validation of associations between genetic variation and clinically meaningful outcomes and (2) evaluation of interventions based on genomic information in prospective, comparative trials. However, cost-effectiveness analysis can help guide this process in several important ways:

- By identifying genomic technologies that have potential value in early stages of drug and test development, thus aiding research portfolio optimization.
- By providing a framework for value-based reimbursement of genomic technologies, thus conferring appropriate incentives to the market.
- By identifying areas where future research would provide significant value to the health care system even once a technology has been developed.

Thus, quantitative economic analyses in combination with scientific research and policy analysis can help optimize the benefits of genomic medicine to society. In summary, genomic medicine offers significant hope for fundamental improvements in health care outcomes. These gains can likely be achieved in a cost-effective manner, but the challenges of providing evidence of and reward for value will prove to be critical in the coming decade.

2009 UPDATE

There have been a number of recent publications addressing methodological issues pertaining to the economic evaluation of genetic testing, incentives for development and reimbursement, as well as studies of specific genetic testing strategies. Deverka and McLeod (2008) propose that two levels of evidence are needed before pharmacogenetic tests can have a major impact on clinical practice: genetic association data and outcomes data – the latter including improvements in patient outcomes and cost-effectiveness. The authors emphasize the importance of focusing on the latter evidence to facilitate the translation of pharmacogenomic technologies. Garrison and colleagues provide an overview of public policy issues related to the development and commercialization of personalized medicine, some of which is covered in this chapter (Garrison et al., 2008). Phillips et al. (2008) review existing evidence on utilization, preferences, and economic value of two case studies: HER2/neu antigen testing with trastuzumab treatment and genetic testing for Lynch syndrome.

Grosse, Wordsworth, and Payne address the challenges of valuing genetic testing – particularly in regard to non-clinical outcomes, such as impacts on patient quality of life, preferences, and 'information for information's sake' (Grosse et al., 2008). The authors discuss the relative merits of different economic measures and methods and conclude that because important outcomes of genetic testing do not fit easily within traditional measures of health, cost-effectiveness analysis do

not necessarily capture the full range of outcomes of genetic testing that are important to decision makers and consumers. Payne and colleagues conducted a systematic review of validated outcome measures of clinical genetics services (Payne et al., 2008). Thirty genetics-specific measures were identified; however, no single validated outcome measure encompassed all potential patient benefits, and the authors conclude that additional research in this area is needed. Vetger and colleagues performed a systematic review and critical evaluation of recent pharmacoeconomic analyses in pharmacogenomics (Vetger et al., 2008). Twenty studies were identified, the majority of which reported that genetic screening was cost-effective or cost-saving. However, several studies failed to provide sufficient evidence for an association between genotype and phenotype. The authors provide a list of recommendations for good pharmacoeconomic practice in pharmacogenomics.

Worth noting, recent research has provided greater clarification of thresholds for cost-effectiveness in the United States. Braithwaite and colleagues identified lower and upper bounds for the value society places on a quality-adjusted life-year (QALY) (Braithwaite et al., 2008). A lower bound was inferred by estimating the incremental cost-effectiveness of recent versus pre-"modern era" medical care in the United States. To infer an upper bound, they estimate the incremental cost-effectiveness of unsubsidized health insurance versus self-pay for nonelderly adults without health insurance. The authors estimate the value of a QALY between $109,000 and $297,000. The authors, and an accompanying editorial (Weinstein, 2008), conclude it is very unlikely that $50,000 per QALY is consistent with societal preferences in the United States.

Several evaluations of specific genetic testing applications are worth noting. First, a federally funded evidence review of gene expression profiling tests in breast cancer was reported (Marchionni et al., 2008). The review included an assessment of the evidence for the cost-effectiveness of these tests and concluded that "since the overall body of evidence is inconclusive about the economic outcomes associated with use of breast cancer gene expression tests, this is an area that will require further investigation." Schackman and colleagues conducted a cost-effectiveness study of HLA-B*5701 screening to guide antiretroviral selection for treatment of HIV (Schackman et al., 2008). The analysis was informed by data from the Prospective Randomized Evaluation of DNA Screening in a Clinical Trial study. The authors report that HLA-B*5701 testing added 0.04 quality-adjusted months (1.2 days) at an incremental cost of $110, and that testing remained the preferred strategy only if abacavir-based treatment had equal efficacy and cost less per month than tenofovir-based treatment.

Holland, Huston, and Noyes conducted a study to determine at what risk of mutation it is cost-effective to test women for BRCA1/2 mutations (Holland et al., 2008). The evaluation included estimates of health-care costs, life expectancy, likelihood of obtaining a mastectomy or oophorectomy, and patient preferences for treatment and certainty about their BRCA1/2 status. The costs and effectiveness of both the test and no-test strategies were very similar even when there was a small probability of mutation. Heitjan and colleagues explored the potential cost-effectiveness of a range of smoking cessation drug treatments, including varenicline, transdermal nicotine, bupropion, and the use of a genetic test to choose between transdermal nicotine and bupropion (Heitjan et al., 2009). The authors found that only under favorable assumptions was the genetically tailored approach competitive. Finally, Carlson and colleagues evaluated the potential cost-utility of EGFR testing to guide non-small cell lung cancer treatment. The authors found that that EGFR pharmacogenomic testing has the potential to provide clinically meaningful improvements in patient outcomes at a value similar to currently approved therapies, and additional research in this area is warranted (Carlson et al., 2009).

2009 UPDATE REFERENCES

Braithwaite, R.S., Meltzer, D.O., King, J.T., Jr., Leslie, D. and Roberts, M.S. (2008). What does the value of modern medicine say about the $50,000 per quality-adjusted life-year decision rule? *Med Care* 46(4), 349–356.

Carlson, J.J., Garrison, L.P., Ramsey, S.D. and Veenstra, D.L. (2009). The potential clinical and economic outcomes of pharmacogenomic approaches to EGFR-tyrosine kinase inhibitor therapy in non-small-cell lung cancer. *Value Health* 12(1), 20–27

Deverka, P.A. and McLeod, H.L. (2009). Harnessing economic drivers for successful clinical implementation of pharmacogenetic testing. *Clin Pharmacol Ther* 84(2), 191–193.

Garrison, L.P., Jr., Carlson, R.J., Carlson, J.J., Kuszler, P.C., Meckley, L.M. and Veenstra, D.L. (2008). A review of public policy issues in promoting the development and commercialization of pharmacogenomic applications: Challenges and implications. *Drug Metab Rev* 40(2), 377–401.

Grosse, S.D., Wordsworth, S. and Payne, K. (2008). Economic methods for valuing the outcomes of genetic testing: beyond cost-effectiveness analysis. *Genet Med* 10(9), 648–654.

Heitjan, D.F., Asch, D.A., Ray, R., Rukstalis, M., Patterson, F. and Lerman C. (2008). Cost-effectiveness of pharmacogenetic testing to tailor smoking-cessation treatment. *Pharmacogenomics J* 8(6), 391–399.

Holland, M.L., Huston, A. and Noyes, K. (2009). Cost-effectiveness of testing for breast cancer susceptibility genes. *Value Health* 12(2), 207–216.

Marchionni, L., Wilson, R., Marinopoulos, S. et al. (2008). *Impact of Gene Expression Profiling Tests on Breast Cancer Outcomes*. Evidence Report/Technology Assessment No. 160. AHRQ Publication No. 08-E002. Agency for Healthcare Research and Quality, Rockville, MD.

Payne, K., Nicholls, S., McAllister, M., Macleod, R., Donnai, D. and Davies, L.M. (2008). Outcome measurement in clinical genetics

services: a systematic review of validated measures. *Value Health* 11(3), 497–508.

Phillips, K.A., Liang, S.Y. and Van Bebber, S. (2008). Challenges to the translation of genomic information into clinical practice and health policy: Utilization, preferences and economic value. *Curr Opin Mol Ther* 10(3), 260–266.

Schackman, B.R., Scott, C.A., Walensky, R.P., Losina, E., Freedberg, K.A. and Sax, P.E. (2008). The cost-effectiveness of HLA-B*5701 genetic screening to guide initial antiretroviral therapy for HIV. *AIDS* 22(15), 2025–2033.

Vegter, S., Boersma, C., Rozenbaum, M., Wilffert, B., Navis, G. and Postma, M.J. (2008). Pharmacoeconomic evaluations of pharmacogenetic and genomic screening programmes: A systematic review on content and adherence to guidelines. *Pharmacoeconomics* 26(7), 569–587.

Weinstein, M.C. (2008). How much are Americans willing to pay for a quality-adjusted life year? *Med Care* 46(4), 343–345.

REFERENCES

Bast, R.C., Jr., Ravdin, P. et al. (2001). 2000 update of recommendations for the use of tumor markers in breast and colorectal cancer: Clinical practice guidelines of the American Society of Clinical Oncology. *J Clin Oncol* 19, 1865–1878.

Blue Cross Blue Shield Association. (2005). Medical technology as a driver of healthcare costs: Diagnostic imaging. Retrieved August, 2005, from http://www.bcbs.com/coststudies/reports/Medical_Tech_Drivr_Rept_10.pdf.

Buxton, M.J., Drummond, M.F. et al. (1997). Modelling in economic evaluation: An unavoidable fact of life. *Health Econ* 6(3), 217–227.

CADTH (2007). Canadian Agency for Drugs and Technologies in Health.

Califf, R.M. (2004). Defining the balance of risk and benefit in the era of genomics and proteomics. *Health Aff (Millwood)* 23(1), 77–87.

Detsky, A.S., Naglie, G. et al. (1997). Primer on medical decision analysis: Part 2—Building a tree. *Med Decis Making* 17(2), 126–135.

Drummond, M., Dubois, D. et al. (1999). Current trends in the use of pharmacoeconomics and outcomes research in Europe. *Value Health* 2(5), 323–332.

Eifel, P., Axelson, J.A. et al. (2001). National Institutes of Health Consensus Development Conference Statement: Adjuvant therapy for breast cancer, November 1–3, 2000. *J Natl Cancer Inst* 93, 979–989.

Feinstein, A.R. (2002). Misguided efforts and future challenges for research on diagnostic tests. *J Epidemiol Community Health* 56(5), 330–332.

Flowers, C.R. and Veenstra, D. (2004). The role of cost-effectiveness analysis in the era of pharmacogenomics. *Pharmacoeconomics* 22(8), 481–493.

Fry, R., Avey, S. et al. (2003). The Academy of Managed Care Pharmacy Format for Formulary Submissions: An evolving standard–a Foundation for Managed Care Pharmacy Task Force report. *Value Health* 6(5), 505–521.

Fryback, D.G. and Thornbury, J.R. (1991). The efficacy of diagnostic imaging. *Med Decis Making* 11(2), 88–94.

Garber, A.M. and Phelps, C.E. (1997). Economic foundations of cost-effectiveness analysis. *J Health Econ* 16(1), 1–31.

Garrison, L.P., Jr. and Austin, M.J.F. (2006). Linking pharmacogenetics-based diagnostics and drugs for personalized medicine: Scientific and economic challenges. *Health Aff (Millwood)* 25(5), 1281–1290.

Garrison, L.P. and Austin, M.J.F. (2007). The economics of personalized medicine: A model of incentives for value creation and capture. *Drug Information journal* 41, 501–509.

Garrison, L.P., Veenstra, D.L. et al. (2007). Backgrounder on Pharmacogenomics for the Pharmaceutical and Biotechnology Industries: Basic Science, Future Scenarios, Policy Directions, University of Washington, Pharmaceutical Outcomes Research and Policy Program.

Henrikson, N.B., Burke, W. et al. (2007). Ancillary risk information and pharmacogenetic tests: Social and policy implications. *Pharmacogenomics J* 8, 85–89.

Hornberger, J., Cosler, L.E. et al. (2005). Economic analysis of targeting chemotherapy using a 21-gene RT-PCR assay in lymph-node-negative, estrogen-receptor-positive, early-stage breast cancer. *Am J Manag Care* 11(5), 313–324.

King, H.C. and Sinha, A.A. (2001). Gene expression profile analysis by DNA microarrays: Promise and pitfalls. *JAMA* 286, 2280–2288.

Lo, S., Norton, J. et al. (2007). Prospective multi-center study of the impact of the 21-gene Recurrence Score (RS) assay on medical oncologist (MO) and patient (pt) adjuvant breast cancer (BC) treatment selection. *Am Soc Clin Oncol*. Comment: Chicago

NCCN (1999). Update: NCCN practice guidelines for the treatment of breast cancer. National Comprehensive Cancer Network. *Oncology (Huntington)* 13, 41–66.

Neumann, P.J. (2004). Evidence-based and value-based formulary guidelines. *Health Aff (Millwood)* 23(1), 124–134.

Neumann, P. (2007). The Cost-Effectiveness Analysis Registry [Internet], Tufts-New England Medical Center, ICRHPS.

NICE (2007). National Institute for Health and Clinical Excellence.

Oestreicher, N., Ramsey, S.D. et al. (2005). Gene expression profiling and breast cancer care: What are the potential benefits and policy implications? *Genet Med* 7(6), 380–389.

Paik, S., Shak, S. et al. (2004). A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 351(27), 2817–2826.

Paik, S., Tang, G. et al. (2006). Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol* 24(23), 3726–3734.

PBAC. (2007). Pharmaceutical Benefits Advisory Committee. Retrieved 7/10/07, from http://www.health.gov.au/internet/wcmspublishing.nsf/Content/health-pbs-general-listing-committee3.htm#esc.

Raab, G.G. and Logue, L.J. (2001). Medicare coverage of new clinical diagnostic laboratory tests: The need for coding and payment reforms. *Clin Leadersh Manag Rev* 15(6), 376–387.

Ramsey, S.D., Clarke, L. et al. (2001). Cost-effectiveness of microsatellite instability screening as a method for detecting hereditary nonpolyposis colorectal cancer. *Ann Intern Med* 135(8 Pt 1), 577–588.

Ramsey, S.D., Veenstra, D.L. et al. (2006). Toward evidence-based assessment for coverage and reimbursement of laboratory-based diagnostic and genetic tests. *Am J Manag Care* 12(4), 197–202.

Reid, M.C., Lachs, M.S. et al. (1995). Use of methodological standards in diagnostic test research. Getting better but still not good. *JAMA* 274(8), 645–651.

Sauter, G. and Simon, R. (2002). Predictive molecular pathology. *N Engl J Med* 347, 1995–1996.

Shapiro, C.L. and Recht, A. (1994). Late effects of adjuvant therapy for breast cancer. *J Natl Cancer Inst Monogr*, 101–112.

Shapiro, C.L. and Recht, A. (2001). Side effects of adjuvant treatment of breast cancer. *N Engl J Med* 344, 1997–2008.

Spooner, J.J., Gandhi, P.K. et al. (2007). AMCP Format dossier requests: Manufacturer response and formulary implications for one large health plan. *J Manag Care Pharm* 13(1), 37–43.

The Lewin Group. (2005). The Value of Diagnostics Innovation, Adoption and Diffusion into Health Care. Retrieved August, 2005, from http://www.advamed.org/publicdocs/thevalueofdiagnostics.pdf.

The Royal Society (2005). *Personalised Medicine: Hopes and Realities*. The Royal Society, London.

van de Vijver, M.J., He, Y.D. et al. (2002). A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347, 1999–2009.

Watkins, J., Bresnahan, B. et al. (2007). *Determining the Value of Diagnostic/Genetic Testing to Inform Health Plan Decision Making: Oncotype DX as a Case Study*. International Society of Pharmacoeconomics and Outcomes Research, Washington DC.

Weinstein, M.C., O'Brien, B. et al. (2003). Principles of good practice for decision analytic modeling in health-care evaluation: Report of the ISPOR Task Force on Good Research Practices—Modeling Studies. *Value Health* 6(1), 9–17.

Weinstein, M.C., Siegel, J.E. et al. (1996). Recommendations of the Panel on Cost-effectiveness in Health and Medicine. *JAMA* 276(15), 1253–1258.

Weinstein, M.C. and Stason, W.B. (1977). Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med* 296(13), 716–721.

Weinstein, S., Obuchowski, N.A. et al. (2005). Clinical evaluation of diagnostic tests. *Am J Roentgenol* 184(1), 14–19.

CHAPTER

Public Health Genomics

Alison Stewart, Marta Gwinn, Ron Zimmern and Muin Khoury

INTRODUCTION

In the mid-1990s, as the Human Genome Project gathered pace, some within the profession of public health began to realize that, in time, new knowledge and technologies stemming from this endeavor would have profound implications for the health of populations and the organization and delivery of health care services. These developments would mean that public health could no longer focus exclusively on social and environmental determinants of health, but would need to incorporate relevant knowledge from the molecular and cellular life sciences in the development of public health policies and programs.

Public health genomics is a new discipline that brings together genetic and genomic science, genetic epidemiology, and a recognition that scientific and technological advance must go hand in hand with an understanding of its ethical, legal, and social dimensions. It is distinct from "genomic medicine" in that its focus is primarily on populations, health services, and public health programs rather than on individual clinical care.

This chapter outlines the development of public health genomics over the last decade, sets out its key concepts, and describes examples of public health genomics programs and activities. New national, regional, and international networks are being established to further the goals of this new field.

THE DEFINITION OF PUBLIC HEALTH GENOMICS

The early practitioners of public health genetics suggested a variety of definitions for their discipline. At the University of Washington Institute of Public Health Genetics it was defined as "The application of advances in human genetics and molecular biotechnology to improve public health and prevent disease," while in the United Kingdom a modified form of the Acheson definition of public health was adopted, describing public health genetics as "The application of genetics on the art and science of promoting health and preventing disease through the organized efforts of society."

All of the early definitions emphasized that the subject matter of public health genetics was how the health of populations and the practice of public health and clinical medicine were affected by genetic science and technology. They also highlighted the importance of harnessing genetic knowledge for disease prevention; the word "prevention" in this context included not just measures to prevent or delay disease onset (primary prevention), but also, if disease was already present, interventions that enabled early detection and treatment, reduced disability, and delayed progression (secondary and tertiary prevention). The UK definition also made it clear that practitioners of public health genetics needed, as well as a working knowledge of genetic science, an understanding of the diverse views within society about genetics and its consequences, and an ability to work across a range of disciplines and cultures.

As the field broadened its outlook beyond the traditional connotations of "genetics" and began to think in terms of genomics, a new definition became necessary. As a result of the 2005 Bellagio workshop, public health genomics is now defined as: "The responsible and effective translation of genome-based knowledge and technologies for the benefit of population health" (Burke et al., 2006; Stewart, 2006).

The Bellagio definition makes a number of important points. By using the term "genome-based" it indicates that the

Copyright © 2010, Elsevier Inc.
All rights reserved.

scope of the knowledge base includes not only genes but also their protein products, the metabolites synthesized by those proteins, and the complex molecular and cellular interactions that make up a biological system. It also includes bioinformatics: the management and analysis of large amounts of biological information using advanced computing techniques. "Technologies" are mentioned explicitly in the definition because biotechnological developments such as drug delivery platforms, high-throughput sequencing technologies, and microarray applications are vital to the clinical exploitation of genomic science.

The adjectives "responsible" and "effective" convey two concepts: firstly, that any clinical tests or interventions resulting from genomic research must be fully validated before they are implemented and, second, that this evidence base must include due consideration of their ethical, legal, and social implications (often shortened to the acronym ELSI) (Clayton, 2003; Ojha and Thertullien, 2005).

Finally, the Bellagio definition emphasizes that public health genomics focuses on the health and health care of populations. This may at first sight seem to be at odds with the individual or familial nature of genomic information and the prospects for a future era of "personalized medicine." However, public health genomics recognizes that populations are not genetically homogeneous and that programs and policies incorporating differences in individual susceptibility to disease and response to treatment offer new opportunities that are complementary to the traditional "one size fits all" approach of public health.

KEY CONCEPTS IN PUBLIC HEALTH GENOMICS

Genes and Environmental Factors as Determinants of Health

Implicit in the various definitions of public health genomics is the recognition that all human characteristics – including susceptibility to disease – result from the combined effects of genes and environment. Figure 21.1 is a conceptual representation of determinants of health, showing the complex array of possible interactions between genomic determinants and the components of the physical and social environment.

The relative contributions of genes and environment to disease risk can vary widely. In the case of highly penetrant heritable conditions such as cystic fibrosis or Huntington's disease, a single genetic mutation may be sufficient to cause disease, but even for these diseases the range and severity of symptoms may vary widely among individuals, partly as a result of environmental factors. In the case of common chronic diseases, where a variety of low-penetrance genomic variants are implicated, the effect of environment is generally much more evident.

It is a mistake, however, to think of any disease as resulting from a fixed and quantifiable combination of genomic and environmental determinants. For example, phenylketonuria, a condition inherited in a Mendelian fashion, is classified as a genetic (heritable) disease because the environmental determinant (dietary phenylalanine) is ubiquitous and the mutation rare and highly penetrant. However, the disease could also be thought of as 100% environmental, because it is only manifest when phenylalanine is present. We might also characterize the situation by stating that the genomic defect is a necessary factor in the pathogenesis of phenylketonuria but not in itself sufficient.

From the representation in Figure 21.1 it is clear that the health status of an individual may be influenced by altering either genomic or environmental factors, or both. In the context of disease prevention, Juengst (1995) distinguished these two types of intervention as "genotypic" and "phenotypic." Genotypic intervention may be appropriate in some circumstances, such as the use of genetic testing to avoid the birth of a child affected by a highly penetrant genetic disease. However, in the context of common chronic disease, which does not usually develop until some time during adult life, only phenotypic intervention will generally be either feasible or ethically acceptable.

Public health genomics aims to prepare the ground for a future era when we have a much more complete understanding of the full range of genomic and environmental determinants of health and – crucially – how they work together. The hope is that this fuller understanding of disease etiology and of the mechanisms of disease at the molecular and cellular levels will enable development of new diagnostic tools, new therapies, and perhaps new preventive options (Guttmacher and Collins, 2005), and that this new era of genomic medicine will offer opportunities to achieve benefits for population health.

Enthusiasm about the potential benefits from genomic research must be balanced by realism about the likely time scale of these developments (Davey Smith et al., 2005; Haga et al., 2003). Unraveling the genomic contribution to disease susceptibility is an immensely complex task, and many applications in mainstream health care are unlikely to materialize for some decades to come. In the meantime, it is a key function of public health genomics to protect patients and health services from the damaging effects of interventions that are premature, not supported by robust evidence or do not fulfill the criterion of public acceptability.

Avoiding Genetic Exceptionalism

It is important to avoid "hype" about advances in genomics and it is equally important to dispel the notion that genetic or genomic factors have a significance and predictive power beyond that of other determinants of health, a view that has been dubbed "genetic exceptionalism" (Murray, 1997). A related pitfall is genetic reductionism, the tendency to over-simplify the relationship between a genetic factor and a disease or other phenotypic trait (Sankar, 2003). Genetic exceptionalism and reductionism demand special protection for genomic information on the grounds of its predictive power and implications for other family members but this thinking derives from a misguided attempt to extrapolate from the highly penetrant mutations associated with Mendelian disease to the much more weakly predictive variants associated with common disease.

Public health genomics recognizes that genes must be included among the determinants of health but their significance

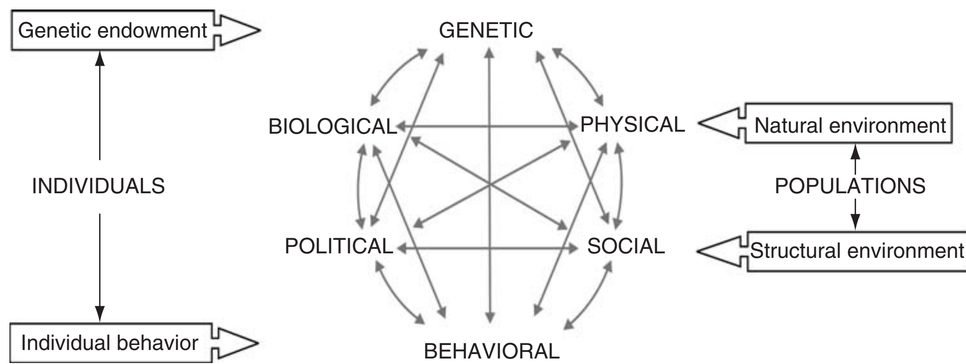


Figure 21.1 Determinants of health. The figure makes the point that external factors such as natural environmental and structural determinants are modifiable by interventions at the population level, whereas genetic factors and behavior are essentially individual (although they are, of course, influenced by external factors).

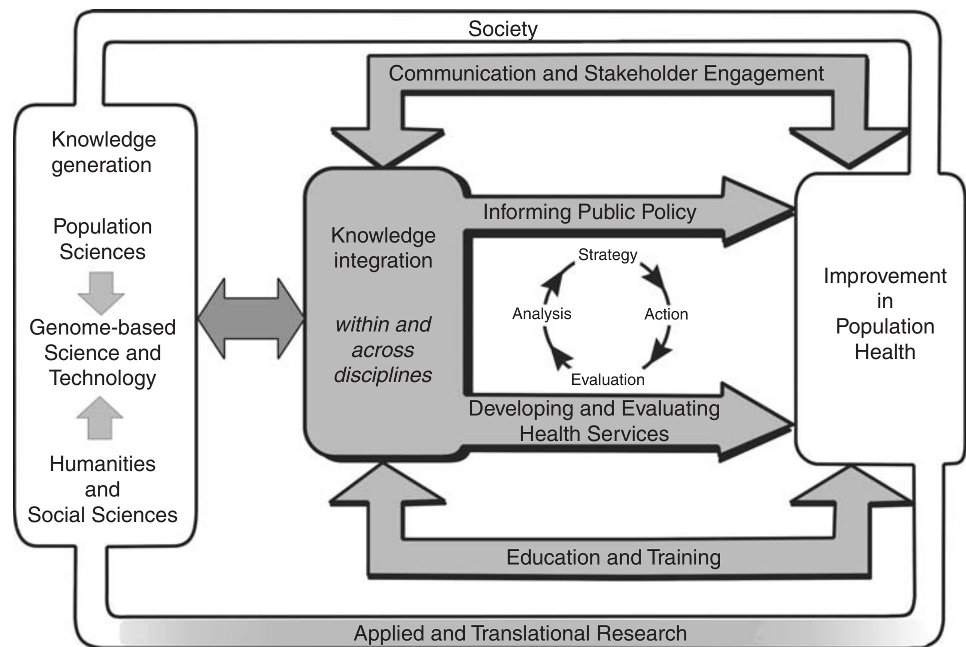


Figure 21.2 Strategy for the effective translation of genome-based knowledge and technologies for the benefit of population health. The shaded areas represent the "enterprise" of public health genomics.

should be kept in proportion: they should neither be overemphasized nor demonized. Genetic exceptionalism can lead to policy responses that are both disproportionate and irrational, in that they place restrictions on uses of DNA test information but leave other physiological or biomarker tests – which may be equally or even more highly predictive – unregulated.

THE "ENTERPRISE" OF PUBLIC HEALTH GENOMICS

The Bellagio workshop participants developed a visual representation of the "enterprise" of public health genomics, which is represented by the shaded regions of Figure 21.2; the aim of this representation is to convey a consensus understanding of what sorts of activities come under the banner of public health genomics, and how its work is conducted.

The starting point for the "enterprise" (but not part of the enterprise itself) is genomics research, shown at the left of Figure 21.2. Basic research is also needed in the population sciences (including, e.g., epidemiology and biostatistics) and in the humanities and social sciences (including law, philosophy, social anthropology, theology, and so on). Basic research can be thought of as the phase of knowledge generation.

Public health genomics comes into play with the phase of knowledge integration, both within and across disciplines. The

raw material for this process is the output from the research phase. Knowledge integration is defined as the process of selecting, storing, collating, analyzing, integrating, and disseminating information. This includes computer-based approaches to classifying, linking, and retrieving information (essentially, what an informatician would understand by the term "knowledge integration") but it is also a broader and more qualitative process encompassing selection, critical analysis, and synthesis of concepts and information from many different fields in the arts, humanities, biological, and social sciences. As the means by which information is transformed into knowledge, it is an essential step in making that information usable in the practical world of clinical medicine and public health practice.

The knowledge base generated by the knowledge integration function of public health genomics is used to underpin four core sets of activities: communication and stakeholder engagement; informing public policy; developing and evaluating health services; and education and training. Examples of some of these activities will be discussed later in this chapter.

Knowledge integration and the four core sets of activities set out *what* public health genomics does. *How* it carries out its work may be described by the cycle of public health action developed by the Institute of Medicine Committee for the Study of the Future of Public Health (1998) and others. This cycle consists of an initial phase of analysis followed by development of a strategy, implementation of that strategy, and evaluation of its outcomes. This is not a once-and-for-all process: successive rounds of the cycle lead to continued refinement of the strategy, both before and after implementation.

Research does have a place in the enterprise of public health genomics, as shown at the bottom of the diagram. Distinct from basic academic research, this is applied and translational research with direct applicability to health service implementation; such research can also identify gaps in the knowledge base that need to be filled by further basic research.

Finally, Figure 21.2 emphasizes that the whole enterprise of public health genomics – and the phase of basic research – is embedded within a societal context. The views and priorities of society play a part in determining what research is carried out and how it is used and, in turn, the perceived results of those activities shape societal attitudes.

CORE ACTIVITIES IN PUBLIC HEALTH GENOMICS

Knowledge Integration

Knowledge integration can be thought of as the driving force of public health genomics. A number of initiatives to support knowledge integration are underway at various centers. These activities aim to bring knowledge together from different sources across the sciences, arts, and humanities so that a multi-disciplinary resource is available for application to projects in public health genomics. For example, the HumGen database at the Centre de recherche en droit public (CRDP) at the University of Montreal, Canada, is a searchable international database of laws and policies relevant to the application of human genetics. The PolicyDB database curated by the Public Health Genetics Unit in Cambridge also deals with the policy field but differs from HumGen in that it concentrates mainly on the United Kingdom and the EU, and coverage includes policy for scientific research, public health, and health services as well as legal and bioethical documents.

A highly developed example of knowledge integration in public health genomics is the Human Genome Epidemiology Network (HuGE Net) initiative (Ioannidis, 2005). HuGE Net was set up by Muin Khoury in 1999 with the aim of assembling and critically analyzing population data on genotype prevalences and gene–disease associations emerging from research in genetic epidemiology. This task is an essential prerequisite to the development of any clinical or public health applications of this information. HuGE Net, discussed in more detail in another chapter of this book, has grown into an international multi-disciplinary initiative comprising around 800 collaborators in 43 countries. HuGE Net's core activities are information exchange through its website and newsletter, training and technical assistance in systematic review and meta-analysis, methodological development, knowledge base development, and information dissemination.

Another international initiative, spearheaded by the CRDP in Montreal, is the Public Population Project in Genomics (P3G) Consortium. The aim of P3G is "to provide the international population genomics community with the resources, tools, and know-how to facilitate data management for improved methods of knowledge transfer and sharing." P3G members include representatives from the major biobank projects underway across the world. A key feature is the P3G Observatory, a knowledge transfer platform that includes information on the design and status of the various projects and their governance arrangements, to enable comparisons and sharing of best practice.

Communication and Stakeholder Engagement

Productive dialogue with the public and other stakeholders, including patient groups, government, and industry, is important to ensure that scientific and clinical developments proceed in harmony with the attitudes and expectations of other sectors of society.

Among examples of work in this area, the Public Health Genetics Unit (PHGU) in Cambridge established a public panel to provide an independent perspective on its work and actively involved patients and carers in its service development projects. For example, a project on learning disability employed a mixture of online discussion, telephone interviews, and focus groups to develop a document, Parents as Partners, which provides detailed evidence of the value that parents place on a genetic diagnosis in learning disability, and sets out recommendations for the health service (Gogarty, 2006).

Service Development and Evaluation

This activity includes development of policies, programs, and services – both clinical and preventive – in the health sector;

strategic planning; service organization; manpower planning and capacity building; service review and evaluation; and guideline development.

The evaluation of population genetic screening programs, either proposed or current, is an example of work in this strand of public health genomics. In the United States, for example, public health leadership was instrumental in securing a rational approach to the proposed introduction of population screening for hereditary hemochromatosis. A multi-agency process produced a consensus statement that the current evidence base did not justify screening and made recommendations for further research (Burke et al., 1997). Further population-based research was funded to establish the penetrance of hemochromatosis-associated mutations and the prevalence of symptomatic hemochromatosis, and a web page is available on the website of the National Office of Public Health Genomics providing education about hereditary hemochromatosis for health care providers and the general public.

The public health genomics community has also taken the lead on the evaluation of DNA tests for clinical use, developing both the theoretical foundations for test evaluation and protocols for implementation in practice. For example, CDC's National Office of Public Health Genomics has convened a project known as EGAPP (Evaluation of Genomic Applications in Practice and Prevention) that aims "to develop a coordinated process for evaluating genetic tests and other genomic applications that are in transition from research to clinical practice." The EGAPP project draws on experience gained from application of an earlier evaluation protocol, the ACCE framework (standing for analytical validity, clinical validity, clinical utility, and ethical, legal, and social implications; Haddow and Palomaki, 2003). The UK's Public Health Genomics Foundation also has a strong involvement in genetic test evaluation, working with the UK Genetic Testing Network to develop and apply a protocol for the evaluation of tests funded by the National Health Service (Sanderson et al., 2005).

Informing Public Policy

The term "policy" refers to a broad range of public policies and programs that have a direct or indirect impact on the application of genomics in health care. Activities may encompass legal, philosophical, and social analysis at an applied level; development of regulatory frameworks; engagement in the policy making process; promoting relevant research; and seeking international comparisons.

Current policy issues include legal and regulatory frameworks for genetic testing; the funding of science and the prioritization of relevant research; consent, confidentiality, data protection, and the use of human tissue; attitudes to and relationships with the pharmaceutical and biotechnology industries; and the patenting of genes and genomic technologies. There is considerable academic research activity on these topics. For example, the US National Human Genome Research Institute "ELSI" program has an annual budget of around $18 million to support research on topics including privacy and fair use of genomic information, DNA banking, ethical conduct of genomic research, genetic discrimination, and the psychosocial impact of genomic testing and technologies. In the United Kingdom, the Economic and Social Research Council has also invested heavily in this field, supporting social science research in genomics through six Genomics Centers that together form the ESRC Genomics Network.

Despite the volume of theoretical work that has been done on these issues, however, there has been less attention to finding practical policy solutions; an aim of public health genomics is to fill this gap. The PHGU's policy team, for example, carried out detailed scrutiny of draft UK legislation governing the storage and use of human tissue samples for purposes including DNA analysis (Liddell and Hall, 2005). Flaws in the draft legislation were identified that could have impeded both research and clinical practice. Working with partner organizations in clinical genetics and biomedical research, PHGU was able to help influence the political process to achieve amendments that at least partly resolved these problems, and was involved in initiatives to produce practical guidance for health professionals needing to comply with the new legislation.

PHGU's mode of working on the Human Tissue Bill highlights another key aspect of public health genomics: the importance of forging working alliances with relevant stakeholder groups from research, clinical medicine, industry, patients, and the general public to achieve a consensus approach that is vital for securing "ownership" of a project and its conclusions.

Education and Training

As genomics and genomic technologies begin to have an impact on clinical practice, all health professionals will need to become "literate" in the general principles of genomics and its specific applications in their own specialist field. Public health professionals and health service managers will also need a working knowledge of genomics in order to make sound judgments about the planning and evaluation of health services (Burke, 2005).

Public health genomics contributes to these goals by promoting programs of genomic literacy for health professionals (as well as generally within society); specific training for public health genomics specialists; and development of educational materials, courses, workshops, and seminars. For example, a set of competencies in genomics for the public health workforce has been developed by the National Office of Public Health Genomics in Atlanta. Competencies are documented for the public health workforce as a whole and for specific groups including leaders/administrators, clinicians, epidemiologists, health educationalists, laboratory staff, and environmental health workers. For public health genomics specialists, the multi-disciplinary Masters and PhD programs developed at the Universities of Michigan and Washington offer a solid grounding in the field: the University of Washington Masters course, for example, includes core modules in genetic epidemiology; ethical and social issues; pharmacogenetics and toxicogenomics; genetics and the law; public perspectives on ethical issues; introduction to genetic services and bioinformatics; and economic and policy issues for genetic technologies and services.

In the United Kingdom, the PHGU led a project to develop a national strategy for genetics education for health professionals; the project's report (Burton, 2003) was one of the influences that led to the establishment of a government-funded NHS National Genetics Education and Development Centre in Birmingham. A multi-disciplinary post-graduate course in public health genomics is under development at Cranfield University, based on a textbook authored by members of the Public Health Genetics Unit (Stewart et al., 2007).

THE ROLE OF PUBLIC HEALTH IN THE TRANSLATION OF HUMAN GENOME DISCOVERIES INTO HEALTH APPLICATIONS

Public health has been traditionally identified with state, federal and local public health agencies; however, recent reports by the Institute of Medicine in the United States and others have adopted a more inclusive view, defining public health professionals as those working on improving health from a population perspective (IOM, 2003). "Public health professionals" are those employed not only in government but also in health care delivery, academia, community organizations, and the private sector; together, they are actors in the "public health system" (IOM, 2002), which is working to assure the conditions under which a population can be healthy. If we adopt the expanded view of public health as population health, it becomes much easier to consider genomics at the interface of the translational research agenda between medicine and public health.

Another reason to adopt an expanded view of population health is that future applications of genomics will occur primarily in the health care delivery system and not in the context of the national or state-mandated screening programs (traditional public health). Unfortunately, the sad history of eugenics will serve as constant reminder of horrifically failed applications for the perceived benefits of population health. Furthermore, the only current model of intersection of genetics and public health, namely newborn screening, should not necessarily serve as a model of future applications and joint partnership between medicine and public health. Undoubtedly, while newborn screening programs make important contributions to health, they continue to raise important ethical questions about informed consent, especially in the light of proposed expanded programs where health benefits are absent or less clear cut than medical emergencies provided by phenylketonuria (PKU) and congenital hypothyroidism. Thus, under the rubric of population health, collaboration between public health and medicine in genomics will emphasize the importance of an individual-oriented perspective but also the need for a population-based evaluation strategy and evidence-based health care with access of validated information to all segments of the population.

Advances in genomics have also been accompanied by an emerging emphasis on translational research, a movement fueled by the NIH road map initiative (Zerhouni, 2003, 2005) to accelerate the translation from basic science to clinical applications. The translation framework is based primarily on the discovery of new drugs and their accelerated use in human clinical trials, with little or no emphasis on prevention. However, the "bench to bedside" paradigm covers only part of the distance from discovering new knowledge to delivering health benefits at the population level (Green and Seifert, 2005; Horig and Pullman, 2004). In 2003, Claude Lenfant (2003), the retiring director of the National Heart, Lung and Blood Institute, presented the Shattuck lecture titled "Clinical research to clinical practice – lost in translation?" He described many discoveries of curative or preventive interventions that nevertheless do not reach the end of the translation highway and asked, "If we can't do it with aspirin, how will we do it with DNA?"

The "lost in translation" problem is complicated by the increasing costs of health care delivery and persistent inequities in access. Some have called the next stage in translation "translation type II" (Rohrbach et al., 2006), which requires more applied research on the best ways to deliver or disseminate interventions that work in real life settings (delivery research) and to evaluate health outcomes and population impact (outcomes research). In 2007, Westfall et al. (2007) proposed that the evaluation of evidence-based interventions in practice can be called "type 3 translation research." As shown in Table 21.1 we have also extended (Khoury et al., 2007b) the translation pathway to type 4 translation (T4) research, which seeks to evaluate the "real world" health outcomes of a genomic application in practice. We have reported that most human genomics research between 2001 and 2006, inclusive, was not translational and have estimated that no more than 3% of published research focuses on T2 and beyond. Indeed, evidence-based guidelines and T3 and T4 research currently are rare (Khoury et al., 2007b).

As discussed below, an enhanced focus on translation and translation research from T1 through T4 can foster a true partnership between medicine and public health in the genomics era in four major areas (see Figure 21.3).

THE FOCUS ON DISEASE PREVENTION AND HEALTH PROMOTION

Advances in genomics will provide new opportunities for disease prevention and health promotion – the main focus of public health, regardless of whether it is delivered at the individual level or through population-wide interventions. Understanding genetic effects and gene–environment interactions in disease processes could lead to the recommendation that certain subgroups avoid defined exposures or receive targeted interventions. Stratification by genotype or family history already provides a means for tailoring screening tests for early disease detection (e.g., colorectal cancer screening in genetically susceptible persons), a paradigm likely to be extended to early detection of other conditions.

Table 21.1 The continuum of translation research in human genomics: Types of research and *BRCA* example

Translation research phase	Notation	Types of research	Examples
TI	Discovery to candidate health application	Phases I and II clinical trials Observational studies	Is there an association between *BRCA* mutations and breast cancer?
T2	Health application to evidence-based practice guidelines	Phase III clinical trials Observational studies Evidence synthesis and guidelines development	What is the positive predictive value of *BRCA* mutations in at-risk women?
T3	Practice guidelines to health practice	Dissemination research Implementation research Diffusion research Phase IV clinical trials	What proportion of women who meet the family history criteria are tested for *BRCA* and what are the barriers to testing?
T4	Practice to population health impact	Outcomes research (includes many disciplines) Population monitoring of morbidity, mortality, benefits, and risks	Does *BRCA* testing in asymptomatic women reduce breast cancer incidence/improve outcomes?

Adapted from Khoury et al. (2007b). See also Chapter 14.

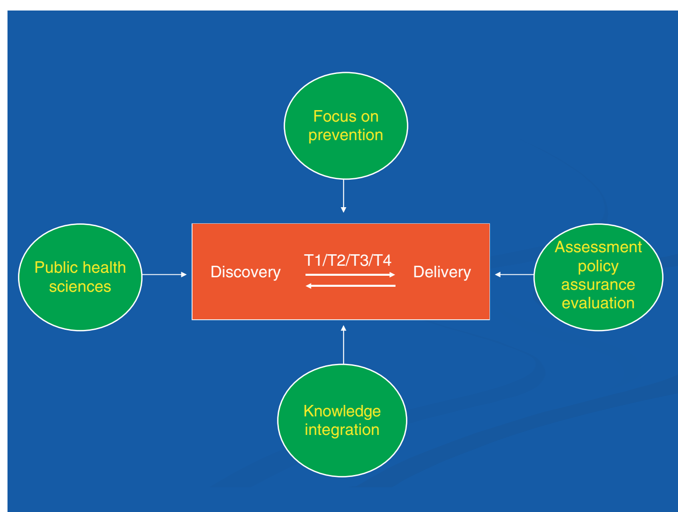


Figure 21.3 A framework for medicine-public health partnership in the genomics era (Adapted from Khoury et al., 2007a, b).

A review of the public health implications of genomic research related to asthma (UWCGPH, 2004) illustrates the potential opportunities and challenges for translating new knowledge into improved prevention and treatment of a common disease. Strong evidence supports a causal role for both genetic and environmental factors. Genomics research has identified numerous gene loci associated with asthma, and further studies of biological pathways associated with asthma, are likely to yield new approaches to prevention and therapy. The earliest clinical applications will be in pharmacogenomics, using genetic information to optimize therapy and prevent adverse events. The path for translating results of genomic research to population-level interventions will not always lead through genetic testing and knowledge of individual genotypes. For example, a study in Mexico of children with asthma found that supplementation with the antioxidant vitamins C and E improved lung function in children with a common polymorphism of *glutathione S-transferase M1* (*GSTM1*) who are exposed to ozone (Romieu et al., 2004). These findings might suggest a simple intervention – antioxidant vitamin supplementation – for children with asthma who are exposed to ozone. Without genotype-specific analysis, a potentially important population-level intervention could have been overlooked.

New gene discoveries are reported daily and have initiated dialog on the value of genetic information with respect to prevention. Consider, for example, the discovery in 2006 that a variant of the *TCFT7L2* gene is associated with increased risk of type 2 diabetes (Grant et al., 2006). This finding was noteworthy for several reasons. First, type 2 diabetes is a serious disease and a major public health problem. Family history is an important risk factor but until now, few genetic associations have been identified. The investigators replicated the association with *TCF7L2* in three independent populations and presented molecular evidence that the gene product is a high mobility group box-containing transcription factor related to blood glucose homeostasis. The gene product may act through regulation of proglucagon gene expression via the Wnt signaling pathway (Grant et al., 2006). The senior investigator commented to the *New York Times* that a practical consequence of the discovery could be a diagnostic test to identify people who are at increased risk of type 2 diabetes. He speculated that "these people, knowing their risk, would be motivated to exercise more and adopt a healthier diet" (Jannssens et al., 2006). However, a relatively simple analysis is sufficient to show that a test for *TCF7L2* variants by itself would have very poor predictive value, adding little to the risk

information already provided by family history, and would have limited utility for prevention (Jannssens et al., 2006). In particular, the hypothesis that genetic risk information of this kind will motivate behavioral change remains to be tested.

Once a genetic variant is discovered, producing a molecular test to detect it is theoretically straightforward. However, the clinical validity of such a test is highly dependent on population characteristics; these include not only prevalence of the genetic variant and the strength of its association with disease, but prevalence of the disease and interactions between this genetic variant and many other risk factors. Clinical utility is even more difficult to establish because it depends on the availability of a specific, effective intervention that adds value to existing practice. Although most gene discoveries for common diseases are not ready for prevention applications, they have, nevertheless, initiated a common interest in medical and public health to develop a common framework for how such discoveries can be evaluated for their potential applications for disease prevention.

2009 UPDATE

Two major developments set the stage for genome-wide association studies (GWAS): the International HapMap Project, which produced a database of genome-wide variation, and high-throughput genotyping technology, which has become increasingly efficient and accessible. Only a handful of GWAS appeared before 2007, when more than 100 such studies were published, followed by more than 120 in 2008 (Yu et al., 2009). Already, GWAS have identified more than 100 genetic associations with a wide range of common diseases and traits (Hindorff et al., 2009; Manolio et al., 2008). Very large study populations are required to detect the relatively small associations (i.e., per-allele odds ratios from 1.1 to 2.0) typically found in GWAS; thus, these studies have required unprecedented cooperation and collaboration among researchers in genetics, medicine, and epidemiology. Increasing interaction and interdependence among these disciplines is leading to a convergence in research approaches (McCarthy et al., 2008).

So far, most published human genomics research has focused on discovery: fewer than 3% of publications address the potential clinical validity or utility of genomic information for improving health or preventing disease (Khoury et al., 2007). Efforts to bridge the "translation gap" begin with knowledge synthesis and evaluation. For example, a consensus group has proposed grading the cumulative epidemiologic evidence for genetic associations as weak, moderate, or strong according to three criteria – amount of evidence, extent of replication, and protection from bias (Ioannidis et al., 2008). Applying this approach to evaluate all genetic associations with a specific disorder, such as schizophrenia, provides a "field synopsis" (Allen et al., 2008). Systematic review and synthesis (e.g., by meta-analysis) is indispensable for interpreting the burgeoning literature on genetic associations, which contains many false positive results, even in the GWAS era (Attia et al., 2009). For example, a field synopsis of DNA repair genes and susceptibility to cancer found that of 1123 studied associations, no more than three were supported by strong evidence (Vineis et al., 2009).

Increasing numbers of genetic tests are finding their way into clinical use, often before their validity and utility have been assessed. An initiative from the US Centers for Disease Control, the Evaluation of Genomic Applications for Practice and Prevention (EGAPP), has developed methods for evaluating the evidence for genetic tests proposed for use in a particular clinical setting (Teutsch et al., 2008). The first such evaluation addressed testing for cytochrome P450 polymorphisms to guide treatment with selective serotonin reuptake inhibitors in depressed adults (AHRQ, 2007); several additional recommendations have been published recently (EGAPP Working Group, 2008a, b, c). Information about pending EGAPP reviews and recommendations is available from the EGAPP website (http://www.egappreviews.org).

Recently, several companies have begun to offer personal genome scans directly to individuals via the Internet. These tests are based on the same arrays used for GWAS, and consumers are provided with summaries of research findings; individual results are most often presented as relative risks or odds ratios (usually less than 2.0). Although personal genome scans have received much popular attention, they so far provide little basis for medical decision-making or lifestyle advice (Hunter et al., 2008). It remains to be seen how consumers and/or healthcare providers will use such information.

Although it may seem paradoxical, the path to predictive personalized medicine passes through population-based research, because probabilistic risk estimates for complex disorders require inference from larger population samples. As observed before by critics of individualized preventive medicine, few risk factors for common chronic diseases have sufficient predictive ability to serve as screening tools; in this respect, common genetic variants associated with disease susceptibility are unlikely to be different.

Information about genetic variants, when integrated with other clinical and epidemiologic data, may be useful for defining groups of persons most likely to benefit from particular interventions. Tests for multiple genetic variants could also be expected to have greater predictive value than a test for a single variant. For example, a recent modeling study found that a test for seven common breast cancer susceptibility alleles would predict risks of incident disease that varied sixfold between women with the most and the fewest alleles; although the test provided insufficient discrimination to be useful at the individual level, the authors of this study suggested that it could be used to stratify

the population for risk-based targeting of mammographic screening (Pharoah et al., 2008).

Like other tests, genetic tests must be evaluated in the context of their application for a specific purpose. In general, clinical decisions hinge on the absolute risk of a particular outcome during a specified time interval and on the relative benefits and harms of different interventions at each level of risk – the measures that together define clinical utility (Zimmern, 2009). Even highly predictive tests may not be useful if their results have no bearing on intervention or treatment; even clinically useful tests, such as genotype-guided warfarin dosing, may not be cost-effective (Eckman et al., 2009). Some basic safeguards on translation of human genomic research include transparent reporting of study methods and results, systematic evaluation of cumulative evidence for association, careful assessment of clinical validity and utility, and clear communication to the public about what a genetic test result can or cannot tell us (Ioannidis et al., 2009).

2009 UPDATE REFERENCES

Agency for Healthcare Research and Quality (AHRQ) (2007). Testing for cytochrome P450 polymorphisms in adults with non-psychotic depression treated with selective serotonin reuptake inhibitors (SSRIs), Structured Abstract. Rockville, MD. http://www.ahrq.gov/clinic/tp/cyp450tp.htm, accessed January 18, 2009.

Allen, N.C., Bagade, S., McQueen, M.B., Ioannidis, J.P., Kavvoura, F.K., Khoury, M.J., Tanzi, R.E. and Bertram, L. (2008). Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: The SzGene database. *Nat Genet* 40, 827–834.

Attia, J., Ioannidis, J.P., Thakkinstian, A., McEvoy, M., Scott, R.J., Minelli, C., Thompson, J., Infante-Rivard, C. Guyatt, G. (2009). How to use an article about genetic association: B: Are the results of the study valid? *JAMA* 301, 191–197.

Eckman, M.H., Rosand, J., Greenberg, S.M. and Gage, B.F. (2009). Cost-effectiveness of using pharmacogenetic information in warfarin dosing for patients with nonvalvular atrial fibrillation. *Ann Intern Med* 150, 73–83.

Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group. Recommendations from the EGAPP Working Group: Can tumor gene expression profiling improve outcomes in patients with breast cancer? *Genet Med* 2008a December 31; [Epub ahead of print].

Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group. Recommendations from the EGAPP Working Group: Genetic testing strategies in newly diagnosed individuals with colorectal cancer aimed at reducing morbidity and mortality from Lynch syndrome in relatives. *Genet Med* 2008b December 31; [Epub ahead of print].

Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group. Can UGT1A1 genotyping reduce morbidity and mortality in patients with metastatic colorectal cancer treated with irinotecan? *Genet Med* 2008c December 31; [Epub ahead of print].

Hindorff, L.A., Junkins, H.A., Mehta, J.P. and Manolio, T.A. A catalog of published genome-wide association studies. www.genome.gov/26525384, accessed January 18, 2009.

Hunter, D.J., Khoury, M.J. and Drazen, J.M. (2008). Letting the genome out of the bottle–will we get our wish? *N Engl J Med* 358, 105–107.

Ioannidis, J.P., Boffetta, P., Little, J., O'Brien, T.R., Uitterlinden, A.G., Vineis, P., Balding, D.J., Chokkalingam, A., Dolan, S.M., Flanders, W.D. et al. (2008). Assessment of cumulative evidence on genetic associations: Interim guidelines. *Int J Epidemiol* 37, 120–132.

Ioannidis, J.P. (2009). Personalized genetic prediction: Too limited, too expensive, or too soon? *Ann Intern Med* 150, 139–141.

Khoury, M.J., Gwinn, M., Yoon, P.W., Dowling, N., Moore, C.A. and Bradley, L. (2007). The continuum of translation research in genomic medicine: how can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention? *Genet Med* 9, 665–674.

Manolio, T.A., Brooks, L.D. and Collins, F.S. (2008). A HapMap harvest of insights into the genetics of common disease. *J Clin Invest* 118, 1590–1605.

McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P. and Hirschhorn, J.N. (2008). Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nat Rev Genet* 9, 356–369.

Pharoah, P.D.P., Antoniou, A.C., Easton, D.F. and Ponder, B.A.J. (2008). Polygenes, risk prediction, and targeted prevention of breast cancer. *N Eng J Med* 358, 45–52.

Teutsch, S.M., Bradley, L.A., Palomaki, G.E., Haddow, J.E., Piper, M., Calonge, N., Dotson, W.D., Douglas, M.P. and Berg, A.O. Chair, on behalf of the EGAPP Working Group. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) initiative: Methods of the EGAPP Working Group. *Genet Med* 2008 December 31; [Epub ahead of print].

Vineis, P., Manuguerra, M., Kavvoura, F.K., Guarrera, S., Allione, A., Rosa, F., Di Gregorio, A., Polidoro, S., Saletta, F., Ioannidis, J.P. et al. (2009). A field synopsis on low-penetrance variants in DNA repair genes and cancer susceptibility. *J Natl Cancer Inst* 101, 24–36.

Yu, W., Gwinn, M., Clyne, M., Yesupriya, A. and Khoury, M.J. (2008). A navigator for human genome epidemiology. *Nat Genet* 40, 124–125. www.hugenavigator.net, accessed January 18, 2009.

Zimmern, R.L. (2009). Testing challenges: Evaluation of novel diagnostics and molecular biomarkers. *Clin Med* 9, 68–73.

REFERENCES

Burke, W. et al. (1997). Hereditary haemochromatosis: Gene discovery and its implications for population-based screening. *JAMA* 280, 172–178.

Burke, W. (2005). Contributions of public health to genetics education for health professionals. *Health Educ Behav* 32, 668–675.

Burke, W., Khoury, M.J., Stewart, A. and Zimmern, R.L. (2006). The path from genome-based research to population health: Development of an international collaborative public health genomics initiative. *Genet Med* 8, 451–458.

Burton, H. (2003). Addressing Genetics, Delivering Health. Public Health Genetics Unit, Cambridge. Available at http://www.phgfoundation.org.

Clayton, E.W. (2003). Genomic medicine: Ethical, legal, and social implications of genomic medicine. *N Engl J Med* 349, 562–569.

Davey Smith, G.D., Ebrahim, S., Lewis, S., Hansell, A.L., Palmer, L.J. and Burton, P.R. (2005). Genetic epidemiology and public health: Hope, hype, and future prospects. *Lancet* 366, 1484–1498.

EGAPP project. National Office of Public Health Genomics, Centers for Disease Control and Prevention. http://www.cdc.gov/genomics/gtesting/egapp.htm.

Gogarty, B. (2006). Parents as Partners. Public Health Genetics Unit, Cambridge. Available at http://www.phgfoundation.org.

Grant, S.F., Thorleifsson, G., Reynisdottir, I. et al. (2006). Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet* 38, 320–323.

GRaPH *Int*. http://www.graphint.org.

Green, L.A. and Seifert, C.M. (2005). Translation of research into practice: Why we can't just do it. *J Am Board Fam Pract* 18, 541–545.

Guttmacher, A.E. and Collins, F.S. (2005). Realizing the promise of genomics in biomedical research. *N Engl J Med* 347, 1512–1520.

Haddow, J.E. and Palomaki, G.E. (2003). ACCE: A model process for evaluating data on emerging genetic tests. In *Human Genome Epidemiology: A Scientific Foundation for Using Genetic Information to Improve Health and Prevent Disease* (M.J. Khoury, J. Little and W. Burke, eds), Oxford University Press, New York, pp. 217–233.

Haga, S.B., Khoury, M.J. and Burke, W. (2003). Genomic profiling to promote a healthy lifestyle: Not ready for prime time? *Nat Genet* 34, 347–350.

Horig, H. and Pullman, W. (2004). From bench to clinic and back: Perspectives on the 1st IQPC translational research conference. *J Trans Med* 2, 44.

Human Genome Epidemiology Network (HuGE Net). http://www.cdc.gov/genomics/hugenet/default.htm.

HumGen database. http://www.humgen.umontreal.ca/int/.

Ioannidis, J.P. (2005). A road map for efficient and reliable human genome epidemiology. *Nat Genet* 38, 3–5.

Institute of Medicine (2003). *Who Will Keep the Public Healthy? Educating Public Health Professionals for the 21st Century*. National Academies Press, Washington, DC.

Institute of Medicine (2002). *The Future of the Public's Health in the 21st Century*. National Academies Press, Washington, DC.

Institute of Medicine Committee for the Study of the Future of Public Health (1998). *The Future of Public Health*. National Academies Press, Washington DC.

Jannssens, A.C.W., Gwinn, M.L., Valdez, R. et al. (2006). Predictive genetic testing for type 2 diabetes may raise unrealistic expectations. *BMJ* 333, 509–510.

Juengst, E.T. (1995). 'Prevention' and the goals of genetic medicine. *Hum Gene Ther* 6, 1595–1605.

Khoury, M.J., Gwinn, M., Burke, W., Bowen, M.S. and Zimmern, R. (2007a). Will genomics widen or heal the schism between medicine and public health. *Am J Prev Med* 33, 310–317.

Khoury, M.J., Gwinn, M., Yoon, P.W., Dowling, N., Moore, C.A. and Bradley, C.A. (2007b). The continuum of translation research in genomic medicine how can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention? *Genet Med* 9, 665–674.

Lenfant, C. (2003). Shattuck lecture: Clinical research to clinical practice – lost in translation?. *N Engl J Med* 349, 868–874.

Liddell, K. and Hall, A. (2005). Beyond Bristol and Alder Hey: The future regulation of human tissue. *Med Law Rev* 13, 170–223.

Murray, T. (1997). Genetic exceptionalism and 'future diaries': Is genetic information different from other medical information? In *Genetic Secrets: Protecting Privacy and Confidentiality in the Genetic Era* (M.A. Rothstein, eds.), Yale University Press, New Haven, pp. 60–73.

National Office of Public Health Genomics, Centers for Disease Control and Prevention. Genomic Competencies for the Public Health Workforce. http://www.cdc.gov/genomics/training/competencies/default.htm.

National Office of Public Health Genomics, Centers for Disease Control and Prevention. Hereditary Haemochromatosis: A Public Health Perspective. http://www.cdc.gov/genomics/training/perspectives/hemo.htm.

NHS National Genetics Education and Development Centre. http://www.geneticseducation.nhs.uk/.

Ojha, R.P. and Thertullien, R. (2005). Health care policy issues as a result of the genomic revolution: Implications for public health. *Am J Public Health* 95, 385–388.

PHG Foundation Policy DB Database. http://www.phgfoundation.org/policydb.

Public Health Genomics European Network. http://www.phgen.nrw.de/.

Public Population Project in Genomics (P3G). http://www.p3gconsortium.org/.

Rohrbach, L.A., Grana, R., Sussman, S. and Valente, T.W. (2006). Type II translation: Transporting prevention interventions from research to real world settings. *Eval Health Prof* 29, 302–333.

Romieu, I., Sienra-Monge, J.J., Ramirez-Aguilar, M. et al. (2004). Genetic polymorphism of *GSTM1* and antioxidant supplementation influence lung function in relation to ozone exposure in asthmatic children in Mexico City. *Thorax* 59, 8–10.

Sanderson, S., Zimmern, R., Kroese, M., Higgins, J., Patch, C. and Emery, J. (2005). How can the evaluation of genetic tests be enhanced? Lessons learned from the ACCE framework and evaluating genetic tests in the United Kingdom. *Genet Med* 7, 495–500.

Sankar, P. (2003). Genetic privacy. *Annu Rev Med* 54, 393–407.

Stewart, A. (2006). Genome based Research and Population Health Report of an expert workshop held at the Rockefeller Foundation study and Conference Centre, Bellagio, Italy, 14–20 April, 2005. Available at http://www.phgfoundation.org.

Stewart, A., Brice, P., Burton, H., Pharoah, P., Sanderson, S. and Zimmern, R. (2007). *Genetics, Health Care and Public Policy*. Cambridge University Press, Cambridge UK. (This textbook sets out the knowledge base for public health genetics and the policy and service issues raised by genomic advances, from a UK perspective.)

University of Michigan School of Public Health Interdepartmental Concentration in Public Health Genetics. http://www.sph.umich.edu/genetics/.

University of Washington Center for Genomics and Public Health (UWCPGH) (2004). Asthma genomics: Implications for public health. 2004. Accessed online on March 15, 2007 at: http://depts.washington.edu/cgph/workinggroups/Asthma%20Genomics_Implications%20for%20Public%20Health.pdf

University of Washington Institute of Public Health Genetics, Degree Programs in Public Health Genetics. http://depts.washington.edu/phgen/degreeprograms/DegreeProgs_OV.shtml.

Westfall, J.M., Mold, J. and Faqnan, L. (2007). Practice-based research – blue highways on the NIH road map. *JAMA* 297, 403–406.

Zerhouni, E.A. (2003). The NIH Roadmap. *Science* 302, 63–72.

Zerhouni, E.A. (2005). Translational and clinical science: Time for a new vision. *New Engl J Med* 353, 1621–1623.

RECOMMENDED RESOURCES

Gwinn, M. and Khoury, M.J. (2006). Genomics and public health in the United States: Signposts on the translation highway. *Community Genet* 9, 21–26. (A recent review of progress in public health genomics in the United States.)

Halliday, J.L., Collins, V.R., Aitken, M.A., Richards, M.P. and Olsson, C.A. (2004). Genetics and public health – evolution, or revolution? *J Epidemiol Community Health* 58, 894–899. (A discussion of the prospects and challenges raised by a convergence between public health and genetics.)

Institute of Medicine Committee on Genomics and the Public's Health in the 21st Century (2005). In *Implications of Genomics for Public Health* (Hernandez, L., ed.), National Academies Press, Washington DC. (This report summarizes the "state of the art" for public health genomics in the United States, and prioritizes the issues that need to be addressed.)

Khoury, M.J., Burke, W. and Thomson, E., eds. (2000). *Genetics and Public Health in the 21st Century*. Oxford University Press, New York. (An excellent overview of the conceptual framework of public health genetics and its early development during the 1990s.)

National Office of Public Health Genomics website. http://www.cdc.gov/genomics. Provides information on NOPHG projects, training resources, and the GDPInfo database of published literature relating genes, diseases, and environmental exposures. Hosts the HuGE Net website. Users can sign up for a weekly e-mail update of news, event, and publications in public health genomics.

PHG Foundation website. http://www.phgfoundation.org. Features information on the PHGU work program, and an online newsletter of genomics policy news, linked to related websites and records in the PolicyDB database. Users can request a monthly e-mail compilation and commentary of genomics policy news.

Zimmern, R. and Cook, C. (2000). Genetics and health. Policy issues for genetic science and their implications for health services. The Nuffield Trust, London. (A useful and particularly broad-ranging discussion of service and policy issues raised by genomics.)

This page intentionally left blank

Clinical (Cardiology)

Section 4

22. The Genomics of Hypertension

23. Lipoprotein Disorders

24. Genomics of Myocardial Infarction

25. Acute Coronary Syndromes

26. Heart Failure in the Era of Genomic Medicine

27. Genomic Assessment of Cardiac Transplant Rejection

28. Genetics and Genomics of Hypertrophic Cardiomyopathy

29. Genetics and Genomics of Arrhythmias

30. Genetics and Genomics in the Management of Hemostasis and Thrombosis

31. Genomics of Congenital Heart Disease

This page intentionally left blank

CHAPTER

The Genomics of Hypertension

Chana Yagil and Yoram Yagil

INTRODUCTION

One of the promises of genomics has been that it would provide a better understanding and improve our ability to treat common diseases that afflict humanity. Hypertension is one of the most common and perhaps more important diseases that carry a high rate of morbidity and mortality. Hypertension, which is defined as systolic blood pressure above 140 mmHg and/or diastolic blood pressure above 90 mmHg, has a worldwide prevalence estimated at over 600 million people and in the United States alone at over 65 million. A wide variety of populations are affected worldwide, ranging over both sexes and all ethnic groups (Kearney et al., 2004).

Despite decades of intensive research in human populations and in experimental animal models, the pathophysiology underlying hypertension remains unresolved. In the majority of cases, hypertension is thought to result from an interaction of genes with environmental factors. In only a small minority of cases, genetic dissection of hypertension has uncovered single gene mutations that lead to the development of the so-called "monogenic" forms of the disease (Luft, 2001). The pathophysiology of these rare monogenic forms of hypertension has been elucidated during the past decade, but the pathophysiology underlying the common form of hypertension remains incomplete and elusive, hence the term "essential hypertension." Our current understanding of essential hypertension is limited to the knowledge that a yet undetermined number of "causative" genes encode for a probably larger number of proteins that are involved in a wide variety of pathophysiological pathways that somehow lead together to a rise in blood pressure and hypertension. These causative genes interact with a multitude of environmental factors, not all of which have yet been identified, and an unknown number of susceptibility genes. The susceptibility genes in turn modulate the response of the individual to the environment. Causative genes, environmental factors and susceptibility genes interact in a highly complex network (Figure 22.1), the outcome of which is a rise in blood pressure to abnormal levels that are clinically defined as hypertension. Hypertension thus belongs to the group of common multifactorial complex diseases.

The complexity of hypertension is perhaps best illustrated by Guyton's traditional model of cardiovascular dynamics (Figure 22.2a) in which multiple circuits of physiological mechanisms act in conjunction with one another to cause hypertension (Guyton and Coleman, 1969). The advent of genomics has allowed the translation of each component of Guyton's model into its underlying genetic make-up. Instead of simplifying the model, however, this allegedly "forward" step only renders the highly complex physiological scheme into an even more complicated micro-circuitry of genes and proteins that interact with

Copyright © 2010, Elsevier Inc.
All rights reserved.

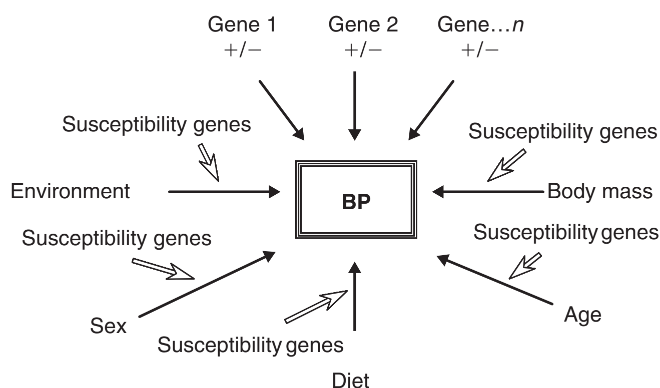


Figure 22.1 The complexity of hypertension. BP, blood pressure.

one another in multiple ways that eventually cause hypertension (Figure 22.2b).

Unraveling the huge complexity of the pathophysiology underlying the most common form of hypertension, the so called "essential hypertension," has been a daunting task which still remains far from being completed. The hope is that where traditional physiology has failed at large, genomics along with physiological genomics will provide more detailed insight into the mechanisms underlying essential hypertension, which will in turn facilitate a more rational therapeutic approach to the treatment of the disease, as well as development of new drugs for hypertension.

In this chapter, we will focus on how genomics has affected our understanding of the predisposition to hypertension, on our ability to screen, diagnose and determine the prognosis of patients with hypertension, on the applicability of genomics to the monitoring of hypertension and on the present and anticipated future impact of genomics on the treatment of hypertension.

PREDISPOSITION

In hypertension, two questions arise with regards to "predisposition." The first is what factors predispose the individual to develop hypertension. This issue will be discussed henceforth in depth. The second issue, which is of no lesser importance, is what predisposes a hypertensive individual to develop end-organ damage, the clinically significant and often devastating direct result of high blood pressure. End-organ damage, which determines the "clinical outcome" of hypertension, will be discussed in the section on "prognosis."

There is a major difficulty in dealing with the question what predisposes the individual to develop hypertension if one considers that if a subject lives to the age of 90, his/her likelihood to develop hypertension approaches 80–90% (Table 22.1). Most human beings are therefore "naturally" inclined and genetically predisposed to develop hypertension sometime in their lifetime, assuming that they live long enough. The incidence and prevalence of hypertension increase with age, and the immediate sequelae of hypertension, cardiovascular, cerebrovascular and renal disease indeed constitute important causes of morbidity and mortality in the elderly. The development of hypertension at an earlier age, however, perhaps "pre-maturely," is less frequent but more intriguing. The real issue of interest is therefore what confers the susceptibility of the individual to develop hypertension at an early age.

The issue of predisposition to develop hypertension, as with other multifactorial diseases, can also be viewed in a different way. One can ask the question what confers resistance, as opposed to susceptibility, of the individual to develop hypertension, resistance which prevents the development of hypertension at an early age as well as at a later age. It is possible, after all, that individuals are physiologically predisposed to hypertension and that blood pressure is maintained at "normal" levels through protective mechanisms, unless these fail.

Irrespective of whether sensitivity or resistance prevail, there is generalized agreement that the likelihood that an individual will develop hypertension sometimes along his/her lifetime is dependent on a multitude of factors, most likely a combination of genetic and environmental factors, notably "bad" genes in a "bad" environment (Geller, 2004).

Researchers have used genomics in an attempt to dissect the predisposition of the individual to environmental factors that leads to the development of hypertension and identify the culprit genes. A classical example has been the study of the susceptibility to dietary salt that causes hypertension. It has long been recognized that individuals are either salt-sensitive or salt-resistant. Those who are salt-sensitive develop hypertension when salt-intake is increased, and those who are salt-resistant remain normotensive, irrespective of dietary salt-intake. It has also been recognized that what renders individuals salt-sensitive and others salt-resistant is their genetic make-up. Studies in animal models such as the Dahl (Rapp, 1982) and Sabra (Yagil et al., 1996) rat models of salt-sensitivity provide unequivocal experimental proof that the trait of salt-susceptibility that leads to the development of hypertension is genetically inherited.

Have decades of research into the genetic basis of salt-susceptibility resulted in the detection of the genes or proteins that account for salt-susceptibility? Have genomics provided the solution to the question which mechanisms account for salt-sensitivity? There have been major problems in dissecting the genetic basis of salt-susceptibility directly in human populations where it is next to impossible to separate and isolate the various modulators of hypertension, including environmental factors, and focus on one at a time. It is certainly even more complicated in humans to combine interventional studies with genomic investigation of the underlying genes. Thus, there have been only few genomic investigations in humans that have aimed specifically at detecting susceptibility genes, and much of the investigation

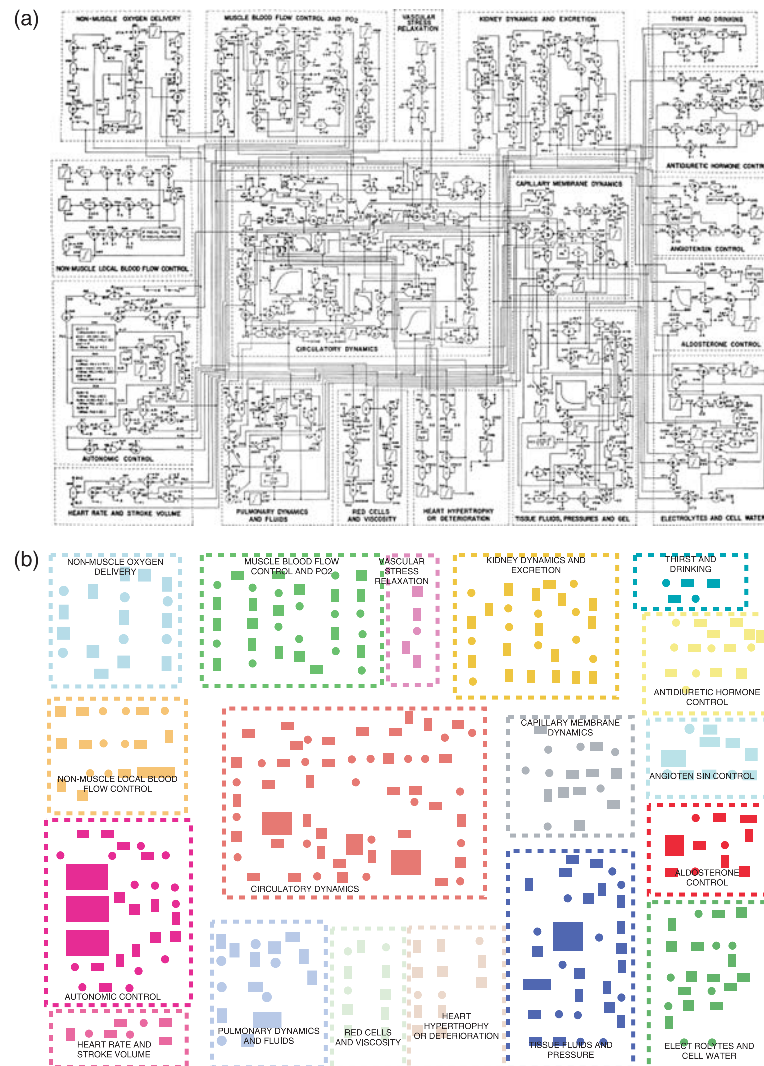


Figure 22.2 (a) Guyton's model of cardiovascular dynamics (adapted from Guyton, A.C., and Coleman, T.G. (1967). Long-term regulation of the circulation: interrelationships with body fluid volumes. In: *Physical Bases of Circulatory Transport Regulation and Exchange*. Saunders, Philadelphia, PA) and (b) after genomic annotation and translation of the model components into individual genes (colored boxes) .

had to be carried out in experimental models of the disease that allow not only isolation but also perturbations and alterations of single environmental factors, one at a time. Indeed, investigations of the genetic basis of salt-sensitivity using genomic approaches such as positional cloning in the experimental animal models of salt-sensitivity have yielded a large number of quantitative trait loci (QTLs), at least one on each autosome and on chromosome X (Garrett et al., 2002). A number of these QTLs, which allegedly incorporate within them the salt-sensitivity-related genes, have been confirmed through the use of genetically designed consomic and congenic strains (Cowley, Jr. et al., 2004a, b; Yagil and Yagil, 2003). A considerable number of genes have been identified within these QTLs as possible candidate genes for salt-susceptibility. Nonetheless, the investigation aiming to elucidate the genetic basis of salt-sensitivity or resistance is nowhere resolved and is actively ongoing. The search for the genetic basis of salt-susceptibility or the genetic basis for the predisposition to salt-sensitivity, however, is only one of the many directions that need to be investigated, as environmental factors other than dietary salt also predispose individuals to hypertension. What mechanisms predispose individuals to develop hypertension in face of stress? Climate? Obesity? Very little is known about these

TABLE 22.1 Prevalence of hypertension in the United States population during 1999–2000

Age (years)	Males (%)	Females (%)
20–34	11.8	3.1
35–44	19.2	18.6
45–54	36.9	33.4
55–64	50.7	57.9
65–74	68.3	73.4
>75	70.7	84.9

Health United States, 2003, NCHS.

other factors, and once again most of the data on susceptibility genes originate from studies in experimental models.

There have been attempts, nonetheless, to study genetic predisposition to hypertension directly in humans using a seemingly less sophisticated alternative approach aiming to link the hypertensive phenotype to variations in the composition of the genome by focusing directly on candidate genes and polymorphisms within those genes. It was thought that such association would allow identification of a predisposition of the individual who is not yet hypertensive but who carries the specific genetic variant to develop hypertension at a later age, the predisposition being expressed as "odds ratio." The list of candidate genes is long and has even merited a website listing 150 such candidate genes for hypertension (http://cmbi.bjmu.edu.cn/genome/candidates/candidates.html). Genetic polymorphisms within these candidate genes have been identified. Little insight has been gained, however, from attempts to associate these polymorphisms with hypertension, and no clinically useful information has been derived as to the ability to identify those individuals who are more prone to develop hypertension "prematurely." Possible explanations for this failure to identify clinically useful genetic polymorphisms may be related to the small population size that have been used in many of the studies, the possibility that the "wrong" candidate genes were chosen, or the inability to identify the polymorphisms that actually account for the development of hypertension. The link between genetic polymorphisms of candidate genes and hypertension remains thus elusive at this stage. Nevertheless, it is quite likely that the correct and relevant candidate gene variants will eventually be identified and these will help determine who within a population is predisposed to hypertension.

When dealing with predisposition to hypertension, one cannot overlook in this era of globalization the interesting, although emotionally laden, issue of ethnic/racial predisposition to hypertension. There is a disparity in the prevalence of hypertension amongst different racial/ethnic populations worldwide. Such variations in the prevalence of hypertension amongst different populations raises the question of whether there is ethnic/racial predisposition to hypertension and what has been the contribution of genomics to resolve the issue. The racial issue in relation to predisposition to disease in general has been difficult to deal with, as it may imply the existence of an inborn genetic flaw (Cooper and Kaufman, 1998). However, arguments regarding the existence or absence of racial or ethnic differences must also consider the large body of human genetic research in rarer Mendelian forms of diseases, some of which have been described as being largely restricted to single racial or ethnic populations. Racial differences may be within genetic subtypes that are particularly sensitive to environmental factors that can raise blood pressure. Race, however, cannot be regarded as an etiological factor *per se*, but rather as a "risk factor" for hypertension, without inferring causality. In this sense, race, as is age, should be considered a risk factor (O'Donnell and Kannel, 1998).

DIAGNOSIS

The diagnosis of hypertension at present is solely based on the measurement of blood pressure, using a blood pressure cuff, and on threshold levels above which subjects are defined as "hypertensive." Whether hypertension is best defined based on sporadic office measurements using standard blood pressure measuring devices or on 24-h intermittent measurements of blood pressure using an ambulatory blood pressure device (Pickering et al., 2006) is a matter that is still hotly debated and thus yet unresolved. The threshold values for hypertension, whether based on standard or ambulatory blood pressure measurements, are determined in many countries by national guidelines that base their decisions on observations that above a certain level, target organ damage occurs.

What is the place of genomics in the process of diagnosing hypertension? Because of the simplicity in the diagnosis of hypertension, genomics have no place in the process of determining whether a subject is hypertensive or not.

Genomics may have a place though, at least theoretically and conceptually, in diagnosing what type of hypertension the subject is suffering from. If one can grossly divide hypertension into "primary" or "essential" hypertension, two terms that are used interchangeably, and "secondary" hypertension, then it is theoretically possible that genomic tools might help eventually determine what kind of hypertension the subject has. Unfortunately, no genomic tools are yet available that differentiate between primary and secondary hypertension. This matter is further complicated by the fact that the term "essential hypertension" is used to describe hypertension of unknown or "idiopathic" origin. In fact, essential hypertension is likely to represent several separate clinical entities, the number and nature of which are presently unknown. Once the pathophysiological mechanisms of "essential hypertension" are elucidated, however, it is likely that virtually all cases of hypertension will be classified as "secondary," and the term "essential" for hypertension

will have become obsolete. But until that is achieved, all subjects with hypertension in whom a primary cause of hypertension has not been identified continue to be labeled "essential hypertension" by default, and the need to differentiate between primary and secondary hypertension prevails.

Genomics can be of help in identifying within the "secondary" hypertension group specific disease entities that are due to genetic alterations. This screening can be based on the search and identification of known mutations that are known to be associated with disease, or of patterns of gene expression that are typical for the disease. By identifying known specific genetic mutations, genomics can already be useful in the diagnosis of specific genetic variants that cause hypertension, the rarer forms of monogenic hypertension (Lifton et al., 2001). For example, Liddle's syndrome, a disorder that is associated with hypertension, low plasma renin and aldosterone levels, and hypokalemia, is due to a mutation in the sodium channel gene, specifically to a deletion or missense mutations of a PPPxY motif in the cytoplasmic COOH terminus of either the beta or gamma subunit of the epithelial Na channel (ENaC). Type 2 pseudohypoaldosteronism or Gordon's syndrome, is another hypertensive syndrome characterized by hypertension, hyperkalemia, normal renal function, and low or low-normal plasma renin activity and aldosterone concentrations; this clinical entity is due to mutations in WNK kinases 1 and 4 (Wilson et al., 2001). Glucocorticoid-remediable aldosteronism is a disorder in which a chimeric gene formed from the 11β-hydroxylase and aldosterone synthase genes (Lifton et al., 1992) results in ACTH stimulating aldosterone synthase, leading to persistent hyperaldosteronism. In congenital adrenal hyperplasia with 11β-hydroxylase deficiency, 10 different mutations of the *CYP11B1* gene have been identified (White et al., 1994), causing steroid 11β-hydroxylase deficiency with signs of androgen excess and hypertension. The syndrome of apparent mineralocorticoid excess arises from mutations in the gene encoding the kidney enzyme 11β-hydroxysteroid dehydrogenase, allowing normal circulating concentrations of cortisol to activate the mineralocorticoid receptors (Mune and White, 1996). Are there readily available clinically useful means to diagnose these genetic mutations that cause secondary hypertension? Theoretically, a specially designed DNA microarray could provide the answer. Unfortunately, such high-throughput genomic tools are not available yet to the clinician, not even for the known forms of monogenic hypertension.

PROGNOSIS

Hypertension constitutes a major risk factor for the development of heart disease and stroke, end-stage renal disease, and peripheral vascular disease and is a chief contributor to adult disability. The prognosis of hypertension is solely and entirely dependent on the development of such target organ damage. Even though the rule in general is that the higher the blood pressure, the greater the likelihood to develop organ damage, clinical observations indicate that not all individuals are similarly susceptible to develop organ damage. Certain individuals develop severe hypertension-related organ damage even when blood pressure levels are not high, and others are highly resistant to end-organ damage despite very high levels of blood pressure. The explanation for this interindividual variability in the susceptibility to develop end-organ damage lies in genetic determinants of target organ damage. In the following sections, we will discuss what is known about genetic determinants of end-organ susceptibility in hypertension.

Stroke, the most dramatic expression of cerebrovascular accident (CVA) and one of the most common complications of hypertension, is a complex trait and not as a mere consequence of hypertension. Like other complex diseases, stroke appears to result from an interaction between several genetic and environmental factors. The identification of the genetic determinants of stroke, much like hypertension, is a difficult task in humans, due to the genetic heterogeneity of human populations and the confounding presence of other risk factors. It is nonetheless clear that there is a genetic predisposition to the interaction between stroke and hypertension. Clinical observations have shown that subjects with well-controlled hypertension can develop strokes, whereas others with severe hypertension may not develop stroke. What renders some individuals sensitive to hypertension-related stroke whereas others resistant is unclear, although individual genetic susceptibility must come into play. The genes that render susceptibility to stroke have been subject of intensive research over the past decade (Rubattu et al., 2004b). Much of the research has been using an animal model of stroke, the stroke-prone spontaneously hypertensive rat (Nagaoka et al., 1976; Yamori et al., 1992). Among the genes that these studies have uncovered and that predispose to hypertension-related stroke is the gene encoding atrial natriuretic peptide (ANP) which has been identified in the stroke-prone spontaneously hypertensive rat but also in two different human populations (Rubattu et al., 1999a, b). The gene encoding fibrinogen is another stroke-related gene. Elevated fibrinogen levels have also been suggested as a factor that increases the risk of stroke and carriers of the A allele of the fibrinogen-455G/A polymorphism have increased plasma fibrinogen levels. Hypertensive patients carrying the A allele have been found to have a four-fold increased risk for lacunar infarcts (Martiskainen et al., 2003). The role of the renin-angiotensin-aldosterone system (RAAS) genes in predisposing to hypertension-related stroke has been under intensive scrutiny, but the only positive association that has been detected so far is between ischemic stroke and the AT1 receptor C1166/AT1 gene allelic variant (Rubattu et al., 2004a). No other RAAS gene has been identified as contributing to stroke.

Hypertensive heart disease is a term applied generally to heart diseases that are caused by direct or indirect effects of elevated blood pressure. Prolonged elevation of blood pressure can lead to a variety of changes in the myocardial structure, coronary vasculature, and conduction system of the heart leading to

the development of left ventricular hypertrophy, coronary artery disease, various conduction system diseases, and systolic and diastolic dysfunction of the myocardium (Diamond and Phillips, 2005; Gradman and Alfayoumi, 2006; Prisant, 2005). The clinical manifestations are angina pectoris, myocardial infarction, cardiac arrhythmias (especially atrial fibrillation) and congestive heart failure. Even though hypertension predisposes to hypertensive heart disease, there appears not always to be a direct correlation between the level of blood pressure and the type and severity of damage to the heart. Individual predisposition, based on the genetic make-up, is likely to determine whether an individual with hypertension will develop hypertensive heart disease, the specific subtype of cardiac damage and its severity. What are the genes underlying the individual susceptibility to hypertensive-induced cardiac disease? Even though there is a vast amount of basic and clinical research that has been carried out on the genes involved in left ventricular hypertrophy, coronary heart disease, diseases of the conduction system and dysfunction of the myocardium, very little if any data are available at this time on the individual genetic susceptibility to develop these complications as result of hypertension, and certainly none is of any clinical usefulness. Stated otherwise, a vast amount of information is available on the genes involved in the various forms of heart disease secondary to hypertension, well beyond the scope of this chapter, but no data are yet available that might allow prediction who will and who will not develop these cardiac complications.

Hypertensive nephrosclerosis, another important complication of hypertension, is characterized by long-term essential hypertension, minimal proteinuria and progressive renal insufficiency (Luke, 1999). This type of nephrosclerosis accounts for 26% of patients reaching end-stage renal disease (ESRD) each year in the United States. Hypertension is considered the second most common cause of ESRD in white people (24%) and the leading cause of ESRD in black people (33%) (http://www.emedicine.com/med/topic1611.htm). Much like cerebrovascular disease and hypertensive heart disease, nephrosclerosis develops in susceptible individuals with hypertension, at times irrespective of their level of blood pressure (Rostand et al., 1989). Part of the individual susceptibility to develop nephrosclerosis is due to genetic factors. The presence of such genetic factors has been demonstrated in an experimental animal model in which two genetically different yet histocompatible kidneys were chronically and simultaneously exposed to the same blood pressure profile and metabolic environment within the same host. The kidney of the one strain was inherently much more susceptible to hypertension-induced damage than the kidney of the other strain (Churchill et al., 1997). There have been several attempts to identify the genes that underlie the susceptibility to hypertensive nephrosclerosis in both experimental models (Griffin and Bidani, 2004) and in humans (Griffin and Bidani, 2004; Hayden et al., 2003). Linkage studies in the fawn-hooded hypertensive rat followed by construction of congenic strains have suggested that genes influencing susceptibility to hypertension-associated renal failure may exist on rat chromosome 1q (St Lezin et al., 1999). In humans, several genes have been proposed as predisposing to hypertensive nephrosclerosis, including the homozygous 677TT mutation of the *MTHFR* gene (Koupepidou et al., 2005). A genetic predisposition to hypertensive nephrosclerosis has also been attributed to ethnic\racial groups. For example, in the Multiple Risk Factor Intervention Trial (MRFIT), a significant loss in kidney function was observed in African-American but not in Caucasian subjects despite similar control of blood pressure (Flack et al., 1993). Similarly, the Modification of Diet in Renal Diseases (MDRD) study demonstrated that at equivalent blood pressure, African-Americans had a greater rate of reduction in glomerular filtration rate than Caucasians (Hebert et al., 1997). One of the possible explanations for the discrepant susceptibility of African-Americans and Caucasian subjects has the prevalence of the *DD* genotype, which is more common in former than in the latter population and which has been associated with a higher prevalence of progressive renal disease (Duru et al., 1994). African-Americans with hypertension also have increased angiotensinogen mutations compared with hypertensive Caucasians (Bloem et al., 1995), but whether these are related to the increased susceptibility to nephrosclerosis remains to be determined.

Although some genetic variants have been identified or at least suggested for stroke and nephrosclerosis, the data have been so far derived from studies in select populations and have not been validated at large. Their relevance to the general population remains unclear, and their usefulness will have to be validated in future studies in well-defined populations before they can be applied to clinical practice and determination of the prognosis of the hypertensive patient.

PHARMACOGENOMICS

The complexity of essential hypertension and the uncertainty as to the underlying pathophysiology render the management of hypertension problematic. Currently, such management is based mostly on our clinical perception of high blood pressure, on our limited knowledge of the pathophysiology of hypertension and on experiments in clinical pharmacology that have successfully achieved a reduction in blood pressure. How does pharmacogenomics apply to hypertension and how can it affect our therapeutic approach to the disease?

Based on the achievements of pharmacogenomics in clinical fields other than hypertension, it is clear that this field of genomics has the potential to significantly improve the clinical management of hypertension. Pharmacogenomics can be used to predict the targeted therapeutic response of anti-hypertensive drugs, the hypotensive effect, as well as the occurrence of untoward side effects which may limit patient compliance and use of the drug. Pharmacogenomics can also be useful in the development of new anti-hypertensive drugs. The first two applications of pharmacogenomics will be discussed in this section.

The blood pressure response of individuals to any single drug is highly variable and clinically very difficult to predict.

Ideally, pharmacogenomics could improve our ability to predict the effectiveness of an anti-hypertensive drug in any given patient by correlating a genetic profile with the type of response. Such prediction, however, has been very difficult to come by, as hypertension is in most cases multifactorial and involves a complex interaction between multiple genes and environmental factors. One should also not overlook the potential contribution of gender, ethnic and racial factors to the individuals' response to anti-hypertensive drugs, which complicates matters even further. Any piece of evidence for a genomic effect may therefore need to be validated in each of the numerous population subtypes before it can be applied to clinical use. This may turn out to be a huge cost-ineffective task, unless a major gene effect is discerned that heavily impacts the anti-hypertensive response of a wide variety of populations.

Investigators in the field of pharmacogenomics, nonetheless, have made sincere efforts to try to correlate genetic variability within known candidate genes with therapeutic effects of anti-hypertensive drugs, hoping to utilize this information to improve the clinical response of patients to those drugs (Trotta et al., 2004). Are these efforts truly relevant to the field of pharmacogenomics, or do they pertain to the much older field of pharmacogenetics? There is a fundamental conceptual difference between the two terms (Yagil and Yagil, 2002). Pharmacogenetics relates to the study of how *individual genes* affect the way individuals respond to medications, whereas pharmacogenomics refers to how *the individual's genomic composition as a whole* affects the way individuals respond to medicines. Pharmacogenetics focus on how a single gene modulates the effect of drugs. Pharmacogenomics deals, on the other hand, with how the genome as a whole modulates the action of drugs, involving multiple genes at a time. The tasks of pharmacogenomics and pharmacogenetics are nonetheless similar. They consist of tailoring drug therapy to the individual by developing specific tests that allow the clinician to optimize the drug regimen. These tests ultimately aim to identify the most suitable patient-specific therapy that can reduce adverse events and improve outcome. The promise of pharmacogenomics has been to use genomic data to achieve these goals. The published literature attests, however, to the achievements of pharmacogenetics, and only little if at all to those of pharmacogenomics (Turner and Schwartz, 2005). The matter is further complicated by the differentiation within the field of pharmacogenetics between pharmacokinetics and pharmacodynamics. Pharmacokinetics deal with mechanisms that affect the level of the drug in blood and ultimately at its target and that are influenced by drug absorption, distribution, excretion and metabolism. Pharmacodynamics deal with mechanisms which determine the interaction of the drug with its target and the subsequent events in the cells, organs and systems (Schwartz and Turner, 2004).

Reviewing the achievements of pharmacogenetics in hypertension on the basis of published reports, it appears that the results have so far been limited mostly to the blood pressure response to drugs, and even there only a limited number of genetic polymorphisms of a limited number of candidate genes have so far been successfully associated with the blood pressure response to drugs (Marteau et al., 2005; Turner and Schwartz, 2005). Most of the data relate to genetic polymorphisms within the RAAS. Polymorphisms of the genes that encode for renin, angiotensinogen, angiotensin converting enzyme, angiotensin II receptor type 1, angiotensin II receptor type 2 and aldosterone synthase genes have all been associated with a hypotensive or a lack of hypotensive response to anti-hypertensive drugs including diuretics, beta-blockers, ACE inhibitors, angiotensin II receptor blockers and calcium channel blockers (Turner and Schwartz, 2005). Additional data relate to polymorphisms of genes involved in signal transduction pathways including the G-protein alpha and beta-3 subunits, the alpha-2 adrenergic and beta-1 and 2 adrenoreceptors, endothelin and adducin (Turner and Schwartz, 2005). Of particular interest is the association of the G460W adducin polymorphism and the anti-hypertensive response to thiazides which appears to be valid for some but not all populations (Schelleman et al., 2006).

Are these pharmacogenetic data useful in clinical practice? Theoretically, yes, as it should be possible to obtain the genetic profile of each patient and find the best matching hypotensive therapy for the individual patient. The difficulty lies, however, in that most patients whose blood pressure is difficult to manage require more than one anti-hypertensive drug for blood pressure control, and very often up to three or four drugs. Much of the available data that associate gene polymorphisms with the magnitude of the blood pressure response stem from studies that did not take into account drug interaction. Therefore, a gene polymorphism that has been associated with a lack of hypotensive response during monotherapy may be associated with a much improved response during combination therapy. An additional problem in applying currently available data is that the studies that have reported positive associations have used different combinations and different drug doses in populations of varying ages, sizes and ethnicity, each and all having a potential impact on the resulting data (Marteau et al., 2005). As a result, the clinical usefulness of the currently available pharmacogenetics data with regard to tailoring anti-hypertensive therapy to the patient is at best questionable at this time. One possible exception is the adducin paradigm in which six linkage studies, 18 of 20 association studies, and four of five follow-up studies that measured organ damage in hypertensive patients suggest a potential clinical impact of adducin polymorphism on the management of hypertension (Manunta and Bianchi, 2006). Are patients currently being tested for the adducin polymorphism prior to being prescribed a diuretic? The answer lies in the fact that it is currently much simpler to prescribe the diuretic and wait 1 month to assess the response than to send DNA for adducin genotyping. Should patient genotyping become in the future a routine procedure, it is possible that clinicians might take into consideration the adducin genotype prior to prescribing a thiazide diuretic for hypertension.

In addition to predicting the therapeutic response to drugs, an important task of pharmacogenomics would be to predict untoward side effects to drugs. One example is the

development of cough when using angiotensin converting enzyme (ACE) inhibitors; another is orthostatic hypotension when using alpha blockers. It is likely that some individuals are more prone to develop such side effects than others, and that genetically mediated factors are involved. The task of pharmacogenomics is to identify the genes involved and specific polymorphisms that are associated with such side effects. Some data are already available, for example with respect to ACEI and cough, which has been associated with the I/D ACE gene polymorphism in some studies (Takahashi et al., 2001) but not in others (Zee et al., 1998). Attempts to associate this cough with polymorphisms of the genes encoding for chymase and the B2-bradykinin receptors have failed (Zee et al., 1998). The clinical implementation of such data, however scarce they may be, has not materialized as yet.

The conclusion at this time is that genetic variants of candidate genes are promising tools for individualizing anti-hypertensive therapy, but that they are currently of little value in the clinical practice of hypertension. It is nonetheless important to be aware of the available data, as they most likely represent early reports of a field that is bound to break out and expand from pharmacogenetics into pharmacogenomics. Further well-disciplined research is required at a global scale to produce data that will become useful for the clinical management of hypertension.

2009 UPDATE

Despite the increasing availability of novel technological and bioinformatic tools, the search for the genes that are responsible for the common forms of hypertension is continuing to evolve into an ever-increasing challenge. The reason may lie in the accruing awareness of the complexity of the genome that bears direct relevance to the pathophysiology of hypertension, including mounting evidence for pleiotropy (one gene accounting for multiple traits), epistasis (gene-to-gene interaction) and molecular heterosis (Shih and O'Connor, 2008); for genome-environment (ecogenetic) (Franks, 2008), genome–sex (Cowley, 2006) and genome–ethnic interactions (Minor et al., 2008); as well as for the existence of genomic variations within protein-coding genes and within regulatory regions, including micro-RNAs (McBride et al., 2006) and copy number variations (Hamet and Seda, 2007). The task of unraveling the genetic basis of hypertension is further complicated by our current understanding that there is need to incorporate into the investigation not only the nuclear genome but also the mitochondrial genome (Puddu et al., 2007), the transcriptome, the proteome and the phenome, each representing a highly challenging task in itself. Nonetheless, the search for the genetic and pathophysiological basis of hypertension is relentless.

Currently, the emphasis is shifting from studies in experimental models of hypertension, which are amenable to genetic manipulations and which continue to generate large amount of genomic data, to large-scale studies of human populations in which associations are sought with genetic loci. There have been multiple such studies, notable amongst which are attempts to perform high-resolution mapping using microsatellite markers by the Japanese (Yatsu et al., 2007) and by the US NHLBI Family Blood Pressure Program (Gu et al., 2007), which have successfully yielded multiple additional hypertension-related loci. Also notable are recent major efforts in the form of genome-wide association studies using SNPs, such as the attempt by the Wellcome Trust Case Control Consortium (2007) to identify genes involved in common human diseases, including hypertension, which unfortunately failed to identify genes from physiological systems previously implicated by clinical or experimental studies in hypertension. The Framingham Heart Study of the US NHLBI (Levy et al., 2007), which similarly attempted to find hypertension-related genes, also failed to detect significant associations relating to blood pressure. In both studies, however, when the level of stringency was reduced, associations could be detected between multiple genomic loci and blood pressure, some of which confirmed previously identified loci, but none of which were truly robust. Hypertension research now faces the challenge of how to proceed after such major genome-wide association experiments failed to provide robust candidate genes.

The question arises whether hypertension-related genes of major impact may ever be found, as the concept that a large number of genes with small effects is becoming increasingly entrenched in our understanding of the pathophysiology of hypertension (Charchar et al., 2008). A realistic view is that, despite the advances in genomics, biotechnology and bioinformatics, a full understanding of the genetic basis of hypertension will not be quickly achieved (Cowley, 2006). Much more effort and investment is required both in experimental models of hypertension and in studies directly in humans with cross validation between the two, before genetic and genomic variants are found that will lead to new insights into the genes and regulatory pathways underlying essential hypertension. Applying large-scale genomic analyses combined with pathway-centered physiological profiling and novel computational approaches in animal models and in humans should enable the multidimensional integration that is required to further our understanding of the pathophysiology of hypertension.

2009 UPDATE REFERENCES

Charchar, F., Zimmerli, L. and Tomaszewski, M. (2008). The pressure of finding human hypertension genes: New tools, old dilemmas. *J Hum Hypertens* 22(12), 821–828.

Cowley, A.W., Jr. (2006). The genetic dissection of essential hypertension. *Nat Rev Genet* 7(11), 829–840.

Franks, P.W. (2008). Identifying genes for primary hypertension: methodological limitations and gene-environment interactions. *J Hum Hypertens* 23(4), 227–237.

Gu, C.C., Hunt, S.C., Kardia, S. et al. (2007). An investigation of genome-wide associations of hypertension with microsatellite markers in the family blood pressure program (FBPP). *Hum Genet* 121(5), 577–590.

Hamet, P. and Seda, O. (2007). Current status of genome-wide scanning for hypertension. *Curr Opin Cardiol* 22(4), 292–297.

Levy, D., Larson, M.G., Benjamin, E.J. et al. (2007). Framingham Heart Study 100K Project: Genome-wide associations for blood pressure and arterial stiffness. *BMC Med Genet* 8(Suppl 1), S3.

McBride, M.W., Graham, D., Delles, C. and Dominiczak, A.F. (2006). Functional genomics in hypertension. *Curr Opin Nephrol Hypertens* 15(2), 145–151.

Minor, D.S., Wofford, M.R. and Jones, D.W. (2008). Racial and ethnic differences in hypertension. *Curr Atheroscler Rep* 10(2), 121–127.

Puddu, P., Puddu, G.M., Cravero, E. et al. (2007). The putative role of mitochondrial dysfunction in hypertension. *Clin Exp Hypertens* 29(7), 427–434.

Shih, P.B. and O'Connor, D.T. (2008). Hereditary determinants of human hypertension: Strategies in the setting of genetic complexity. *Hypertension* 51, 1456–1464.

Wellcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447(7145), 661–678.

Yatsu, K., Mizuki, N., Hirawa, N. et al. (2007). High-resolution mapping for essential hypertension using microsatellite markers. *Hypertension* 49(3), 446–452.

ACKNOWLEDGEMENTS

Disclosure of conflicts of interests: The authors of this manuscript have no affiliation associated with financial benefit or ownership of a company relating to the contents of this manuscript.

REFERENCES

Bloem, L.J., Manatunga, A.K., Tewksbury, D.A. and Pratt, JH. (1995). The serum angiotensinogen concentration and variants of the angiotensinogen gene in white and black children. *J Clin Invest* 95, 948–953.

Churchill, P.C., Churchill, M.C., Bidani, A.K., Griffin, K.A., Picken, M., Pravenec, M., Kren, V., St Lezin, E., Wang, J.M., Wang, N. et al. (1997). Genetic susceptibility to hypertension-induced renal damage in the rat. Evidence based on kidney-specific genome transfer. *J Clin Invest* 100, 1373–1382.

Cooper, R.S. and Kaufman, J.S. (1998). Race and hypertension: Science and nescience. *Hypertension* 32, 813–816.

Cowley, A.W. Jr, Liang, M., Roman, R.J., Greene, A.S. and Jacob, H.J. (2004a). Consomic rat model systems for physiological genomics. *Acta Physiol Scand* 181, 585–592.

Cowley, A.W. Jr, Roman, R.J. and Jacob, H.J. (2004b). Application of chromosomal substitution techniques in gene-function discovery. *J Physiol* 554, 46–55.

Diamond, J.A. and Phillips, R.A. (2005). Hypertensive heart disease. *Hypertens Res* 28, 191–202.

Duru, K., Farrow, S., Wang, J.M., Lockette, W. and Kurtz, T. (1994). Frequency of a deletion polymorphism in the gene for angiotensin converting enzyme is increased in African-Americans with hypertension. *Am J Hypertens* 7, 759–762.

Flack, J.M., Neaton, J.D., Daniels, B. and Esunge, P. (1993). Ethnicity and renal disease: Lessons from the Multiple Risk Factor Intervention Trial and the Treatment of Mild Hypertension Study. *Am J Kidney Dis* 21, 31–40.

Garrett, M.R., Joe, B., Dene, H. and Rapp, J.P. (2002). Identification of blood pressure quantitative trait loci that differentiate two hypertensive strains. *J Hypertens* 20, 2399–2406.

Geller, D.S. (2004). A genetic predisposition to hypertension? *Hypertension* 44, 27–28.

Gradman, A.H. and Alfayoumi, F. (2006). From left ventricular hypertrophy to congestive heart failure: Management of hypertensive heart disease. *Prog Cardiovasc Dis* 48, 326–341.

Griffin, K.A. and Bidani, A.K. (2004). Hypertensive renal damage: Insights from animal models and clinical relevance. *Curr Hypertens Rep* 6, 145–153.

Guyton, A.C. and Coleman, T.G. (1969). Quantitative analysis of the pathophysiology of hypertension. *Circ Res* 24, 1–19.

Hayden, P.S., Iyengar, S.K., Schelling, J.R. and Sedor, J.R. (2003). Kidney disease, genotype and the pathogenesis of vasculopathy. *Curr Opin Nephrol Hypertens* 12, 71–78.

Hebert, L.A., Kusek, J.W., Greene, T., Agodoa, L.Y., Jones, C.A., Levey, A.S., Breyer, J.A., Faubert, P., Rolin, H.A. and Wang, S.R. (1997). Effects of blood pressure control on progressive renal disease in blacks and whites Modification of Diet in Renal Disease Study Group. *Hypertension* 30, 428–435.

Kearney, P.M., Whelton, M., Reynolds, K., Whelton, P.K. and He, J. (2004). Worldwide prevalence of hypertension: A systematic review. *J Hypertens* 22, 11–19.

Koupepidou, P., Deltas, C., Christofides, T.C., Athanasiou, Y., Zouvani, I. and Pierides, A. (2005). The MTHFR 677TT and 677CT/1298AC

genotypes in Cypriot patients may be predisposing to hypertensive nephrosclerosis and chronic renal failure. *Int Angiol* 24, 287–294.

Lifton, R.P., Dluhy, R.G., Powers, M., Rich, G.M., Cook, S., Ulick, S. and Lalouel, J-M. (1992). A chimaeric 11B-hydroxylase/aldosterone synthase gene causes glucocorticoid-remediable aldosteronism and human hypertension. *Nature* 355, 262–265.

Lifton, R.P., Gharavi, A.G. and Geller, D.S. (2001). Molecular mechanisms of human hypertension. *Cell* 104, 545–556.

Luft, F.C. (2001). Monogenic hypertension: Lessons from the genome. *Kidney Intl* 60, 381–390.

Luke, R.G. (1999). Hypertensive nephrosclerosis: Pathogenesis and prevalence: Essential hypertension is an important cause of end-stage renal disease. *Nephrol Dial Transplant* 14, 2271–2278.

Manunta, P. and Bianchi, G. (2006). Pharmacogenomics and pharmacogenetics of hypertension: Update and perspectives – the adducin paradigm. *J Am Soc Nephrol* 17, S30–S35.

Marteau, J.B., Zaiou, M., Siest, G. and Visvikis-Siest, S. (2005). Genetic determinants of blood pressure regulation. *J Hypertens* 23, 2127–2143.

Martiskainen, M., Pohjasvaara, T., Mikkelsson, J., Mantyla, R., Kunnas, T., Laippala, P., Ilveskoski, E., Kaste, M., Karhunen, P.J. and Erkinjuntti, T. (2003). Fibrinogen gene promoter – 455 A allele as a risk factor for lacunar stroke. *Stroke* 34, 886–891.

Mune, T. and White, P.C. (1996). Apparent mineralocorticoid excess: Genotype is correlated with biochemical phenotype. *Hypertension* 27, 1193–1199.

Nagaoka, A., Iwatsuka, H., Suzuoki, Z. and Okamoto, K. (1976). Genetic predisposition to stroke in spontaneously hypertensive rats. *Am J Physiol* 230, 1354–1359.

O'Donnell, C.J. and Kannel, W.B. (1998). Is there a racial predisposition to hypertension?. *Hypertension* 32, 817–819.

Pickering, T.G., Shimbo, D. and Haas, D. (2006). Ambulatory blood-pressure monitoring. *New Engl J Med* 354, 2368–2374.

Prisant, L.M. (2005). Hypertensive heart disease. *J Clin Hypertens (Greenwich)* 7, 231–238.

Rapp, J.P. (1982). Dahl salt-susceptible and salt-resistant rats. *Hypertension* 4, 753–763.

Rostand, S.G., Brown, G., Kirk, K.A., Rutsky, E.A. and Dustan, H.P. (1989). Renal insufficiency in treated essential hypertension. *New Engl J Med* 320, 684–688.

Rubattu, S., Di Angelantonio, E., Stanzione, R., Zanda, B., Evangelista, A., Pirisi, A., De Paolis, P., Cota, L., Brunetti, E. and Volpe, M. (2004a). Gene polymorphisms of the renin-angiotensin-aldosterone system and the risk of ischemic stroke: a role of the A1166C/AT1 gene variant. *J Hypertens* 22, 2129–2134.

Rubattu, S., Gigante, B., Stanzione, R., De Paolis, P., Tarasi, D. and Volpe, M. (2004b). In the search for stroke genes: A long and winding road. *Am J Hypertens* 17, 197–202.

Rubattu, S., Lee-Kirsch, M.A., DePaolis, P., Giliberti, R., Gigante, B., Lombardi, A., Volpe, M. and Lindpaintner, K. (1999a). Altered structure, regulation, and function of the gene encoding the atrial natriuretic peptide in the stroke-prone spontaneously hypertensive rat. *Circ Res* 85, 900–905.

Rubattu, S., Ridker, P., Stampfer, M.J., Volpe, M., Hennekens, C.H. and Lindpaintner, K. (1999b). The gene encoding atrial natriuretic peptide and the risk of human stroke. *Circulation* 100, 1722–1726.

Schelleman, H., Stricker, B.H., Verschuren, W.M., de Boer, A., Kroon, A.A., de Leeuw, P.W., Kromhout, D. and Klungel, O.H. (2006). Interactions between five candidate genes and antihypertensive drug therapy on blood pressure. *Pharmacogenomics J* 6, 22–26.

Schwartz, G.L. and Turner, S.T. (2004). Pharmacogenetics of antihypertensive drug responses. *Am J Pharmacogenomics* 4, 151–160.

St Lezin, E., Griffin, K.A., Picken, M., Churchill, M.C., Churchill, P.C., Kurtz, T.W., Liu, W., Wang, N., Kren, V., Zidek, V. et al. (1999). Genetic isolation of a chromosome 1 region affecting susceptibility to hypertension-induced renal damage in the spontaneously hypertensive rat. *Hypertension* 34, 187–191.

Takahashi, T., Yamaguchi, E., Furuya, K. and Kawakami, Y. (2001). The ACE gene polymorphism and cough threshold for capsaicin after cilazapril usage. *Respir Med* 95, 130–135.

Trotta, R., Donati, M.B. and Iacoviello, L. (2004). Trends in pharmacogenomics of drugs acting on hypertension. *Pharmacol Res* 49, 351–356.

Turner, S.T. and Schwartz, G.L. (2005). Gene markers and antihypertensive therapy. *Curr Hypertens Rep* 7, 21–30.

White, P.C., Curnow, K.M. and Pascoe, L. (1994). Disorders of steroid 11 beta-hydroxylase isozymes. *Endocr Rev* 15, 421–438.

Wilson, F.H., Disse-Nicodeme, S., Choate, K.A., Ishikawa, K., Nelson-Williams, C., Desitter, I., Gunel, M., Milford, D.V., Lipkin, G.W., Achard, J.M. et al. (2001). Human hypertension caused by mutations in WNK kinases. *Science* 293, 1107–1112.

Yagil, C., Katni, G., Rubattu, S., Stolpe, C., Kreutz, R., Lindpaintner, K., Ganten, D., Ben-Ishay, D. and Yagil, Y. (1996). Development, genotype and phenotype of a new colony of the Sabra hypertension prone (SBH/y) and resistant (SBN/y) rat model of slat sensitivity and resistance. *J Hypertens* 14, 1175–1182.

Yagil, Y. and Yagil, C. (2002). Insights into pharmacogenomics and its impact upon immunosuppressive therapy. *Transpl Immunol* 9, 203–209.

Yagil, Y. and Yagil, C. (2003). Congenics in the pathway from quantitative trait loci detection to gene identification: Is that the way to go? *J Hypertens* 21, 2009–2011.

Yamori, Y., Nara, Y., Mizushima, S., Murakami, S., Ikeda, K., Sawamura, M., Nabika, T. and Horie, R. (1992). Gene-environment interaction in hypertension, stroke and atherosclerosis in experimental models and supportive findings from a world-wide cross-sectional epidemiological survey: A WHO-cardiac study. *Clin Exp Pharmacol Physiol* (Suppl) 20, 43–52.

Zee, R.Y.L., Rao, V.S., Paster, R.Z., Sweet, C.S. and Lindpainter, K. (1998). Three candidate genes and angiotensin-converting enzyme inhibitor-cough related: A pharmacogenetic analysis. *Hypertension* 31, 925–928.

RECOMMENDED RESOURCES

http://cmbi.bjmu.edu.cn/genome/candidates/candidates.html
A list of 150 candidate genes for hypertension with relevant genomic data including referral to known polymorphisms within those genes.

http://www.emedicine.com/med/topic1611.htm
An extensive and updated overview of various important aspects of hypertensive nephrosclerosis.

http://www.sin-italy.org/jnonline/Vol14s4/Fogo/FOGO.htm
Another extensive and updated overview of various important aspects of hypertensive nephrosclerosis.

http://www.emedicine.com/med/topic3432.htm
An extensive and updated overview of various important aspects of hypertensive heart disease.

CHAPTER

Lipoprotein Disorders

Sekar Kathiresan and Daniel J. Rader

INTRODUCTION

Plasma lipoproteins are integral to energy and cholesterol metabolism, but disorders involving lipoprotein metabolism can predispose to atherosclerotic vascular disease (ASCVD). Genetic factors play an important role in influencing lipoprotein metabolism and therefore plasma levels of the major lipoproteins and risk for cardiovascular disease. Molecular characterization of classic Mendelian monogenic lipoprotein disorders has provided major insights into the physiology and regulation of lipoprotein metabolism and new targets for therapeutic drug development. Much attention is now focused on greater elucidation of the genetic factors that influence the complex lipoprotein phenotypes that are much more common and important in influencing cardiovascular risk in the general population. Lipoprotein metabolism is a ripe area for the application of genomic medicine because of frequency with which plasma lipids are measured in clinical practice, the quantitative importance of genetics in determining their levels, the large number of gene products involved in lipoprotein metabolism, and the broad clinical relevance of the field to the most important cause of morbidity and mortality in most of the world.

OVERVIEW OF LIPOPROTEIN METABOLISM

Lipoproteins are large macromolecular complexes that transport hydrophobic lipids (primarily triglycerides, cholesterol, and fat-soluble vitamins) through body fluids (plasma, interstitial fluid, and lymph) to and from tissues. Lipoproteins play an essential role in the absorption of dietary cholesterol, long-chain fatty acids, and fat-soluble vitamins; the transport of triglycerides, cholesterol, and fat-soluble vitamins from the liver to peripheral tissues; and the transport of cholesterol from peripheral tissues to the liver. Lipoproteins contain a core of hydrophobic lipids (triglycerides and cholesteryl esters) surrounded by hydrophilic lipids (phospholipids, unesterified cholesterol) and proteins that interact with body fluids. The plasma lipoproteins are divided into five major classes based on their relative density: chylomicrons, very-low-density lipoproteins (VLDL), intermediate-density lipoproteins (IDL), low-density lipoproteins (LDL), and high-density lipoproteins (HDL). Each lipoprotein class comprises a family of particles that vary slightly in density, size, migration during electrophoresis, and protein composition. The density of a lipoprotein is determined by the amount of lipid per particle. HDL is the smallest and most dense lipoprotein, whereas chylomicrons and VLDL are the largest and least dense lipoprotein particles. Most plasma triglyceride is transported in chylomicrons or VLDL and most plasma cholesterol is carried as cholesteryl esters in LDL and HDL.

The proteins associated with lipoproteins, called apolipoproteins, are required for the assembly, structure, and function of lipoproteins. Apolipoproteins activate enzymes important in lipoprotein metabolism and act as ligands for cell surface receptors. ApoA-I, which is synthesized in the liver and intestine, is found on virtually all HDL particles. ApoA-II is the second most abundant HDL apolipoprotein and is on approximately two-thirds of all HDL particles. ApoB is the major structural

Copyright © 2010, Elsevier Inc.
All rights reserved.

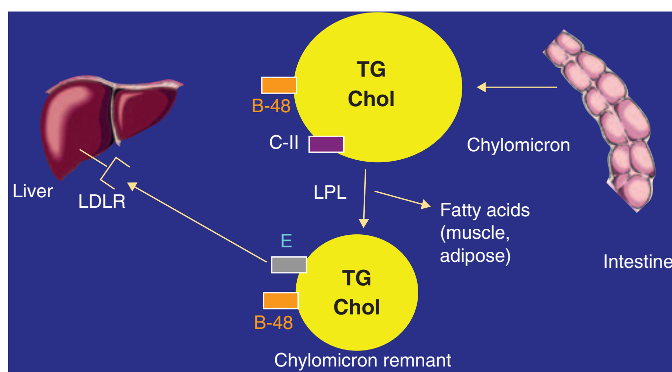


Figure 23.1 The exogenous pathway of apoB-containing lipoprotein metabolism. This pathway transports exogenous dietary fat from intestine to peripheral tissues and ultimately to the liver.

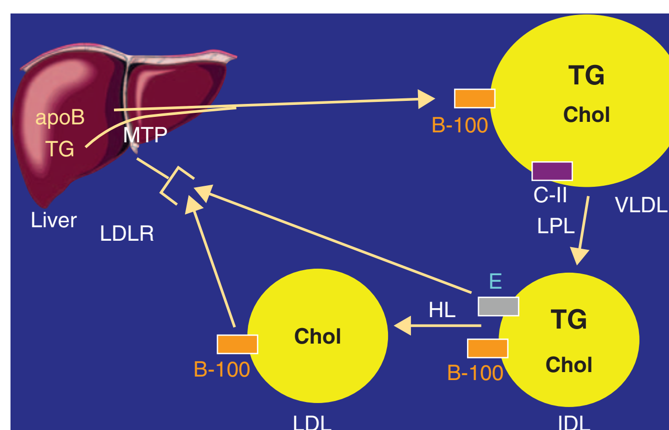


Figure 23.2 The endogenous pathway of apoB-containing lipoprotein metabolism. This pathway transports endogenous stored fat from liver to peripheral tissues and ultimately back to the liver.

protein of chylomicrons, VLDL, IDL, and LDL; one molecule of apoB, either apoB-48 (chylomicrons) or apoB-100 (VLDL, IDL, or LDL), is present on each lipoprotein particle. The human liver synthesizes apoB-100 and the intestine makes apoB-48, which is derived from the same gene by mRNA editing. ApoE is present in multiple copies on chylomicrons, VLDL and IDL and plays a critical role in the metabolism and clearance of triglyceride-rich particles. ApoC-I, apoC-II, and apoC-III also participate in the metabolism of triglyceride-rich lipoproteins.

The exogenous pathway of lipoprotein metabolism involves the absorption and transport of dietary lipids to appropriate sites within the body (Figure 23.1). Dietary triglycerides are hydrolyzed by lipases within the intestinal lumen and emulsified with bile acids to form micelles. Dietary cholesterol, fatty acids, and fat-soluble vitamins are absorbed in the proximal small intestine. Cholesterol and retinol are esterified (by the addition of a fatty acid) in the enterocyte to form cholesteryl esters and retinyl esters, respectively. Longer-chain fatty acids (>12 carbons) are incorporated into triglycerides and packaged with apoB-48, cholesteryl esters, retinyl esters, phospholipids, and cholesterol to form chylomicrons. Nascent chylomicrons are secreted into the intestinal lymph and delivered through the thoracic duct directly to the systemic circulation, where they are extensively processed by peripheral tissues before reaching the liver. The particles encounter lipoprotein lipase (LPL), which is anchored to proteoglycans that decorate the capillary endothelial surfaces of adipose tissue, heart and skeletal muscle. The triglycerides of chylomicrons are hydrolyzed by LPL and free fatty acids (FFAs) are released; apoC-II, which is transferred to circulating chylomicrons from HDL, acts as a cofactor for LPL in this reaction. The released FFAs are taken up by adjacent myocytes or adipocytes and either oxidized to generate energy or re-esterified and stored as triglyceride. Some of the released FFAs bind albumin before entering cells, and are transported to other tissues, especially the liver. The chylomicron particle progressively shrinks in size as the hydrophobic core is hydrolyzed and the hydrophilic lipids (cholesterol and phospholipids) and apolipoproteins on the particle surface are transferred to HDL, creating chylomicron remnants. Chylomicron remnants are rapidly removed from the circulation by the liver through a process that requires apoE as a ligand for receptors in the liver.

The endogenous pathway of lipoprotein metabolism refers to the hepatic secretion of apoB-containing lipoproteins and their metabolism (Figure 23.2). VLDL particles resemble chylomicrons in protein composition but contain apoB-100 rather than apoB-48 and have a higher ratio of cholesterol to triglyceride (~1 mg of cholesterol for every 5 mg of triglyceride). The triglycerides of VLDL are derived predominantly from the esterification of long-chain fatty acids in the liver. The packaging of hepatic triglycerides with the other major components of the nascent VLDL particle (apoB-100, cholesteryl esters, phospholipids, and vitamin E) requires the action of the enzyme microsomal triglyceride transfer protein (MTP). After secretion into the plasma, VLDL acquires multiple copies of apoE and apolipoproteins of the C series by transfer from HDL. As with chylomicrons, the triglycerides of VLDL are hydrolyzed by LPL, especially in muscle and adipose tissue. After the VLDL remnants dissociate from LPL, they are referred to as IDL, which contain roughly similar amounts of cholesterol and triglyceride. The liver removes approximately 40–60% of IDL by LDL receptor–mediated endocytosis via binding to apoE. The remainder of IDL is remodeled by hepatic lipase (HL) to form LDL; during this process most of the triglyceride in the particle is hydrolyzed and all apolipoproteins except apoB-100 are transferred to other lipoproteins. The cholesterol in LDL accounts for over half of the plasma cholesterol in most individuals. Approximately 70% of circulating LDL is cleared by LDL receptor–mediated endocytosis in the liver. Lipoprotein(a) [Lp(a)] is a lipoprotein similar to LDL in lipid and protein composition, but contains an additional protein called apolipoprotein(a) [apo(a)]. Apo(a) is synthesized

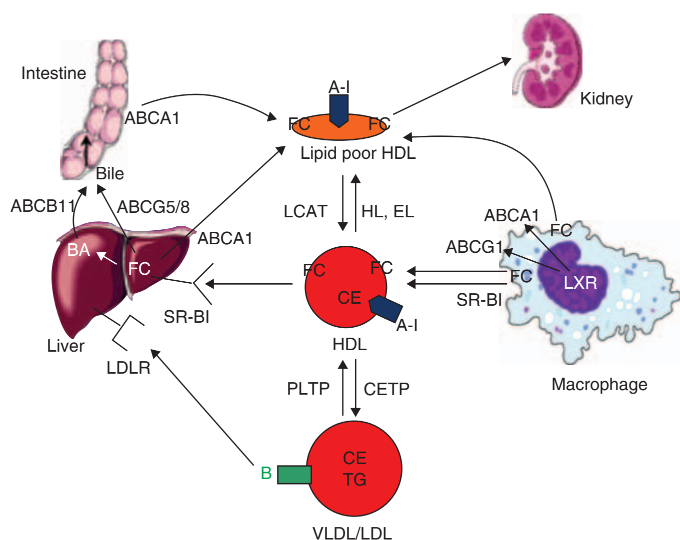


Figure 23.3 HDL metabolism and reverse cholesterol transport. This pathway transports excess cholesterol from the periphery back to the liver for excretion in the bile and feces.

in the liver and is attached to apoB-100 by a disulfide linkage. The major site of clearance of Lp(a) is the liver but the uptake pathway is not known.

HDL metabolism is complex (Figure 23.3) (Rader, 2006). Nascent HDL particles are synthesized by the intestine and the liver. Newly secreted apoA-I rapidly acquires phospholipids and unesterified cholesterol from its site of synthesis (intestine or liver) via efflux promoted by the membrane protein ATP-binding cassette protein A1 (ABCA1). This process results in the formation of discoidal HDL particles, which then recruit additional unesterified cholesterol from the periphery. Within the HDL particle, the cholesterol is esterified by lecithin-cholesterol acyltransferase (LCAT), a plasma enzyme associated with HDL, and the more hydrophobic cholesteryl ester moves to the core of the HDL particle. As HDL acquires more cholesteryl ester it becomes spherical, and additional apolipoproteins and lipids are transferred to the particles from the surfaces of chylomicrons and VLDL during lipolysis. HDL cholesterol (HDL-C) is transported to hepatoctyes by both an indirect and a direct pathway. HDL cholesteryl esters can be transferred to apoB-containing lipoproteins in exchange for triglyceride by the cholesteryl ester transfer protein (CETP). The cholesteryl esters are then removed from the circulation by LDL receptor–mediated endocytosis. HDL-C can also be taken up directly by hepatocytes via the scavenger receptor class BI (SR-BI), a cell surface receptor that mediates the selective transfer of lipids to cells. HDL particles undergo extensive remodeling within the plasma compartment by a variety of lipid transfer proteins and lipases. The phospholipid transfer protein (PLTP) has the net effect of transferring phospholipids from other lipoproteins to HDL. After CETP-mediated lipid exchange, the triglyceride-enriched HDL becomes a much better substrate for HL, which hydrolyzes the triglycerides and phospholipids to generate smaller HDL particles. A related enzyme called endothelial lipase (EL) hydrolyzes HDL phospholipids, generating smaller HDL particles that are catabolized faster. Remodeling of HDL influences the metabolism, function, and plasma concentrations of HDL.

PLASMA LIPID AND LIPOPROTEIN LEVELS AND ATHEROSCLEROTIC CARDIOVASCULAR DISEASE

Plasma total cholesterol was firmly established as an independent risk factor for cardiovascular disease in 1961 after researchers at the Framingham Heart Study demonstrated that participants with total cholesterol >245 mg/dl had a threefold increased risk of future coronary heart disease (CHD) compared with participants with a total cholesterol <210 mg/dl (Kannel et al., 1961). Subsequently, it was clarified that the two major lipoproteins carrying cholesterol, LDL and HDL, are associated with opposite influences on risk for cardiovascular disease, with LDL cholesterol (LDL-C) associated with increased risk and HDL-C with decreased risk. In observational studies, every 1 mg/dl increase in LDL-C has been shown to be associated with a 2% increased risk for cardiovascular disease, whereas every 1 mg/dl increase in HDL-C is associated with a 2–3% decreased risk (Gordon and Rifkind, 1989). For both LDL-C and HDL-C, there is a continuous, graded relationship between blood levels and subsequent risk for cardiovascular disease. Thus genetic factors that influence plasma levels of LDL-C and HDL-C are critically important to the risk of developing cardiovascular disease. The independent relationship of triglyceride levels and cardiovascular disease has been a topic of debate over the years, but recently a general consensus has developed that triglycerides are an independent predictor of risk for cardiovascular disease. Thus genetic factors that influence triglycerides are also important determinants of risk for cardiovascular disease.

INHERITED BASIS FOR BLOOD LIPID TRAITS

Though lipid levels are affected by many nongenetic factors, interindividual variability in lipids has been shown to have a strong inherited component. In simplest terms, an important role for shared genes is suggested by the fact that the correlation between family members for LDL-C, HDL-C, or triglycerides is considerably greater than that between unrelated individuals. Heritability, or the proportion of total phenotypic variance that is due to genetic variance, has been consistently estimated to be ~50% for the major blood lipid traits. For example, in the Framingham Heart Study, heritability for single time-point measurements of LDL-C, HDL-C, and triglycerides are 0.59, 0.52, and 0.48 (Kathiresan et al., 2007).

Broadly, traits that have an inherited basis may display a simple (or Mendelian) pattern of inheritance where variation at a single genetic locus is both necessary and sufficient to cause a phenotype. Alternatively, traits may depend on multiple genetic loci (or multifactorial inheritance). For blood lipid disorders, both patterns of inheritance are operational. Mendelian syndromes where LDL-C, HDL-C, or triglyceride levels are extremely high or low have been successfully studied and a number of specific genes have been isolated (see below). Knowledge derived from these genes has transformed our understanding of lipoprotein biology and treatment of cardiovascular disease (Goldstein and Brown, 2001). However, each of these syndromes is individually rare in the population and cannot explain the overall heritability of blood lipids.

Instead, blood lipid variation in the population depends on the additive effects of multiple loci. As initially demonstrated by the British geneticist R.A. Fisher in 1918 and now widely accepted, the additive effects of alleles at multiple loci can lead to a continuous trait that is normally distributed in the population. This is likely to be the case for blood lipid levels. However, several aspects of the underlying genetic architecture for lipid levels are unknown. Important unanswered questions are (1) how many loci affect blood lipid variation? and (2) at each locus, what is the number and frequency of alleles that affect risk? Emerging data from large population-based sequencing and genotyping studies suggests that the number of loci will be large, at least greater than 20. In addition, each locus is likely to harbor a spectrum of alleles rare and common in the population.

SCREENING FOR LIPID DISORDERS

Guidelines for the screening and management of lipid disorders have been established by an expert Adult Treatment Panel (ATP) convened by the National Cholesterol Education Program (NCEP) of the National Heart Lung and Blood Institute. The NCEP ATPIII guidelines published in 2001 recommend that all adults over age 20 to have plasma levels of cholesterol, triglyceride, LDL-C, and HDL-C measured after a 12-h overnight fast. In most clinical laboratories, the total cholesterol and triglycerides in the plasma are measured enzymatically and then the cholesterol in the supernatant is measured after precipitation of apoB-containing lipoproteins to determine the HDL-C. The LDL-C is estimated using the following equation: LDL-C = Total cholesterol − (Triglycerides/5) – HDL-C. (The VLDL-C is estimated by dividing the plasma triglyceride by 5, reflecting the ratio of cholesterol to triglyceride in VLDL particles.) This formula is reasonably accurate if test results are obtained on fasting plasma and if the triglyceride level does not exceed ~300 mg/dl, and by convention cannot be used if the TG are greater than 400 mg/dl. The accurate determination of LDL-C levels in patients with triglyceride levels greater than 300 mg/dl requires application of ultracentrifugal techniques or other direct assays for LDL-C. Further evaluation and treatment is based primarily on the plasma LDL-C level and the assessment of overall cardiovascular risk. The fact that screening for lipids is so widely performed has resulted in frequent identification of extreme lipid phenotypes due to Mendelian syndromes as well as widely available data for genetic studies in populations. Thus the genetics and genomics of lipids is among the most widely studied of complex genetic traits that are broadly clinically important.

GENETICS OF LDL-C

Studies of the genetic basis of substantially elevated or reduced LDL-C levels have provided major insights into the regulation of LDL metabolism and novel targets for therapeutic development. We first review the Mendelian disorders of high and low LDL-C (Table 23.1) and then review the current state of knowledge regarding common polymorphisms and their association with LDL-C levels.

Mendelian Disorders Primarily Causing Elevated LDL-C Levels

Familial Hypercholesterolemia (FH)

Familial hypercholesterolemia (FH) (Rader et al., 2003) is caused by loss-of-function mutations in the LDL receptor. Homozygous FH, caused by mutations in both LDL-receptor alleles, is a rare (approximately 1 in 1 million persons) condition characterized by markedly elevated cholesterol (greater than 500 mg/dl), cutaneous and tendon xanthomas, and accelerated atherosclerosis developing in childhood. Because they work through upregulation of the LDL receptor, statins, cholesterol absorption inhibitors, and bile acid sequestrants have only modest effects in reducing cholesterol. Liver transplantation is effective in decreasing LDL-C levels and gene therapy has been attempted. LDL apheresis is the therapy of choice at this time. Heterozygous FH is one of the most common (approximately 1 in 500 persons) single gene disorders. It is characterized by substantial elevations in LDL-C (usually 200–400 mg/dl), tendon xanthomas, and premature atherosclerotic cardiovascular disease. Treatment usually requires more than one drug, usually a statin plus a cholesterol absorption inhibitor and often a bile acid sequestrant and/or niacin. In some cases, LDL apheresis should be considered.

Familial Defective Apolipoprotein B-100 (FDB)

Familial defective apoB-100 (FDB) (Tybjaerg-Hansen and Humphries, 1992) is caused by mutations in the receptor binding region of apoB-100, which impairs its binding to the LDL receptor and delays the clearance of LDL. FDB is generally recognized as an autosomal dominant condition and occurs in approximately 1 in 700 persons of European descent. Like heterozygous FH, FDB is associated with elevated LDL-C and normal triglycerides. The most common mutation causing FDB is a substitution of glutamine for arginine at position 3500 in apoB-100; other mutations have also been reported that have a similar effect on apoB

TABLE 23.1 Mendelian disorders of lipoprotein metabolism

Genetic disorder	Gene	Lipoproteins affected	Clinical findings	Genetic transmission
LDL cholesterol				
Familial hypercholesterolemia	(*LDLR*)	↑ LDL	Tendon xanthomas, CHD	ACD
Familial defective apoB-100	(*APOB*)	↑ LDL	Tendon xanthomas, CHD	AD
Autosomal recessive hypercholesterolemia	(*ARH*)	↑ LDL	Tendon xanthomas, CHD	AR
Sitosterolemia	*ABCG5* or *ABCG8*	↑ LDL	Tendon xanthomas, CHD	AR
Autosomal dominant hypercholesterolemia	*PCSK9* (gain of function)	↑ LDL	Tendon xanthomas, CHD	AD
Abetalipoproteinemia	*MTP*	↓ LDL	Fat malabsorption, spinocerebellar degeneration, retinopathy, possible hepatic steatosis	AR
Hypobetalipoproteinemia	(*APOB*)	↓ LDL	Fat malabsorption, spinocerebellar degeneration, retinopathy, possible hepatic steatosis	ACD
PCSK9 deficiency	*PCSK9* (loss of function)	↓ LDL		ACD
HDL cholesterol				
APOA-I mutations	*APOA1*	↓ HDL		AD
Tangier disease	*ABCA1*	↓ HDL	Hepatosplenomegaly, enlarged orange tonsils, CHD	ACD
LCAT deficiency	*LCAT*	↓ HDL	Corneal opacification, hemolytic anemia, progressive renal insufficiency	AR
CETP deficiency	*CETP*	↑ HDL		AR
Triglycerides				
Lipoprotein lipase deficiency	(*LPL*)	↑ Chylomicrons	Eruptive xanthomas, hepatosplenomegaly, pancreatitis	AR
Familial apolipoprotein C-II deficiency	(*APOC2*)	↑ Chylomicrons	Eruptive xanthomas, hepatosplenomegaly	AR
Familial hepatic lipase deficiency	(*LIPC*)	↑ VLDL remnants	Premature atherosclerosis, pancreatitis	AR
Familial dysbetalipoproteinemia	(*APOE*)	↑ Chylomicron, and VLDL remnants	Palmar and tuberoeruptive xanthomas, CHD, PVD	AD

AD, autosomal dominant; ACD, autosomal co-dominant; AR, autosomal recessive.

binding to the LDL receptor. FDB is treated with statins and often additional drugs, similar to the treatment of heterozygous FH.

Autosomal Recessive Hypercholesterolemia (ARH)

Autosomal recessive hypercholesterolemia (ARH) (Garcia et al., 2001) is caused by mutations in the *ARH* gene, which produces a protein that regulates LDL receptor–mediated endocytosis in hepatocytes. ARH is very rare and in some ways the clinical presentation resembles homozygous FH. However, in contrast to FH, the condition is formally recessive and obligate heterozygotes have normal cholesterol levels. Statins and other LDL upregulating therapy sometimes result in partial LDL lowering response.

Autosomal Dominant Hypercholesterolemia (ADH)

Autosomal dominant hypercholesterolemia (ADH) is caused by apparent gain-of-function mutations in the proprotein convertase subtilisin/kexin type 9 (*PCSK9*) gene (Abifadel et al., 2003). PCSK9 is secreted by hepatocytes and appears to downregulate the density of functional LDL receptors in hepatocytes by promoting endosomal degradation rather than recycling of the receptor (Horton et al., 2007). Interestingly, loss-of-function mutations in this gene appear to cause low LDL-C levels (see below). The discovery of the molecular basis of ADH ultimately led to the identification of PCSK9 as a novel therapeutic target.

Sitosterolemia

Sitosterolemia is caused by mutations in one of two members of the adenosine triphosphate (ATP)-binding cassette (ABC) transporter family, ABCG5 and ABCG8 (Berge et al., 2000). These genes are expressed in the intestine and liver where they form a functional complex to limit intestinal absorption and promote biliary excretion of plant- and animal-derived neutral sterols. In sitosterolemia, normally the low level of intestinal absorption of plant sterols is markedly increased and biliary excretion of plant sterols is reduced, resulting in increased plasma levels of sitosterol and other plant sterols. Because the hepatic LDL receptor is downregulated, LDL-C levels tend to be high in this condition. Patients with sitosterolemia often have tendon xanthomas and are at risk for premature cardiovascular disease. Treatment of sitosterolemia is focused on dietary counseling, cholesterol-absorption inhibitors, and bile acid sequestrants, rather than statins.

Mendelian Disorders Primarily Causing Reduced LDL-C Levels

Abetalipoproteinemia

Abetalipoproteinemia is caused by mutations in the gene encoding microsomal transfer protein (MTP) (Sharp et al., 1993), a protein that transfers lipids to apoB in the ER, forming nascent chylomicrons and VLDL in the intestine and liver, respectively. It is a very rare autosomal recessive disease characterized by extremely low plasma levels of cholesterol and no detectable apoB-containing lipoproteins in plasma (Rader and Brewer, 1993). Abetalipoproteinemia is characterized clinically by fat malabsorption, spinocerebellar degeneration, pigmented retinopathy, and acanthocytosis. Most clinical manifestations of abetalipoproteinemia result from defects in the absorption and transport of fat-soluble vitamins, especially alpha tocopherol (vitamin E) which is dependent on VLDL for efficient transport out of the liver. The discovery of MTP as the basis of abetalipoproteinemia led to the concept of MTP inhibition as a novel therapeutic target for lowering LDL-C levels.

Familial Hypobetalipoproteinemia

Familial hypobetalipoproteinemia generically refers to low LDL-C levels that have a genetic basis. Historically it has been used to refer to low LDL-C due to mutations in apoB (Schonfeld et al., 2005). There is a range of missense and nonsense mutations in apoB that have been shown to reduce secretion and/or accelerate catabolism of apoB. Individuals heterozygous for these mutations generally have LDL-C levels less than 100 mg/dl. They appear to be protected from the development of ASCVD. There are rare patients with homozygous hypobetalipoproteinemia who have mutations in both apoB alleles and have plasma lipids similar to those in abetalipoproteinemia.

PCSK9 Deficiency

More recently, loss-of-function mutations in PCSK9 have also been shown to cause low LDL-C levels (Cohen et al., 2005; Kotowski et al., 2006), and therefore PCSK9 is another molecular cause of the generic phenotype of familial hypobetalipoproteinemia. The mechanism is uncertain, but presumably involves reduced PCSK9-mediated targeting of the LDL receptor to degradation pathways, resulting in upregulation of the hepatic LDL receptor and increased catabolism of LDL. This condition, which is more common in people of African descent, provided the opportunity to demonstrate that the effects of lifelong low LDL-C levels are a substantial reduction in CHD with no other adverse consequences (Cohen et al., 2006). This strongly supports the concept that aggressive LDL-C reduction is associated with long-term substantial reduction in cardiovascular risk.

Common Gene Variants Segregating in Populations with LDL-C Levels

Common gene variants within at least seven loci, *APOB*, *APOE*, *CILP2/PBX4*, *HMGCR*, *LDLR*, *PCSK9*, and *SORT1*, have been reproducibly related to LDL-C (Benn et al., 2005; Boright et al., 1998; Kathiresan et al., 2008a; Kotowski et al., 2006; Saxena et al., 2007; Sing and Davignon, 1985; Willer et al., 2008). In Table 23.2, we present representative associations of single nucleotide polymorphisms (SNPs) in these genes with LDL-C from a single community-based cohort study in southern Sweden, the Malmo Diet and Cancer Study-Cardiovascular Cohort (Berglund et al., 1993). These variants vary in frequency from 1% at *PCSK9* to 48% at *APOE*. These SNPs explain 0.5–1.7% of the interindividual variability in LDL-C in the population (Kathiresan et al., 2008b). In comparisons between the major and minor allele homozygote classes, the difference in LDL-C ranges from 4 mg/dl at *HMGCR* to 51 mg/dl at *PCSK9*. Variants in other genes have been reported to be associated with LDL-C but have been reproduced less frequently.

GENETICS OF HDL-C

Studies of the genetic basis of substantially reduced or elevated HDL-C levels have also provided important insights into the regulation of HDL metabolism and novel targets for therapeutic development. We first review the Mendelian disorders of low and high HDL-C (Table 23.1) and then review the current state of knowledge regarding common polymorphisms and their association with HDL-C levels.

Table 23.2 Association of common single nucleotide polymorphisms and low-density lipoprotein, high-density lipoprotein cholesterol or triglycerides in the Malmo Diet and Cancer Study – Cardiovascular Cohort[a]

SNPS	Gene	SNP type	MAF	M/M	M/m	m/m	% of variance explained	P[b]
LDL cholesterol								
rs693	*APOB*	Coding	0.48	AA[c] 167 ± 38 *n* = 1349	AG 160 ± 37 *n* = 2462	GG 157 ± 38 *n* = 1173	0.9	2×10^{-11}
rs4420638	*APOE cluster*	5′ Upstream	0.20	AA 157 ± 37 *n* = 3291	AG 167 ± 38 *n* = 1621	GG 173 ± 38 *n* = 224	1.7	3×10^{-21}
rs12654264	*HMGCR*	Intronic	0.39	AA 158 ± 38 1911	AT 163 ± 38 2405	TT 162 ± 38 764	0.2	0.002
rs688	*LDLR*	Coding	0.42	CC 160 ± 39 *n* = 1717	CT 161 ± 38 *n* = 2551	TT 163 ± 37 *n* = 860	0.1	0.04
rs11591147	*PCSK9*	Coding	0.01	GG 161 ± 38 *n* = 4885	GT 146 ± 30 *n* = 114	TT 89 ± 22 *n* = 2	0.5	7×10^{-7}
HDL cholesterol								
rs3890182	*ABCA1*	Intronic	0.13	GG 54 ± 15 *n* = 3818	GA 53 ± 14 *n* = 1163	AA 51 ± 12 *n* = 82	0.2	0.003
rs28927680	*APOA5*	5′ Upstream	0.07	CC 53 ± 14 *n* = 4444	CG 52 ± 14 *n* = 677	GG 46 ± 8 *n* = 28	1.1	7×10^{-14}
rs1800775	*CETP*	5′ Upstream	0.49	CC 51 ± 13 *n* = 1397	CA 54 ± 14 *n* = 2456	AA 56 ± 15 *n* = 1245	2.5	2×10^{-29}
rs1800588	*LIPC*	5′ Upstream	0.21	CC 53 ± 14 *n* = 3157	CT 54 ± 14 *n* = 1754	TT 57 ± 16 *n* = 247	0.8	4×10^{-10}
rs328	*LPL*	Coding	0.09	CC 53 ± 14 *n* = 4219	CG 56 ± 15 *n* = 863	GG 58 ± 16 *n* = 49	0.9	3×10^{-12}
Triglycerides								
rs693	*APOB*	Coding	0.48	AA 115 ± 65 *n* = 1120	AG 120 ± 73 *n* = 2385	GG 124 ± 69 *n* = 1303	0.3	7×10^{-5}
rs28927680	*APOA5*	5′ Upstream	0.07	CC 117 ± 60 *n* = 4481	CG 131 ± 68 *n* = 684	GG 156 ± 84 *n* = 29	0.4	6×10^{-6}
rs780094	*GCKR*	Intronic	0.34	CC 117 ± 70 *n* = 2207	CG 123 ± 81 *n* = 2457	GG 127 ± 73 *n* = 639	0.5	2×10^{-7}
rs328	*LPL*	Coding	0.09	CC 122 ± 72 *n* = 3996	CG 111 ± 58 *n* = 832	GG 91 ± 34 *n* = 46	0.7	9×10^{-9}

SNP, single nucleotide polymorphism; MAF, minor allele frequency; M/M, major allele homozygote; M/m, heterozygote; m/m, minor allele homozygote.
[a] Plus–minus values are means ± sd. To convert values for cholesterol to millimoles per liter, multiply by 0.02586. To convert values for triglycerides to millimoles per liter, multiply by 0.01129;
[b] Association analyses were conducted using multivariable-adjusted lipid concentration (adjusted for age, age^2, sex, and diabetes status) as the phenotype;
[c] Within each cell are the alleles for the SNP on the forward strand of the human genome reference sequence (from National Center for Biotechnology Information Build 35), the mean unadjusted cholesterol value in mg/dl plus minus standard deviation in mg/dl, and the number of individuals of that genotype class.

Mendelian Disorders Primarily Causing Reduced HDL-C Levels

ApoA-I Deficiency and Structural Mutations

A rare cause of extremely low HDL-C is complete deficiency of apoA-I either from apoA-I gene deletion or nonsense mutations which result in virtually absent plasma HDL (Ng et al., 1994; Norum et al., 1982; Schaefer et al., 1982). Most of these cases are associated with premature CHD, consistent with the concept that apoA-I is atheroprotective and supportive of the concept that apoA-I elevation could be a therapeutic strategy. Another relatively rare cause of low HDL-C are missense or nonsense mutations that result in structurally abnormal or truncated apoA-I proteins. The best known of these mutations is apoA-I_{Milano}, where a substitution of cysteine for arginine at position 173 (Chiesa and Sirtori, 2003) results in increased turnover of the mutant apoA-I_{Milano} protein, as well of the wild-type apoA-I and a substantial reduction in HDL-C. The low HDL-C levels, however, are not associated with an increased risk of atherosclerosis. Animal studies with intravenous infusion of recombinant apoA-I_{Milano} show less atherosclerosis (Chiesa and Sirtori, 2003), and a small trial of the intravenous infusion of apoA-I_{Milano}–phospholipid complexes in humans demonstrated a reduction from baseline in coronary atheroma volume as measured by intravascular ultrasound (Nissen et al., 2003). There have been several other apoA-I structural mutations described that cause low HDL-C (von Eckardstein, 2005) but structural apoA-I mutations are rare, and in the general population apoA-I mutations are not a common source of variation in HDL-C levels.

Tangier Disease (ABCA1 Deficiency)

Tangier disease is caused by loss-of-function mutations in both alleles encoding the gene *ABCA1* (Bodzioch et al., 1999; Brooks-Wilson et al., 1999; Rust et al., 1999). It is characterized by cholesterol accumulation in the reticuloendothelial system causing enlarged orange tonsils, hepatosplenomegaly, intestinal mucosal abnormalities, and peripheral neuropathy, as well as markedly low HDL-C (<5 mg/dl) and apoA-I levels (Hobbs and Rader, 1999). The lack of ABCA1 results in markedly impaired efflux of cholesterol and phospholipids from cells to lipid-free apoA-I. The lack of intestinal and hepatic ABCA1 is probably largely responsible for the low HDL due to impaired lipidation of newly secreted apoA-I by these organs. Poorly lipidated apoA-I is then extremely rapidly catabolized. Impaired cholesterol efflux from other tissues, particularly macrophages, results in cholesterol accumulation leading to many of the typical clinical characteristics of this disorder. However, Tangier disease patients do not develop rapidly accelerated atherosclerosis to the extent one might expect based on the cholesterol efflux defect and the extremely low HDL-C levels. Heterozygotes for ABCA1 mutations have reduced HDL-C levels that are intermediate between Tangier disease and normal but have no evidence of cholesterol accumulation in tissues. However, they are at some increased risk for premature CHD. Mutations in ABCA1 have been found to cause low HDL-C levels in some families in which Tangier disease homozygotes are not found (Brooks-Wilson et al., 1999; Marcil et al., 1999). Rare private mutations in the *ABCA1* gene may even be a cause of low HDL-C levels in the general population (Cohen et al., 2004). As a result of the discovery of the molecular basis of Tangier disease, ABCA1 is now a major target for the development of new therapies intended to upregulate ABCA1 expression (Linsel-Nitschke and Tall, 2005).

Lecithin:Cholesterol Acyltransferase (LCAT) Deficiency

LCAT deficiency is caused by loss-of-function mutations in both alleles of the *LCAT* gene (Kuivenhoven et al., 1997). LCAT is the enzyme that esterifies the free cholesterol present on HDL to cholesteryl ester, creating a cholesteryl ester core and resulting in maturation of HDL. In the absence of functional LCAT and cholesterol esterification, mature HDL particles are not formed and nascent HDL particles containing apoA-I and apoA-II are rapidly catabolized (Rader et al., 1994). Two genetic forms of LCAT deficiency have been described, complete deficiency known as classic LCAT deficiency and partial deficiency known as fish-eye disease (Kuivenhoven et al., 1997). In addition to extremely low HDL-C, both types of LCAT deficiency are characterized by corneal opacification, but only individuals with complete LCAT deficiency have low-grade hemolytic anemia and progressive chronic kidney disease leading to end-stage renal disease. Interestingly, neither form of LCAT deficiency is clearly associated with premature coronary disease despite the markedly reduced HDL-C levels (Kuivenhoven et al., 1997), raising questions about the importance of LCAT in protecting against CHD. LCAT heterozygotes have relatively normal HDL-C levels. Nevertheless, promotion of LCAT activity is of therapeutic interest as an HDL-raising approach.

Mendelian Disorders Primarily Causing Elevated HDL-C Levels

Cholesteryl Ester Transfer Protein (CETP) Deficiency

CETP deficiency is caused by loss-of-function mutations in both alleles of the *CETP* gene (Brown et al., 1989). CETP transfers cholesteryl esters from HDL to apoB-containing lipoproteins in exchange for triglycerides (Rader, 2006). Lack of functional CETP results in markedly elevated HDL-C levels due to lack of HDL remodeling, accumulation of cholesteryl esters in HDL, and slower turnover of apoA-I and apoA-II (Ikewaki et al., 1993). LDL-C levels are also low because of increased catabolism of LDL and apoB with endogenous upregulation of the LDL receptor (Ikewaki et al., 1995). CETP deficiency is extremely rare outside Japan; among the Japanese the most common mutations are a 5′ donor splice site intron 14 G to A substitution and a missense mutation in exon 15 (D442G) (Inazu et al., 1990). Heterozygous individuals for CETP deficiency have 60–70% of normal CETP activity and only a modest increase in HDL-C levels, and otherwise normal LDL-C levels. Whether homozygous or heterozygous CETP deficiency is associated with increased, decreased, or unchanged cardiovascular risk remains to be resolved (Rader, 2004). Nevertheless,

the identification of CETP deficiency led to the concept that CETP is a therapeutic target for inhibition as an HDL-raising strategy (Rader, 2006). The current status of CETP inhibition is unclear due to failure of the CETP inhibitor torcetrapib, which caused increased mortality in a phase III outcomes trial (Tall et al., 2007).

Common Gene Variants Segregating in Populations with HDL-C Levels

Common gene variants within at least nine loci (*ABCA1*, *ANGPTL4*, *APOA5*, *cluster*, *CETP*, *GALNT2*, *LIPC*, *LIPG*, *MVK/MMAB* and *LPL*) have been reproducibly found to be associated with HDL-C (Boekholdt et al., 2005; Frikke-Schmidt et al., 2004; Guerra et al., 1997; Kathiresan et al., 2008a; Lai et al., 2004; Rip et al., 2006; Willer et al., 2008). Common genetic variation in the *ABCA1* gene is clearly associated with variation in HDL-C levels in the general population and may even be associated with CHD risk (Brousseau et al., 2001; Clee et al., 2001). Common *CETP* polymorphisms are also associated with HDL-C levels, though the association with CHD risk is less clear (Boekholdt and Thompson, 2003). The most commonly studied SNPs in *CETP* are the Taq1B polymorphism (associated with increased HDL-C) and the I405V SNP (also associated with modestly increased HDL-C) (Boekholdt and Thompson, 2003). In Table 23.2, we present representative associations of SNPs in these genes with HDL-C from a single community-based cohort study in southern Sweden, the Malmo Diet and Cancer Study-Cardiovascular Cohort (Berglund et al., 1993). The variants associated with HDL-C vary in frequency from 7% at *APOA5* to 49% at *CETP* (Kathiresan et al., 2008b). These SNPs explain 0.2–2.5% of the interindividual variability in HDL-C in the population. In comparisons between the major and minor allele homozygote classes, the difference in HDL-C ranges from 3 mg/dl at *ABCA1* for to 7 mg/dl at *APOA5*. In addition to the five genes in Table 23.2, the nonsynonymous variant in *ANGPTL4* (E40K) has been strongly related to HDL-C (Romeo et al., 2007).

GENETICS OF TRIGLYCERIDES

Studies of the genetic basis of substantially elevated triglyceride levels have been useful in understanding the regulation of triglyceride metabolism. We first review the Mendelian disorders of high triglycerides (Table 23.1) and then review the current state of knowledge regarding common polymorphisms and their association with triglyceride levels.

Mendelian Disorders Influencing Triglyceride Levels

Familial Chylomicronemia Syndrome (LPL Deficiency and ApoC-II Deficiency)

The familial hyperchylomicronemia syndrome (FCS) is caused by homozygosity for loss-of-function mutations in one of two genes, LPL and apoC-II (Santamarina-Fojo, 1992). These conditions are virtual phenocopies and are therefore discussed together. FCS is characterized by extreme hypertriglyceridemia (greater than 1000 mg/dl) usually presenting in childhood with acute pancreatitis, eruptive xanthomas, lipemia retinalis, and/or hepatosplenomegaly.

Chylomicron triglycerides are hydrolyzed in muscle and adipose capillary beds by LPL with apoC-II acting as a required cofactor, thus loss of function of either protein produces the phenotype of hyperchylomicronemia. Interestingly, despite the markedly elevated triglyceride (and cholesterol) levels, premature atherosclerotic cardiovascular disease is not generally a feature of this disease. This observation contributed to our understanding of the importance of the nature of lipoprotein, not just the lipid, in determining cardiovascular risk. Primary therapy is restriction of total dietary fat.

ApoAV Deficiency

Genetic deficiency of apoAV due to loss-of-function mutations on both alleles causes late-onset hyperchylomicronemia (Priore Oliva et al., 2005). In addition, heterozygosity for mutations such as Q139X can lead to the same phenotype due to a dominant negative effect (Marcais et al., 2005). ApoAV is a minor apolipoprotein that promotes LPL-mediated hydrolysis of lipoprotein triglycerides. Subjects with apoAV deficiency have a severe lipolysis defect and markedly reduced VLDL catabolism.

Familial Dysbetalipoproteinemia (Type III Hyperlipoproteinemia)

Familial dysbetalipoproteinemia (FD), or type III hyperlipoproteinemia, is caused by mutations in the gene for apolipoprotein E (apoE) (Mahley et al., 1999). ApoE on chylomicron and VLDL remnants normally mediates their catabolism by binding to receptors in the liver. FD is usually caused by homozygosity for a common variant called apoE2, which differs from the wild-type apoE3 form by a substitution of a cysteine for an arginine at position 158. ApoE2 has impaired binding to lipoprotein receptors such as the LDL receptor, resulting in defective removal of chylomicron and VLDL remnants. About 0.5% of individuals are homozygous for apoE2 but the prevalence of FD is only about 1 in 10,000, indicating that other genetic or environmental factors are required for expression of the phenotype. Because remnant lipoproteins are elevated and contain both triglycerides and cholesterol, plasma levels of both triglycerides and cholesterol are elevated. Palmar xanthomata and tuberoeruptive xanthomata on the elbows, knees, or buttocks are distinctive skin findings in this condition. Importantly, premature atherosclerotic CVD is common in this disorder, an observation that helped to clarify that remnant lipoproteins are highly atherogenic. This disorder is important from a clinical genomics sense, because it is one of the few genetic lipid disorders in which genotyping is clinically indicated for diagnosis; the finding of the *apoE2/E2* genotype is diagnostic in the appropriate clinical setting.

Hepatic Lipase Deficiency

Hepatic lipase (HL) deficiency is caused by loss-of-function mutations in both alleles of the *LIPC* gene (Hegele et al., 1993). HL deficiency is characterized by elevated plasma levels of cholesterol and triglycerides due to the accumulation of circulating lipoprotein remnants as a result of lack of HL activity. HL deficiency is very rare and therefore it is difficult to determine its true relationship to atherosclerotic CVD.

Common Gene Variants Segregating in Populations with Triglyceride Levels

Common gene variants within at least ten loci, *ANGPTL3*, *ANGPTL4*, *APOB*, *APOA5*, *CILP2/PBX4*, *GALNT2*, *GCKR*, *LPL*, *MLXIPL*, and *TRIB1*, have been reproducibly related to triglyceride levels (Benn et al., 2005; Kathiresan et al., 2008a; Kooner et al., 2008; Pennacchio et al., 2001; Rip et al., 2006; Saxena et al., 2007; Willer et al., 2008). In Table 23.2, we present representative associations of SNPs in these genes with triglycerides from a single community-based cohort study in southern Sweden, the Malmo Diet and Cancer Study-Cardiovascular Cohort (Berglund et al., 1993). These variants vary in frequency from 7% at *APOA5* to 48% at *APOB*. These SNPs explain 0.3–0.7% of the interindividual variability in triglyceride in the population (Kathiresan et al., 2008b). In comparisons between the major and minor allele homozygote classes, the difference in triglycerides ranges from 9 mg/dl at *APOB* to 39 mg/dl at *APOA5*. In addition to these variants, the same 3% nonsynonymous variant in *ANGPTL4* noted above with regard to HDL-C has been related to triglyceride level (Romeo et al., 2007).

GENETIC LIPID DISORDERS WITHOUT CURRENT PROVEN MOLECULAR ETIOLOGY

Familial Combined Hyperlipidemia (FCHL)

Familial combined hyperlipidemia (FCHL) is a dominantly inherited condition characterized by elevated triglycerides, elevated cholesterol, and reduced HDL-C (Grundy et al., 1987). Premature ASCVD is common in patients with FCHL. The metabolic basis of FCHL is thought to be hepatic overproduction of VLDL. It is the most common inherited lipid disorder, with a prevalence of approximately 1 in 200 persons. The genetic basis of FCHL is not well understood, and probably several different genes can cause a similar phenotype (Shoulders et al., 2004; Suviolahti et al., 2006). Genome-wide linkage studies have demonstrated a repeated linkage of FCHL to a locus on chromosome 1q21-q23 (Coon et al., 2000; Pajukanta et al., 1998), and association with SNPs in the gene encoding the upstream transcription factor 1 (USF1) has been repeatedly shown to be associated with the FCHL phenotype (Coon et al., 2005; Pajukanta et al., 2004). Because USF1 is a transcription factor known to regulate the expression of genes involved in lipid metabolism (Lee et al., 2006), it is plausible, though unproven, that genetic variation in USF1 expression or function could cause the phenotype of FCHL. However, the mechanisms of this association are unknown, and other genes likely contribute to the phenotype of FCHL. The molecular basis of FCHL is one of the most important questions in the field of the genomics of lipid disorders, and additional discoveries in this area are likely to provide important insights into the pathophysiology of hepatic VLDL overproduction as well as new targets for the development of new therapies for lipid disorders.

Familial Hypertriglyceridemia (FHTG)

Familial hypertriglyceridemia (FHTG) is an autosomal dominant trait characterized by elevated triglycerides with normal or only modestly increased total cholesterol levels, and LDL-C levels are usually normal. In contrast to FCHL, FHTG is often not associated with increased risk of ASCVD. Both VLDL overproduction and reduced VLDL catabolism have been implicated in causing this phenotype, but the pathophysiology is not well understood. The molecular basis of FHTG has not yet been elucidated.

Familial Hypoalphalipoproteinemia

Familial hypoalphalipoproteinemia is a dominantly inherited trait characterized by low HDL-C levels, usually in the setting of relatively normal TG levels. Premature CVD is often, but not always, associated with this condition. Accelerated catabolism of apoA-I is thought to be the metabolic basis of this trait (Lewis and Rader, 2005). ABCA1 mutations can cause this phenotype, but this phenotype occurs in families without ABCA1 mutations and is likely caused by mutations in more than one other gene.

Familial Hyperalphalipoproteinemia

Familial hyperalphalipoproteinemia is a dominantly inherited trait characterized by high HDL-C levels. This trait is usually, but not always, associated with lower risk for CHD and increased longevity. Interestingly, while homozygous CETP deficiency is a cause of hyperalphalipoproteinemia as discussed above, obligate heterozygotes for CETP generally do not have exceptionally elevated HDL-C levels. The molecular basis of this condition is not known, but the discovery of genes that cause this phenotype would increase our understanding of the regulation of HDL metabolism and potentially provide new targets for the development of therapeutics for raising HDL.

INFLUENCE OF LIPID-MODULATING MUTATIONS ON RISK OF ATHEROSCLEROTIC CARDIOVASCULAR DISEASE

Mendelian Syndromes and ASCVD

With regard to LDL, the recognition of FH was instrumental in proving that elevated LDL-C levels cause accelerated ASCVD

without a requirement for additional traditional risk factors. This observation strongly supported the concept that reduction in LDL-C would reduce risk of ASCVD, a hypothesis subsequently proven with large CV outcome trials. Other Mendelian syndromes of elevated LDL-C, such as FDB, ARH, and ADH, though less common, have also been associated with early onset ASCVD, consistent with this concept. Mendelian syndromes of low LDL are generally not common enough to allow definitive demonstration of protection against ASCVD (much more difficult to prove than accelerated disease), but studies of patients with hypobetalipoproteinemia due to apoB mutations are consistent with the concept that the risk of CVD is reduced (Schonfeld et al., 2005). Overall, the study of Mendelian disorders of high and low LDL has had a major impact on the understanding of the role of LDL as a risk factor for atherogenesis.

Mendelian syndromes of HDL are generally not prevalent enough to allow definitive proof of a relationship to CVD. However, inability to synthesize apoA-I, while rare, has been consistently associated with accelerated ASCVD (Norum et al., 1982; Schaefer et al., 1982). In contrast, Tangier disease, while associated with similarly extremely low HDL-C levels, has not been associated with markedly increased ASCVD (Schaefer et al., 1980). In a similar fashion, LCAT deficiency is also associated with extremely low HDL but is also not associated with increased CVD (Kuivenhoven et al., 1997). Conversely, while CETP deficiency is associated with extremely elevated HDL-C levels, its relationship to CVD still remains uncertain.

Finally, the sole Mendelian disorders associated with hyperchylomicronemia, LPL deficiency and apoC-II deficiency, have not been convincingly associated with increased cardiovascular risk, confirming the relative lack of atherogenicity of chylomicrons. Conversely, FD (type III hyperlipoproteinemia) is also associated with elevated triglycerides but in this case due to elevated remnant lipoproteins, and causes substantially increased risk of ASCVD. These Mendelian disorders of elevated triglycerides demonstrate that it is the nature of the lipoprotein that is elevated, not the plasma level of triglycerides *per se*, that determines the risk of atherosclerosis.

Common Mutations in Aggregate Affecting CVD Risk

As common SNPs influence blood lipid levels and blood lipids affect risk for cardiovascular disease, a natural question is whether lipid-modulating SNPs influence risk for cardiovascular disease. One of the best examples illustrating this issue is that of PCSK9, with relatively common variants Y142X or C679X causing reduced LDL-C levels and substantially reduced risk of CHD (Cohen et al., 2006). However, for many other common variants that influence LDL-C (e.g., at the *APOB* and *APOE* loci), the connection to cardiovascular disease has less compelling evidence. Likewise, common SNPs in CETP and HL that reproducibly influence HDL-C levels have not been definitively associated with cardiovascular outcomes; the same is true for common SNPs that influence plasma triglyceride levels. A key limiting factor may be that each common SNP only modestly affects the lipid level. This raises the hypothesis that a combination of lipid-modulating SNPs might contribute to risk for cardiovascular disease. With the recent establishment of several reproducible associations for each of the three lipid traits, it is now feasible to test this hypothesis in prospective studies. The identification of a panel of lipid polymorphisms that influence CVD risk could help target preventive therapies or aid in risk prediction at the population level.

FUTURE DIRECTIONS IN GENETICS AND GENOMICS OF LIPOPROTEINS

Over the next few years, the widespread application of the genome-wide association study (GWAS) method is expected to lead to substantial progress in defining the inherited basis for blood lipids. The GWAS is defined as an experiment in which a substantial set of common SNPs across the genome (~300,000–500,000 SNPs) are simultaneously tested for association with disease or quantitative risk factors for disease (Hirschhorn and Daly, 2005). Blood lipid levels and ASCVD are among the phenotypes being widely studied by this approach and the early results are promising. Table 23.3 summarizes the design features of several GWASs recently completed or in progress for lipid or ASCVD phenotypes.

For blood lipids, an early example is the Diabetes Genetics Initiative of Broad, Lund And Novartis. This study was designed to examine type 2 diabetes but 18 other phenotypes including blood lipids were analyzed as secondary traits (Saxena et al., 2007). Patients with type 2 diabetes and controls were genotyped for ~500,000 SNPs across the genome and these SNPs were tested for association with blood LDL-C, HDL-C, triglycerides, apoB, apoA-I, and apoA-II. Several findings from this study highlight key issues in defining the genetic basis for lipid levels. First, the largest effect size for a common SNP approached ~2% of trait variance explained (e.g., an *APOE* cluster SNP at 2.1% of LDL-C variance and a *CETP* SNP at 2.1% for HDL-C variance). As most other common variants will each individually confer a more modest effect size (<1% of trait variance explained), large sample sizes will be needed to convincingly demonstrate that a gene variant influences lipid levels. Second, several loci with common alleles affecting lipid levels (*APOE*, *ABCA1*, *APOA5*, *CETP*, *LIPC*, and *LPL*) have also been shown to cause Mendelian syndromes or contain multiple rare alleles. Thus, sequencing all loci with validated common alleles will be required to define the spectrum of common and rare alleles at each locus and the full impact of each locus on trait variation. Third, loci previously unsuspected to play a role in lipid phenotypes may be identified using an unbiased discovery approach such as GWAS. For example, a highly significant association with triglycerides was observed for rs780094 ($P = 3.7 \times 10^{-8}$), explaining 1% of residual variance in triglyceride levels. This single SNP was then tested in 5217 individuals from the Malmö

TABLE 23.3 Selected genome-wide association studies for atherosclerotic cardiovascular disease and/or lipid phenotypes

Study	Ascertainment scheme	Ancestry	N	Phenotypes
Diabetes Genetics Initiative (Saxena et al., 2007)	Cases with type 2 diabetes, controls free of diabetes from Sweden and Finland	European	2931	LDL-C, HDL-C, triglycerides, apolipoproteins
FUSION (Scott et al., 2007)	Cases with type 2 diabetes, controls free of diabetes from Finland	European	2457	LDL-C, HDL-C, triglycerides
SardiNIA (Scuteri et al., 2007)	Family-based sample from 4 towns in Sardinia, Italy	European	4305	LDL-C, HDL-C, triglycerides
Framingham Heart Study (Splansky et al., 2007)	Community-based prospective cohort study from single town in Framingham, USA	European	~9000	LDL-C, HDL-C, triglycerides, apolipoproteins, lipoprotein subfractions, others
Atherosclerosis Risk in Communities (ARIC, 1989)	Community-based prospective cohort study, four communities in United States	European, African-American	~16,000	LDL-C, HDL-C, triglycerides, apolipoproteins, others
Kosrae (Bonnen et al., 2006)	Founder population on island in South Pacific	Asian	~3000	LDL-C, HDL-C, triglycerides
Jackson Heart Study (Wilson et al., 2005)	Community-based prospective cohort study from single town in United States	African-American	~4000	LDL-C, HDL-C, triglycerides
McPherson et al. (2007)	Hospital-based case collection; incident cases from two community-based cohort studies, and prevalent coronary artery calcium in a community-based cohort study	European	4277 cases, 20,054 controls	Coronary artery bypass grafting, angioplasty, myocardial infarction, coronary artery calcium
DeCode (Helgadottir et al., 2007)	Population-based registry from Iceland	European	4589 cases, 12,768 controls	Myocardial infarction
WTCCC (WTCCC, 2007; Samani et al., 2007)	UK-wide ascertainment of cases aged <66 with at least one affected sibling	European	1988 cases, 3004 controls	Myocardial infarction, coronary artery bypass surgery or angioplasty
German MI Family Study (Samani et al., 2007)	Premature MI and at least one first-degree relative with premature CAD; cases from cardiac rehabilitation programs	European	875 cases, 1644 controls	Myocardial infarction
Myocardial Infarction Genetics Consortium (MIGen)	Premature MI; 3 hospital-based case collections, a community-based ascertainment of cases, and cases from two prospective cohort studies	European	~3300 cases, ~3300 controls	Early-onset myocardial infarction
PennCath/Medstar	Cardiac cath-lab based collection of premature angiographic CAD with or without history of MI; controls older and free of CAD by angiography	European	~1000 cases with MI, ~1000 cases without MI, ~1000 controls	Obstructive coronary atherosclerosis on angiography, MI, LDL-C, HDL-C, triglycerides, apolipoproteins, others

Diet and Cancer Study–Cardiovascular Cohort and the association was replicated ($P = 8.7 \times 10^{-8}$). This SNP resides within the gene glucokinase regulatory protein (*GCKR*), which regulates glucokinase (GCK), the first enzyme in the glycolytic pathway, and has the potential to influence hepatic TG synthesis.

For ASCVD, three separate studies have identified a locus on chromosome 9p21 as reproducibly and highly significantly associated with myocardial infarction (MI) or coronary artery disease (Helgadottir et al., 2007; McPherson et al., 2007; Samani et al., 2007). The risk allele at 9p21 is common at ~47% frequency in

controls, and each copy of the risk allele increases risk by ~25%. The associated SNPs exist in a 190-kilobase region of strong linkage disequilibrium which contains two cyclin-dependent kinase inhibitors – CDKN2A and CDKN2B. Though the association is robust, the mechanism behind which the associated SNPs confer risk for MI or coronary artery disease remains to be defined. In addition to the 9p21 variant, the study by Samani et al. (2007) highlighted six other loci with either convincing or suggestive evidence for association. From the three initial GWASs for ASCVD, it is clear that the effect sizes of common variants will be less than an odds ratio of 1.4 per copy of the risk allele. Several of the newly discovered loci, including 9p21, do not seem to be acting through established risk factors such as lipids. As additional studies are performed, the number of loci and alleles affecting ASCVD should become clear.

PHARMACOGENETICS OF LIPID-MODULATING THERAPIES

Drug therapy to modulate lipids with the goal of preventing or treating atherosclerotic cardiovascular disease is well established clinically and commonly used in practice. The three major classes of drugs used to reduce LDL-C are HMG CoA reductase inhibitors (statins), cholesterol absorption inhibitors (currently ezetimibe is the only member of this class), and bile acid sequestrants. The three major classes of drugs used to treat the HDL-TG axis are nicotinic acid (niacin), PPARα agonists (fibrates), and omega 3 fatty acids ("fish oils"). Interindividual variability in response to all of these classes is substantial and undoubtedly influenced by genetic factors. A limited number of studies have investigated the genetic variation underlying response to lipid-modifying drugs.

Statins inhibit 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase, the rate-limiting step in cholesterol biosynthesis, and in doing so alter cholesterol metabolism in the hepatocyte, resulting in compensatory upregulation of the LDL receptor, leading to increased LDL catabolism. A few studies have investigated candidate genes with regard to variability in LDL-lowering response to statins. In a study involving the treatment of 1536 healthy persons with moderate hypercholesterolemia with pravastatin 40 mg for 24 weeks, 148 SNPs in 10 candidate genes were assessed with regard to LDL-C reduction. Two common SNPs in strong linkage disequilibrium, rs17244841 and rs17238540, in the gene encoding HMG-CoA reductase were significantly associated with reduced LDL-C reduction on pravastatin (Chasman et al., 2004). In another study in 1360 individuals on atorvastatin therapy, 43 SNPs in 16 candidate genes were assessed with regard to LDL-C reduction. A significant association with HMG-CoA reductase was not found, but a modest significant association with the *apoE2* polymorphism was noted, whereby heterozygotes had 3.5% greater LDL-C reduction (Thompson et al., 2005). Thus the majority of the variability in statin response remains unexplained by the existing candidate gene studies.

Ezetimibe inhibits cholesterol absorption in the small intestine by binding to the enterocyte brush border cholesterol transporter NPC1L1. This reduces cholesterol transport from intestine to liver and results in secondary compensatory upregulation of the LDL receptor, leading to increased LDL catabolism. Hegele and colleagues identified a patient who exhibited extremely poor clinical response to ezetimibe with regard to LDL lowering and sequenced the *NPC1L1* gene, finding two rare nonsynonymous polymorphisms in NPC1L1, V55I and I1233N, that were absent in control subjects and suggesting that compound heterozygosity for these mutations may have caused the lack of response to ezetimibe (Wang et al., 2005). In a follow-up study, 101 dyslipidemic subjects were treated with ezetimibe for 12 weeks and common *NPC1L1* SNPs and haplotype were assessed for association with LDL-C reduction. Subjects treated with ezetimibe lacking the common NPC1L1 haplotype 1735C-25342A-27677T had significantly greater reduction in LDL-C (36%) than subjects with at least one copy of this haplotype (24%) (Hegele et al., 2005). Thus genetic variation in NPC1L1 appears to influence the response to ezetimibe treatment.

Nicotinic acid activates the G protein-coupled receptor GPR109A on adipocytes, resulting in reduced adipocyte triglyceride lipolysis and FFA release and reduced flux of FFA to the liver (Offermanns, 2006). This is believed to be the primary mechanism of TG-lowering of nicotinic acid; the mechanism of HDL-raising remains uncertain. While polymorphisms in GPR109A have been described (Zellner et al., 2005), no studies of the pharmacogenetics of niacin response have been reported to date. Fibrates are PPARα agonists that activate the transcription of a variety of lipid metabolism genes in multiple tissues, including muscle, adipose, and liver. The triglyceride lowering and modest HDL-raising mechanisms of fibrates are due to regulation of several genes, including LPL, apoA-I, and apoC-III. Only a few pharmacogenetic studies of fibrate lipid responses have been reported to date. In 292 hypertriglyceridemic subjects treated with fenofibrate for 3 months, the TG-lowering response to fenofibrate was reduced in carriers of the P207L mutation in LPL and greater in carriers of the apoE2 polymorphism and the L162V polymorphism in PPARα (Brisson et al., 2002). In 791 hypertriglyceridemic subjects treated with fenofibrate for 3 weeks, carriers of the C56G tagSNP in *APOA5* had a greater decrease in TG and increase in HDL-C (Lai et al., 2007). Interestingly, carriers of C56G also had a greater response to fenofibrate after a defined fat load. Thus, genetic variation clearly plays a role in determining fibrate response, though additional work is required.

IMPLICATIONS OF GENOMICS OF LIPOPROTEIN METABOLISM FOR THE DEVELOPMENT OF NOVEL THERAPIES

The study of the genetics and genomics of lipoprotein metabolism has led to several novel therapeutic targets. The discovery of the molecular basis of FH being mutations in the LDL receptor

led to the concept that upregulation of the LDL receptor could be a strategy for reducing plasma levels of LDL-C. While statins were developed based on their ability to inhibit HMG-CoA reductase, the rate-limiting step in cholesterol biosynthesis, their physiological effect in the liver is upregulation of the LDL receptor with consequent reduction in LDL-C levels. Despite the success of statins and other LDL-lowering therapies such as cholesterol absorption inhibitors, there remains a need for additional LDL-lowering therapies. The fact that mutations that impair the biosynthesis of apoB cause reduced LDL-C levels supports the concept of targeting apoB therapeutically. While small molecule approaches to apoB inhibition are impractical, an approach using an antisense oligonucleotides (ASO) has been shown to be effective in animal models as well as in humans. The efficacy of an ASO targeted to apoB (apoB-ASO) was demonstrated in hypercholesterolemic mice (Crooke et al., 2005). In a phase I multiple-dose study in subjects with mild dyslipidemia, subcutaneous injection of an apoB-ASO reduced plasma apoB by up to 50% and LDL-C by up to 44% (Kastelein et al., 2006). Alternatively, small-interfering RNA (siRNA) molecules complementary to apoB mRNA also significantly reduced the plasma concentration of apoB-containing lipoproteins in nonhuman primates (Zimmermann et al., 2006). Thus inhibition of apoB expression using systemic administration of either an ASO or siRNA to apoB reduces LDL-C levels comparable to that seen in subjects heterozygous for familial hypobetalipoproteinemia.

As discussed above, another Mendelian disorder associated with low LDL-C levels is abetalipoproteinemia. The discovery that genetic deficiency of MTP causes abetalipoproteinemia led to the concept that MTP inhibitors would disrupt VLDL assembly and secretion, with consequent reductions in circulating LDL (Burnett and Watts, 2007). Indeed, studies in mouse models have demonstrated that small molecule inhibitors of MTP effectively reduce plasma levels of total and LDL-C (Liao et al., 2003; Spann et al., 2006). Furthermore, an MTP inhibitor was shown to be effective in substantially reducing cholesterol in the Watanabe heritable hyperlipidemic (WHHL) rabbit, a model of homozygous FH (Wetterau et al., 1998). Finally, a small clinical trial in six subjects with homozygous FH definitively demonstrated that MTP inhibition reduces LDL-C levels in humans by reducing hepatic apoB production (Cuchel et al., 2007). After titrating the dose over 4 months, LDL-C was reduced by 51% and apoB by 55% at the highest dose of the MTP inhibitor. ApoB kinetic studies demonstrated a marked reduction in LDL apoB production as the basis for the reduction in LDL-C and apoB. This trial also demonstrated the mechanism-based increase in hepatic steatosis associated with MTP inhibition, which occurred in four of the six subjects. While this clearly requires more study, including longer-term trials, it is possible that high-dose MTP inhibition could be developed as an orphan drug approach for homozygous FH patients and that lower doses of MTP inhibition could potentially have a role in treating patients who are unable to reach LDL-C goal on present therapies.

The story of PCSK9 provides the best example in the lipoprotein field of how human genetics can identify novel therapeutic targets. As discussed above, it was first discovered that rare nonsynonymous variants of PCSK9 can cause a form of ADH (Abifadel et al., 2003). After studies of overexpression of PCSK9 in mice unexpectedly caused elevated LDL-C levels (Benjannet et al., 2004; Maxwell and Breslow, 2004), loss-of-function mutations in PCSK9 were shown to result in substantially reduced LDL-C levels (Cohen et al., 2005). Importantly, individuals heterozygous for loss-of-function mutations in PCSK9 were subsequently shown to have markedly reduced lifetime risk of CHD (Cohen et al., 2006; Kathiresan et al., 2008c). A compound heterozygote with loss-of-function mutations in both *PCSK9* alleles has been reported to have a very low LDL-C but is in good health (Zhao et al., 2006). Thus, inhibition of PCSK9 is an extremely attractive target based on the human genetics. Because PCSK9 is upregulated by statin therapy, the addition of a PCSK9 inhibitor to a statin could result in additive or even synergistic reduction of LDL-C.

Investigation of the human genetics of HDL has provided important information regarding potential therapeutic targets. The fact that LCAT deficiency is a cause of severely low HDL supported the concept that LCAT upregulation or stimulation would be a strategy for raising HDL. More importantly, the discovery of *ABCA1* as the mutated gene in Tangier disease created immediate interest in ABCA1 as a therapeutic target for upregulation as a strategy to raise HDL-C levels and inhibit or regress atherosclerosis. Perhaps the best example of human genetics identifying a potential therapeutic target for HDL is that of CETP deficiency. As noted above, CETP-deficient subjects have markedly elevated HDL levels, directly identifying CETP as a target for inhibition as an approach to raising HDL. While the first CETP inhibitor to enter phase III, torcetrapib, failed to reduce cardiovascular events, it has the off-target property of raising blood pressure. Thus the status of CETP inhibition as a therapeutic approach remains uncertain. Other molecular causes of high HDL could provide additional new targets for therapeutic development.

CLINICAL RECOMMENDATIONS FOR GENETIC TESTING FOR LIPID DISORDERS

The majority of lipid disorders are diagnosed based on the clinical presentation, and genetic testing is not usually clinically indicated. For example, while heterozygous FH has a prevalence of 1 in 500 individuals, the majority of LDL-receptor mutations are (with the exception of a few regions in which a founder effect is present) private and could not be screened for short of sequencing. Furthermore, knowledge of the specific mutation, or even that the gene defect is in the LDL receptor and not, for example, in apoB, has no direct impact on the clinical care of the patient. Therefore, in FH genetic screening is not recommended. One exception to this might be in prenatal diagnosis of homozygous

FH. If both parents have heterozygous FH, the chance of a homozygous child is one in four. Knowledge of the specific mutations in each parent (which would require sequencing) could permit prenatal testing for these mutations. This same rationale and approach could be applied to other serious lipoprotein disorders, for example, abetalipoproteinemia in which knowledge of the specific mutation(s) through sequencing could be used for prenatal diagnosis.

The one major exception in which genetic testing is clinically recommended for diagnosis of lipid disorders is in the definitive diagnosis of FD (also called type III hyperlipoproteinemia). If this diagnosis is suspected on clinical grounds, it can be confirmed by performing apoE genotyping (with a focus on the apoE2 polymorphism). The finding of homozygosity for apoE2/E2 in the appropriate clinical setting is diagnostic for FD, and may help with clinical management decisions and family counseling. Interestingly, clinical apoE genotyping generally includes typing for not only the apoE2 but also the apoE4 polymorphism. This is not associated with FD, but is associated with Alzheimer's disease, raising the issue of providing unwanted genetic prognostic information that has no immediate clinical utility and could impact on issues such as insurability if recorded in the medical record. Thus, while clinically appropriate, apoE genotyping should be performed only after careful consideration, and ideally should be limited to typing for the apoE2 polymorphism when being done for diagnosis of FD.

2009 UPDATE

Progress continues to be made in applying the GWAS approach to the discovery of novel loci associated with variation in plasma lipid levels. Kathiresan et al. (2009) utilized approximately 40,000 persons assembled from cohorts, case-control studies, and clinical trials to identify 30 distinct loci associated with lipoprotein concentrations (at a genome-wide significant p value of $<5 \times 10^{-8}$), including 11 loci that had not previously been identified as genome-wide significant. The 11 newly defined loci included *ABCG8*, *MAFB*, *HNF1A*, and *TIMD4* associated with LDL cholesterol; *ANGPTL4*, *FADS1-FADS2-FADS3*, *HNF4A*, *LCAT*, *PLTP*, and *TTC39B* associated with HDL cholesterol; and *AMAC1L2*, *FADS1-FADS2-FADS3*, and *PLTP* associated with triglycerides. *ABCG8* and *LCAT* have been shown to cause Mendelian forms of dyslipidemia (sitosterolemia and LCAT deficiency, respectively), further strengthening the connection between loci for mendelian dyslipidemic syndromes and those with common variants of modest effect. Indeed, at least 11 of the 30 loci (*ABCG8*, *LCAT*, *APOB*, *APOE*, *LDLR*, *PCSK9*, *CETP*, *LPL*, *LIPC*, *APOA5*, and *ABCA1*) not only harbor common variants associated with variation in plasma lipids but also low-frequency variants and/or rare mutations that cause mendelian syndromes. Finally, the use of an allelic dosage score suggested that the cumulative effect of multiple common variants contributes to polygenic dyslipidemia.

Aulchenko et al. (2009) studied over 20,000 persons from 16 different population-based cohorts and identified 22 distinct loci associated with plasma lipids including six new loci: *ABCG5*, *TMEM57*, *DNAH11*, and *FADS3-FADS2* associated with total and/or LDL cholesterol and the *CTCF-PRMT8* region and *MADD-FOLH1* region associated with HDL. Sex-specific differences in effect size were demonstrated for *HMGCR*, *NCAN*, and *LPL*. Allelic dosage scores were associated with carotid intima media thickness and CHD incidence. The genetic risk score improves the screening of high-risk groups of dyslipidemia over classical risk factors.

Sabatti et al. (2009) used the Northern Finland Birth Cohort 1966 for GWAS of lipid and other metabolic traits and identified three new lipid loci: *AR* and *FADS1-FADS2* with LDL cholesterol and *NR1H3* (*LXRA*) with HDL cholesterol. A low-frequency (MAF = 0.017) variant in the AR was found to be associated with markedly increased LDL in males, associated with a mean LDL elevation of 28 mg/dl.

Low-frequency variants (0.5–5% allele frequency) are likely an important source of variation in plasma lipid levels but are rarely captured by genome-wide genotyping platforms. Pollin et al. (2008) found through a GWAS that about 5% of the Amish in Lancaster County are heterozygous for a null mutation (R19X) in the gene encoding apoC-III (*APOC3*), resulting in reduced plasma apoC-III and lower triglyceride and LDL-C levels and higher HDL-C levels. Furthermore, individuals with R19X had reduced coronary artery calcification compared with noncarriers, consistent with a cardioprotective effect of this rare variant.

Medical resequencing studies will be required to identify the majority of the low frequency variants and all of the rare variants (<0.5% allele frequency) that have a quantitatively large effect size on plasma lipids. As one example of this approach, Romeo et al. (2007) resequenced the coding region and proximal intronic regions of *ANGPTL4* in a multiethnic sample of 3551 individuals and found an excess of nonsynonymous variants in individuals with TG levels in the lowest quartile. One variant, E40K, was present in 3% of Caucasians and was associated with significantly lower plasma levels of TG and LDL cholesterol (LDL-C) and higher levels of HDL cholesterol (HDL-C) in two other large cohorts. More recently, Romeo et al. (2009) reported the resequencing of coding regions in the genes encoding the angiopoietin-like proteins (ANGPTL3, -5, and -6) and identified multiple rare nonsynonymous variants in *ANGPTL3* and *ANGPTL5* associated with low plasma TG levels. These studies have established that *ANGPTL3*, *ANGPTL4*, and *ANGPTL5*, but not *ANGPTL6*, play nonredundant roles in TG metabolism.

Resequencing of individuals ascertained on the basis of extreme lipid traits, rather than those from the tails of a normal distribution, may provide a more efficient method of identifying rare mutations and low-frequency variants with

large effect sizes on lipid phenotypes. Edmondson et al. (2009) resequenced the *LIPG* gene encoding the enzyme endothelial lipase in individuals with extremely high HDL-C levels and controls with relatively low HDL and found a significant excess of rare variants in the high HDL group; analysis of variants created by site-directed mutagenesis proved them to be loss-of-function mutations. In addition, a low-frequency variant N396S was found to be highly associated with elevated HDL-C in multiple cohorts and within families and was also proven to have reduced lipolytic activity.

One of the critical questions regarding loci that influence plasma lipids is whether they also influence risk of CHD. This question can be addressed using mendelian randomization. The key concept is that if a locus altering plasma lipid level is causal for CHD, then gene variants at that locus would also be associated with CHD in a manner consistent with the effect on lipids. Willer et al. (2008) tested whether SNPs associated with LDL cholesterol, HDL cholesterol, or triglycerides also related to risk for CHD. Interestingly, they observed that nearly all variants related to LDL cholesterol were associated with CHD but that this was not the case for variants related to HDL cholesterol or triglycerides. This observation raises the provocative hypothesis that only certain genetic mechanisms that alter plasma HDL cholesterol or triglycerides affect risk for CHD.

Testing this hypothesis for the *CETP* locus is of particular interest as drug development programs targeting this protein are ongoing. Thompson et al. (2008) conducted a meta-analysis of studies reporting on the relationship between *CETP* gene variants, HDL cholesterol, and risk for CHD. They observed that *CETP* variants that increase HDL cholesterol modestly reduced the risk for CHD. Ridker et al. (2009) reported similar findings in a prospective cohort study. These studies are important, as they suggest that the HDL-raising effect of reduced CETP expression may reduce cardiovascular risk, albeit modestly. Even larger studies are needed to definitively evaluate whether genetic loci which affect HDL cholesterol or triglycerides in turn alter risk for CHD.

The explosion of GWAS for lipid traits and related phenotypes (such as response to lipid-lowering therapy) has the potential to impact on the evolution of "personalized medicine." For example, genetic risk score based on the counting of the number of "adverse" alleles influencing lipids may enhance risk prediction compared with measurement of lipids alone (Kathiresan et al., 2008). Furthermore, genetic information could influence treatment decisions. For example, a GWAS study of statin myopathy performed in the SEARCH study of two different doses of simvastatin revealed a highly significant association with a locus encoding the organic ion transporter SLCO1B1 responsible for transporting simvastatin into the liver (Link et al., 2008). Carriers of the rare allele were at significantly increased risk of statin myopathy and also had a reduced response with regard to LDL lowering. In concept, genotyping for this allele could influence the choice of type or dose of statin to reduce the risk of myopathy. The coming year will see a large number of studies addressing the genetic determinants of the response to lipid-altering therapy.

Finally, several GWASs focused directly on atherosclerotic cardiovascular disease phenotypes have been reported and in sum, they have identified ~14 loci for MI or CHD (at genome-wide levels of significance) (Erdmann et al., 2009; Gudbjartsson et al., 2009; Kathiresan et al., 2009; Ozaki et al., 2009; Tregouet et al., 2009). For example, in the largest of these studies, the Myocardial Infarction Genetics Consortium found nine loci associated with early-onset MI with genome-wide significance:21q22 near *MRPS6-SLC5A3-KCNE2*, 6p24 in *PHACTR1*, 2q33 in *WDR12*, 9p21, 1p13 near *CELSR2-PSRC1-SORT1*, 10q11 near *CXCL12*, 1q41 in *MIA3*, 19p13 near *LDLR*, and 1p32 near *PCSK9*. Of note, five of the 14 loci defined in recent studies have been previously related to LDL cholesterol; these include *LDLR*, *PCSK9*, 1p13 near *CELSR2-PSRC1-SORT1*, *TCF1*, and *LPA*. These results further strengthen the now unequivocal causal relationship between LDL cholesterol and CHD.

2009 UPDATE REFERENCES

Aulchenko, Y.S., Ripatti, S., Lindqvist, I., Boomsma, D., Heid, I.M., Pramstaller, P.P., Penninx, B.W.J.H., Janssens, A.C.J.W., Wilson, J.F., Spector, T. et al. (2009). Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet* 41, 47–55.

Edmondson, A.C., Brown, R.J., Kathiresan, S., Cupples, L.A., Demissie, S., Manning, A.K., Jensen, M.K., Rimm, E.B., Wang, J., Rodrigues, A. (2009). Loss-of-function variants in endothelial lipase are a cause of elevated HDL cholesterol in humans. *J Clin Invest* 119(4), 1042–1050.

Erdmann, J., Grosshennig, A., Braund, P.S., Konig, I.R., Hengstenberg, C., Hall, A.S., Linsel-Nitschke, P., Kathiresan, S., Wright, B., Tregouet, D.A. et al. New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat Genet* 41(3), 280–282.

Gudbjartsson, D.F., Bjornsdottir, U.S., Halapi, E., Helgadottir, A., Sulem, P., Jonsdottir, G.M., Thorleifsson, G., Helgadottir, H., Steinthorsdottir, V., Stefansson, H. et al. (2009). Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nat Genet* 41(3), 342–347.

Kathiresan, S., Melander, O., Anevski, D., Guiducci, C., Burtt, N., Roos, C., Hirschhorn, J., Berglund, G., Hedblad, B., Groop, L. et al. (2008). Polymorphisms associated with cholesterol and risk of cardiovascular events. *N Engl J Med* 358, 1240–1249.

Kathiresan, S., Willer, C.J., Peloso, G.M., Demissie, S., Musunuru, K., Schadt, E.E., Kaplan, L., Bennett, D., Li, Y., Tanaka, T. et al. (2009). Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet* 41, 56–65.

Kathiresan, S., Voight, B.F., Purcell, S., Musunuru, K., Ardissino, D., Mannucci, P.M., Anand, S., Engert, J.C., Samani, N.J., Schunkert, H. et al. (2009). Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet* 41(3), 334–341.

Link, E., Parish, S., Armitage, J., Bowman, L., Heath, S., Matsuda, F., Gut, I., Lathrop, M. and Collins, R. (2008). SLCO1B1 variants and statin-induced myopathy – A genomewide study. *N Engl J Med* 359, 789–799.

Ozaki, K., Sato, H., Inoue, K., Tsunoda, T., Sakata, Y., Mizuno, H., Lin, T.H., Miyamoto, Y., Aoki, A., Onouchi, Y. et al. (2009). SNPs in BRAP associated with risk of myocardial infarction in Asian populations. *Nat Genet* 41(3), 329–333.

Pollin, T.I., Damcott, C.M., Shen, H., Ott, S.H., Shelton, J., Horenstein, R.B., Post, W., McLenithan, J.C., Bielak, L.F., Peyser, P.A. et al. (2008). A null mutation in human APOC3 confers a favorable plasma lipid profile and apparent cardioprotection. *Science* 322, 1702–1705.

Ridker, P.M., Pare, G., Parker A., Zee, R.Y.L., Miletich, J. and Chasman, D.I. (2009). Polymorphism in the CETP gene region, HDL cholesterol, and risk of future myocardial infarction. *Circ Cardiovasc Genet* 2009 January 23; [Epubahead of print].

Romeo, S., Pennacchio, L.A., Fu, Y., Boerwinkle, E., Tybjaerg-Hansen, A., Hobbs, H.H. and Cohen, J.C. (2007). Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. *Nat Genet* 39, 513–516.

Romeo, S., Yin, W., Kozlitina, J., Pennacchio, L.A., Boerwinkle, E., Hobbs, H.H. and Cohen, J.C. (2009). Rare loss-of-function mutations in ANGPTL family members contribute to plasma triglyceride levels in humans. *J Clin Invest* 119, 70–79.

Sabatti, C., Service, S.K., Hartikainen, A.-L., Pouta, A., Ripatti, S., Brodsky, J., Jones, C.G., Zaitlen, N.A., Varilo, T., Kaakinen, M. et al. (2009). Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat Genet* 41, 35–46.

Thompson, A., Di Angelantonio, E., Sarwar, N., Erqou, S., Saleheen, D., Dullaart, R.P., Keavney, B., Ye, Z. and Danesh, J. (2008). Association of cholesteryl ester transfer protein genotypes with CETP mass and activity, lipid levels, and coronary risk. *JAMA* 299, 2777–2788.

Tregouet, D.A., Konig, I.R., Erdmann, J., Munteanu, A., Braund, P.S., Hall, A.S., Grosshennig, A., Linsel-Nitschke, P., Perret, C., Desuremain, M. et al. (2009). Genome-wide haplotype association study identifies the SLC22A3-LPAL2-LPA gene cluster as a risk locus for coronary artery disease. *Nat Genet* 41(3), 283–285.

Willer, C.J., Sanna, S., Jackson, A.U., Scuteri, A., Bonnycastle, L.L., Clarke, R., Heath, S.C., Timpson, N.J., Najjar, S.S., Stringham, H.M. et al. (2008). Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* 40, 161–169.

ACKNOWLEDGEMENTS

Dr Rader has received funding from the National Heart Lung and Blood Institute, the National Institute for Diabetes and Digestive and Kidney Diseases, the National Center for Research Resources, the American Heart Association, the Burroughs Wellcome Fund, the Doris Duke Charitable Foundation, and GlaxoSmithKline through an industry-academic alliance (Alternative Drug Discovery Initiative, ADDI) with the University of Pennsylvania School of Medicine. Dr Kathiresan has received funding from the National Heart, Lung, and Blood Institute, the Doris Duke Charitable Foundation, the Fannie E. Rippel Foundation, and the Donovan Family Foundation.

REFERENCES

Abifadel, M., Varret, M., Rabes, J.P., Allard, D., Ouguerram, K., Devillers, M., Cruaud, C., Benjannet, S., Wickham, L., Erlich, D. et al. (2003). Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat Genet* 34, 154–156.

ARIC (1989). The Atherosclerosis Risk in Communities (ARIC) Study: Design and objectives. The ARIC investigators. *Am J Epidemiol* 129, 687–702.

Benjannet, S., Rhainds, D., Essalmani, R., Mayne, J., Wickham, L., Jin, W., Asselin, M.C., Hamelin, J., Varret, M., Allard, D. et al. (2004). NARC-1/PCSK9 and its natural mutants: Zymogen cleavage and effects on the low density lipoprotein (LDL) receptor and LDL cholesterol. *J Biol Chem* 279, 48865–48875.

Benn, M., Nordestgaard, B.G., Jensen, J.S., Grande, P., Sillesen, H. and Tybjaerg-Hansen, A. (2005). Polymorphism in APOB associated with increased low-density lipoprotein levels in both genders in the general population. *J Clin Endocrinol Metab* 90, 5797–5803.

Berge, K.E., Tian, H., Graf, G.A., Yu, L., Grishin, N.V., Schultz, J., Kwiterovich, P., Shan, B., Barnes, R. and Hobbs, H.H. (2000). Accumulation of dietary cholesterol in sitosterolemia caused by mutations in adjacent ABC transporters. *Science* 290, 1771–1775.

Berglund, G., Elmstahl, S., Janzon, L. and Larsson, S.A. (1993). The Malmo Diet and Cancer Study. Design and feasibility. *J Intern Med* 233, 45–51.

Bodzioch, M., Orso, E., Klucken, J., Langmann, T., Bottcher, A., Diederich, W., Drobnik, W., Barlage, S., Buchler, C., Ozcurumez, M.P. et al. (1999). The gene encoding ATP-binding cassette transporter 1 is mutated in Tangier disease. *Nat Genet* 22, 347–351.

Boekholdt, S.M. and Thompson, J.F. (2003). Natural genetic variation as a tool in understanding the role of CETP in lipid levels and disease. *J Lipid Res* 44, 1080–1093.

Boekholdt, S.M., Sacks, F.M., Jukema, J.W., Shepherd, J., Freeman, D.J., McMahon, A.D., Cambien, F., Nicaud, V., de Grooth, G.J., Talmud, P.J. et al. (2005). Cholesteryl ester transfer protein TaqIB variant, high-density lipoprotein cholesterol levels, cardiovascular risk, and efficacy of pravastatin treatment: Individual patient meta-analysis of 13,677 subjects. *Circulation* 111, 278–287.

Bonnen, P.E., Pe'er, I., Plenge, R.M., Salit, J., Lowe, J.K., Shapero, M.H., Lifton, R.P., Breslow, J.L., Daly, M.J., Reich, D.E. et al. (2006). Evaluating potential for whole-genome studies in Kosrae, an isolated population in Micronesia. *Nat Genet* 38, 214–217.

Boright, A.P., Connelly, P.W., Brunt, J.H., Morgan, K. and Hegele, R.A. (1998). Association and linkage of LDLR gene variation with variation in plasma low density lipoprotein cholesterol. *J Hum Genet* 43, 153–159.

Brisson, D., Ledoux, K., Bosse, Y., St-Pierre, J., Julien, P., Perron, P., Hudson, T.J., Vohl, M.C. and Gaudet, D. (2002). Effect of apolipoprotein E, peroxisome proliferator-activated receptor alpha and lipoprotein lipase gene mutations on the ability of fenofibrate to improve lipid profiles and reach clinical guideline targets among hypertriglyceridemic patients. *Pharmacogenetics* 12, 313–320.

Brooks-Wilson, A., Marcil, M., Clee, S.M., Zhang, L.H., Roomp, K., van Dam, M., Yu, L., Brewer, C., Collins, J.A., Molhuizen, H.O. et al. (1999). Mutations in ABC1 in Tangier disease and familial high-density lipoprotein deficiency. *Nat Genet* 22, 336–345.

Brousseau, M.E., Bodzioch, M., Schaefer, E.J., Goldkamp, A.L., Kielar, D., Probst, M., Ordovas, J.M., Aslanidis, C., Lackner, K.J., Bloomfield, R.H. et al. (2001). Common variants in the gene encoding ATP-binding cassette transporter 1 in men with low HDL cholesterol levels and coronary heart disease. *Atherosclerosis* 154, 607–611.

Brown, M.L., Inazu, A., Hesler, C.B., Agellon, L.B., Mann, C., Whitlock, M.E., Marcel, Y.L., Milne, R.W., Koizumi, J. and Mabuchi, H. (1989). Molecular basis of lipid transfer protein deficiency in a family with increased high-density lipoproteins. *Nature* 342, 448–451.

Burnett, J.R. and Watts, G.F. (2007). MTP inhibition as a treatment for dyslipidaemias: Time to deliver or empty promises?. *Expert Opin Ther Targets* 11, 181–189.

Chasman, D.I., Posada, D., Subrahmanyan, L., Cook, N.R., Stanton, V.P., Jr and Ridker, P.M. (2004). Pharmacogenetic study of statin therapy and cholesterol reduction. *JAMA* 291, 2821–2827.

Chiesa, G. and Sirtori, C.R. (2003). Apolipoprotein A-IMilano: Current perspectives. *Curr Opin Lipidol* 14, 159–163.

Clee, S.M., Zwinderman, A.H., Engert, J.C., Zwarts, K.Y., Molhuizen, H.O., Roomp, K., Jukema, J.W., van Wijland, M., van Dam, M., Hudson, T.J. et al. (2001). Common genetic variation in ABCA1 is associated with altered lipoprotein levels and a modified risk for coronary artery disease. *Circulation* 103, 1198–1205.

Cohen, J., Pertsemlidis, A., Kotowski, I.K., Graham, R., Garcia, C.K. and Hobbs, H.H. (2005). Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nat Genet* 37, 161–165.

Cohen, J.C., Kiss, R.S., Pertsemlidis, A., Marcel, Y.L., McPherson, R. and Hobbs, H.H. (2004). Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* 305, 869–872.

Cohen, J.C., Boerwinkle, E., Mosley, T.H., Jr and Hobbs, H.H. (2006). Sequence variations in PCSK9 low LDL, and protection against coronary heart disease. *N Engl J Med* 354, 1264–1272.

Coon, H., Xin, Y., Hopkins, P.N., Cawthon, R.M., Hasstedt, S.J. and Hunt, S.C. (2005). Upstream stimulatory factor 1 associated with familial combined hyperlipidemia, LDL cholesterol, and triglycerides. *Hum Genet* 117, 444–451.

Coon, H., Myers, R.H., Borecki, I.B., Arnett, D.K., Hunt, S.C., Province, M.A., Djousse, L. and Leppert, M.F. (2000). Replication of linkage of familial combined hyperlipidemia to chromosome 1q with additional heterogeneous effect of apolipoprotein A-I/C-III/A-IV locus The NHLBI Family Heart Study. *Arterioscler Thromb Vasc Biol* 20, 2275–2280.

Crooke, R.M., Graham, M.J., Lemonidis, K.M., Whipple, C.P., Koo, S. and Perera, R.J. (2005). An apolipoprotein B antisense oligonucleotide lowers LDL cholesterol in hyperlipidemic mice without causing hepatic steatosis. *J Lipid Res* 46, 872–884.

Cuchel, M., Bloedon, L.T., Szapary, P.O., Kolansky, D.M., Wolfe, M.L., Sarkis A., Millar, J.S., Ikewaki, K., Siegelman, E.S., Gregg, R.E. et al. (2007). Inhibition of microsomal triglyceride transfer protein in familial hypercholesterolemia. *N Engl J Med* 356, 148–156.

Frikke-Schmidt, R., Nordestgaard, B.G., Jensen, G.B. and Tybjaerg-Hansen, A. (2004). Genetic variation in ABC transporter A1 contributes to HDL cholesterol in the general population. *J Clin Invest* 114, 1343–1353.

Garcia, C.K., Wilund, K., Arca, M., Zuliani, G., Fellin, R., Maioli, M., Calandra, S., Bertolini, S., Cossu, F., Grishin, N. et al. (2001). Autosomal recessive hypercholesterolemia caused by mutations in a putative LDL receptor adaptor protein. *Science* 292, 1394–1398.

Goldstein, J.L. and Brown, M.S. (2001). Molecular medicine. The cholesterol quartet. *Science* 292, 1310–1312.

Gordon, D.J. and Rifkind, B.M. (1989). High-density lipoproteins – the clinical implications of recent studies. *N Engl J Med* 321, 1311–1316.

Grundy, S.M., Chait, A. and Brunzell, J.D. (1987). Familial combined hyperlipidemia workshop. *Arteriosclerosis* 7, 203–207.

Guerra, R., Wang, J., Grundy, M.S. and Cohen, C.J. (1997). A hepatic lipase (LIPC) allele associated with high plasma concentrations of high density lioporotein cholesterol. *Proc Natl Acad Sci U S A* 94, 4532–4537.

Hegele, R., Little, J.A., Vezina, C., Maguire, G., Tu, L., Wolever, T., Jenkins, D. and Connelly, P. (1993). Hepatic lipase deficiency: Clinical, biochemical, and molecular genetic characteristics. *Arterioscler Thromb* 13, 720–728.

Hegele, R.A., Guy, J., Ban, M.R. and Wang, J. (2005). NPC1L1 haplotype is associated with inter-individual variation in plasma low-density lipoprotein response to ezetimibe. *Lipids Health Dis* 4, 16.

Helgadottir, A., Thorleifsson, G., Manolescu, A., Gretarsdottir, S., Blondal, T., Jonasdottir, A., Jonasdottir, A., Sigurdsson, A., Baker, A., Palsson, A. et al. (2007). A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*.

Hirschhorn, J.N. and Daly, M.J. (2005). Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 6, 95–108.

Hobbs, H.H. and Rader, D.J. (1999). ABC1: Connecting yellow tonsils, neuropathy, and very low HDL. *J Clin Invest* 104, 1015–1017.

Horton, J.D., Cohen, J.C. and Hobbs, H.H. (2007). Molecular biology of PCSK9: Its role in LDL metabolism. *Trends Biochem Sci* 32, 71–77.

Ikewaki, K., Rader, D.J., Sakamoto, T., Nishiwaki, M., Wakimoto, N., Schaefer, J.R., Ishikawa, T., Fairwell, T., Zech, L.A. and Nakamura, H. (1993). Delayed catabolism of high density lipoprotein apolipoproteins A-I and A-II in human cholesteryl ester transfer protein deficiency. *J Clin Invest* 92, 1650–1658.

Ikewaki, K., Nishiwaki, M., Sakamoto, T., Ishikawa, T., Fairwell, T., Zech, L. A., Nakamura, H., Brewer, H.B., Jr and Rader, D.J. (1995). Increased catabolic rate of low density lipoproteins in humans with cholesteryl ester transfer protein deficiency. *J Clin Invest* 96, 1573–1581.

Inazu, A., Brown, M.L., Hesler, C.B., Agellon, L.B., Koizumi, J., Takata, K., Maruhama, Y., Mabuchi, H. and Tall, A.R. (1990). Increased high-density lipoprotein levels caused by a common cholesteryl-ester transfer protein gene mutation. *N Engl J Med* 323, 1234–1238.

Kannel, W.B., Dawber, T.R., Kagan, A., Revotskie, N. and Stokes, J., 3rd (1961). Factors of risk in the development of coronary heart disease – six year follow-up experience. The Framingham Study. *Ann Intern Med* 55, 33–50.

Kastelein, J.J., Wedel, M.K., Baker, B.F., Su, J., Bradley, J.D., Yu, R.Z., Chuang, E., Graham, M.J. and Crooke, R.M. (2006). Potent reduction of apolipoprotein B and low-density lipoprotein cholesterol by short-term administration of an antisense inhibitor of apolipowprotein B. *Circulation* 114, 1729–1735.

Kathiresan, S., Manning, A.K., Demissie, S., D'Agostino, R.B., Surti, A., Guiducci, C., Gianniny, L., Burtt, N.P., Melander, O., Orho-Melander, M. et al. (2007). A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study. *BMC Med Genet* 8(Suppl 1), S17.

Kathiresan, S. et al. (2008a). Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet* 40, 189–197.

Kathiresan, S. et al. (2008b). Polymorphisms associated with cholesterol and risk of cardiovascular events. *N Engl J Med* 358, 1240–1249.

Kathiresan, S. et al. (2008c). A PCSK9 missense variant associated with a reduced risk of early-onset myocardial infarction. *N Engl J Med* 358, 2299–2300.

Kooner, J.S. et al. (2008). Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides. *Nat Genet* 40, 149–151.

Kotowski, I.K., Pertsemlidis, A., Luke, A., Cooper, R.S., Vega, G.L., Cohen, J.C. and Hobbs, H.H. (2006). A spectrum of PCSK9 alleles contributes to plasma levels of low-density lipoprotein cholesterol. *Am J Hum Genet* 78, 410–422.

Kuivenhoven, J.A., Pritchard, H., Hill, J., Frohlich, J., Assmann, G. and Kastelein, J. (1997). The molecular pathology of lecithin:cholesterol acyltransferase (LCAT) deficiency syndromes. [Review] [102 refs]. *J Lipid Res* 38, 191–205.

Lai, C.Q., Demissie, S., Cupples, L.A., Zhu, Y., Adiconis, X., Parnell, L.D., Corella, D. and Ordovas, J.M. (2004). Influence of the APOA5 locus on plasma triglyceride, lipoprotein subclasses, and CVD risk in the Framingham Heart Study. *J Lipid Res* 45, 2096–2105.

Lai, C.Q., Arnett, D.K., Corella, D., Straka, R.J., Tsai, M.Y., Peacock, J.M., Adiconis, X., Parnell, L.D., Hixson, J.E., Province, M.A. et al. (2007). Fenofibrate effect on triglyceride and postprandial response of apolipoprotein A5 variants: The GOLDN study. *Arterioscler Thromb Vasc Biol* 27, 1417–1425.

Lee, J.C., Lusis, A.J. and Pajukanta, P. (2006). Familial combined hyperlipidemia: Upstream transcription factor 1 and beyond. *Curr Opin Lipidol* 17, 101–109.

Lewis, G.F. and Rader, D.J. (2005). New insights into the regulation of HDL metabolism and reverse cholesterol transport. *Circ Res* 96, 1221–1232.

Liao, W., Hui, T.Y., Young, S.G. and Davis, R.A. (2003). Blocking microsomal triglyceride transfer protein interferes with apoB secretion without causing retention or stress in the ER. *J Lipid Res* 44, 978–985.

Linsel-Nitschke, P. and Tall, A.R. (2005). HDL as a target in the treatment of atherosclerotic cardiovascular disease. *Nat Rev Drug Discov* 4, 193–205.

Mahley, R.W., Huang, Y. and Rall, S.C.J. (1999). Pathogenesis of type III hyperlipoproteinemia (dysbetalipoproteinemia): Questions, quandaries, and paradoxes. *J Lipid Res* 40, 1933–1949.

Marcais, C., Verges, B., Charriere, S., Pruneta, V., Merlin, M., Billon, S., Perrot, L., Drai, J., Sassolas, A., Pennacchio, L.A. et al. (2005). Apoa5 Q139X truncation predisposes to late-onset hyperchylomicronemia due to lipoprotein lipase impairment. *J Clin Invest* 115, 2862–2869.

Marcil, M., Brooks-Wilson, A., Clee, S.M., Roomp, K., Zhang, L.H., Yu, L., Collins, J.A., van Dam, M., Molhuizen, H.O., Loubster, O. et al. (1999). Mutations in the ABC1 gene in familial HDL deficiency with defective cholesterol efflux [see comments]. *Lancet* 354, 1341–1346.

Maxwell, K.N. and Breslow, J.L. (2004). Adenoviral-mediated expression of Pcsk9 in mice results in a low-density lipoprotein receptor knockout phenotype. *Proc Natl Acad Sci U S A* 101, 7100–7105.

McPherson, R., Pertsemlidis, A., Kavaslar, N., Stewart, A., Roberts, R., Cox, D.R., Hinds, D.A., Pennacchio, L.A., Tybjaerg-Hansen, A., Folsom, A.R. et al. (2007). A common allele on chromosome 9 associated with coronary heart disease. *Science* 316, 1488–1491.

Ng, D., Leiter, L., Vezina, C., Connelly, P. and Hegele, R. (1994). Apolipoprotein A-I Q[-2]X causing isolated apolipoprotein A-I deficiency in a family with analphalipoproteinemia. *J Clin Invest* 93, 223–229.

Nissen, S.E., Tsunoda, T., Tuzcu, E.M., Schoenhagen, P., Cooper, C.J., Yasin, M., Eaton, G.M., Lauer, M.A., Sheldon, W.S., Grines, C.L. et al. (2003). Effect of recombinant ApoA-I Milano on coronary atherosclerosis in patients with acute coronary syndromes: A randomized controlled trial. *JAMA* 290, 2292–2300.

Norum, R.A., Lakier, J.B., Goldstein, S., Angel, A., Goldberg, R.B., Block, W.D., Noffze, D.K., Dolphin, P.J., Edelglass, J., Bogorad, D.D. et al. (1982). Familial deficiency of apolipoproteins A-I and C-III and precocious coronary-artery disease. *N Engl J Med* 306, 1513–1519.

Offermanns, S. (2006). The nicotinic acid receptor GPR109A (HM74A or PUMA-G) as a new therapeutic target. *Trends Pharmacol Sci* 27, 384–390.

Pajukanta, P., Nuotio, I., Terwilliger, J.D., Porkka, K.V., Ylitalo, K., Pihlajamaki, J., Suomalainen, A.J., Syvanen, A.C., Lehtimaki, T., Viikari, J.S. et al. (1998). Linkage of familial combined hyperlipidaemia to chromosome 1q21-q23. *Nat Genet* 18, 369–373.

Pajukanta, P., Lilja, H.E., Sinsheimer, J.S., Cantor, R.M., Lusis, A.J., Gentile, M., Duan, X.J., Soro-Paavonen, A., Naukkarinen, J. et al. (2004). Familial combined hyperlipidemia is associated with upstream transcription factor 1 (USF1). *Nat Genet* 36, 371–376.

Pennacchio, L.A., Olivier, M., Hubacek, J.A., Cohen, J.C., Cox, D.R., Fruchart, J.C., Krauss, R.M. and Rubin, E.M. (2001). An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing. *Science* 294, 169–173.

Pollin, T.I., Damcott, C.M., Shen, H., Ott, S.H., Shelton, J., Horenstein, R.B., Post, W., McLenithan, J.C., Bielak, L.F., Peyser, P.A. et al. (2008). A null mutation in human APOC3 confers a favorable plasma lipid profile and apparent cardioprotection. *Science* 322, 1702–1705.

Priore Oliva, C., Pisciotta, L., Li Volti, G., Sambataro, M.P., Cantafora, A., Bellocchio, A., Catapano, A., Tarugi, P., Bertolini, S. and Calandra, S. (2005). Inherited apolipoprotein A-V deficiency in severe hypertriglyceridemia. *Arterioscler Thromb Vasc Biol* 25, 411–417.

Rader, D.J. (2004). Inhibition of cholesteryl ester transfer protein activity: A new therapeutic approach to raising high-density lipoprotein. *Curr Atheroscler Rep* 6, 398–405.

Rader, D.J. (2006). Molecular regulation of HDL metabolism and function: Implications for novel therapies. *J Clin Invest* 116, 3090–3100.

Rader, D.J. and Brewer, H.B., Jr (1993). Abetalipoproteinemia. New insights into lipoprotein assembly and vitamin E metabolism from a rare genetic disease. *JAMA* 270, 865–869.

Rader, D.J., Cohen, J. and Hobbs, H.H. (2003). Monogenic hypercholesterolemia: New insights in pathogenesis and treatment. *J Clin Invest* 111, 1795–1803.

Rader, D.J., Ikewaki, K., Duverger, N., Schmidt, H., Pritchard, H., Frohlich, J., Dumon, M.F., Fairwell, T., Zech, L., Santamarina-Fojo, S. et al. (1994). Markedly accelerated catabolism of apolipoprotein A-II (ApoA-II) and high density lipoproteins containing ApoA-II in classic lecithin: Cholesterol acyltransferase deficiency and fish-eye disease. *J Clin Invest* 93, 321–330.

Rip, J., Nierman, M.C., Ross, C.J., Jukema, J.W., Hayden, M.R., Kastelein, J.J., Stroes, E.S. and Kuivenhoven, J.A. (2006).

Lipoprotein lipase S447X: A naturally occurring gain-of-function mutation. *Arterioscler Thromb Vasc Biol* 26, 1236–1245.

Rust, S., Rosier, M., Funke, H., Real, J., Amoura, Z., Piette, J.C., Deleuze, J.F., Brewer, B., Duverger, N., Denefle, P. et al. (1999). Tangier disease is caused by mutations in the gene encoding ATP-binding cassette transporter 1. *Nat Genet* 22, 352–355.

Samani, N.J., Erdmann, J., Hall, A.S., Hengstenberg, C., Mangino, M., Mayer, B., Dixon, R.J., Meitinger, T., Braund, P., Wichmann, H.E. et al. (2007). Genomewide association analysis of coronary artery disease. *N Engl J Med* 357, 443–453.

Santamarina-Fojo, S. (1992). Genetic dyslipoproteinemias: Role of lipoprotein lipase and apolipoprotein C-II. *Curr Opin Lipidol* 3, 186–195.

Saxena, R., Voight, B.F., Lyssenko, V., Burtt, N.P., de Bakker, P.I., Chen, H., Roix, J.J., Kathiresan, S., Hirschhorn, J.N., Daly, M.J. et al. (2007). Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316, 1331–1336.

Schaefer, E.J., Zech, L.A., Schwartz, D.E. and Brewer, H.B., Jr (1980). Coronary heart disease prevalence and other clinical features in familial high-density lipoprotein deficiency (Tangier disease). *Ann Intern Med* 93, 261–266.

Schaefer, E.J., Heaton, W.H., Wetzel, M.G. and Brewer, H.B., Jr (1982). Plasma apolipoprotein A-1 absence associated with a marked reduction of high density lipoproteins and premature coronary artery disease. *Arteriosclerosis* 2, 16–26.

Schonfeld, G., Lin, X. and Yue, P. (2005). Familial hypobetalipoproteinemia: Genetics and metabolism. *Cell Mol Life Sci* 62, 1372–1378.

Scott, L.J., Mohlke, K.L., Bonnycastle, L.L., Willer, C.J., Li, Y., Duren, W.L., Erdos, M.R., Stringham, H.M., Chines, P.S., Jackson, A.U. et al. (2007). A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316, 1341–1345.

Sharp, D., Blinderman, L., Combs, K.A., Kienzle, B., Ricci, B., Wager-Smith, K., Gil, C.M., Turck, C.W., Bouna, M.E., Rader, D.J. et al. (1993). Cloning and gene defects in microsomal triglyceride transfer protien associated with abetalipoproteinemia. *Nature* 365, 65–69.

Scuteri, A., Sanna, S., Chen, W.M., Uda, M., Albai, G., Strait, J., Najjar, S., Nagaraja, R., Orru, M., Usala, G. et al. (2007). Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. *PLoS Genet* 3, e115.

Shoulders, C.C., Jones, E.L. and Naoumova, R.P. (2004). Genetics of familial combined hyperlipidemia and risk of coronary heart disease. *Hum Mol Genet* 13(Spec No 1), R149–rR160.

Sing, C.F. and Davignon, J. (1985). Role of the apolipoprotein E polymorphism in determining normal plasma lipid and lipoprotein variation. *Am J Hum Genet* 37, 268–285.

Spann, N.J., Kang, S., Li, A.C., Chen, A.Z., Newberry, E.P., Davidson, N.O., Hui, S.T. and Davis, R.A. (2006). Coordinate transcriptional repression of liver fatty acid-binding protein and microsomal triglyceride transfer protein blocks hepatic very low density lipoprotein secretion without hepatosteatosis. *J Biol Chem* 281, 33066–33077.

Splansky, G.L., Corey, D., Yang, Q., Atwood, L.D., Cupples, L.A., Benjamin, E.J., D'Agostino, R.B., Sr, Fox, C.S., Larson, M.G., Murabito, J.M. et al. (2007). The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: Design, recruitment, and initial examination. *Am J Epidemiol* 165, 1328–1335.

Suviolahti, E., Lilja, H.E. and Pajukanta, P. (2006). Unraveling the complex genetics of familial combined hyperlipidemia. *Ann Med* 38, 337–351.

Tall, A.R., Yvan-Charvet, L. and Wang, N. (2007). The failure of torcetrapib: Was it the molecule or the mechanism? *Arterioscler Thromb Vasc Biol* 27, 257–260.

Thompson, J.F., Man, M., Johnson, K.J., Wood, L.S., Lira, M.E., Lloyd, D.B., Banerjee, P., Milos, P.M., Myrand, S.P., Paulauskis, J. et al. (2005). An association study of 43 SNPs in 16 candidate genes with atorvastatin response. *Pharmacogenomics J* 5, 352–358.

Tybjaerg-Hansen, A. and Humphries, S.E. (1992). Familial defective apolipoprotein B-100: A single mutation that causes hypercholestolemia and premature coronary artery disease. *Atherosclerosis* 96, 91–107.

von Eckardstein, A. (2006). Differential diagnosis of familial high density lipoprotein deficiency syndromes. *Atherosclerosis* 186, 231–239.

Wang, J., Williams, C.M. and Hegele, R.A. (2005). Compound heterozygosity for two non-synonymous polymorphisms in NPC1L1 in a non-responder to ezetimibe. *Clin Genet* 67, 175–177.

Wetterau, J.R., Gregg, R.E., Harrity, T.W., Arbeeny, C., Cap, M., Connolly, F., Chu, C., George, R.J., Gordan, D.A., Jamil, H. et al. (1998). An MTP inhibitor that normalizes atherogenic lipoprotein levels in WHHL rabbits. *Science* 282, 751–754.

Willer, C.J. et al. (2008). Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* 40, 161–169.

Willer, C.J., Sanna, S., Jackson, A.U., Scuteri, A., Bonnycastle, L.L., Clarke, R., Heath, S.C., Timpson, N.J., Najjar, S.S., Stringham, H.M. et al. (2008). Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* 40, 161–169.

Wilson, J.G., Rotimi, C.N., Ekunwe, L., Royal, C.D., Crump, M.E., Wyatt, S.B., Steffes, M.W., Adeyemo, A., Zhou, J., Taylor, H.A., Jr et al. (2005). Study design for genetic analysis in the Jackson Heart Study. *Ethn Dis* 15, S6-30–S6-37.

WTCCC (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661–678.

Zellner, C., Pullinger, C.R., Aouizerat, B.E., Frost, P.H., Kwok, P.Y., Malloy, M.J. and Kane, J.P. (2005). Variations in human HM74 (GPR109B) and HM74A (GPR109A) niacin receptors. *Hum Mutat* 25, 18–21.

Zhao, Z., Tuakli-Wosornu, Y., Lagace, T.A., Kinch, L., Grishin, N.V., Horton, J.D., Cohen, J.C. and Hobbs, H.H. (2006). Molecular characterization of loss-of-function mutations in PCSK9 and identification of a compound heterozygote. *Am J Hum Genet* 79, 514–523.

Zimmermann, T.S., Lee, A.C., Akinc, A., Bramlage, B., Bumcrot, D., Fedoruk, M.N., Harborth, J., Heyes, J.A., Jeffs, L.B., John, M. et al. (2006). RNAi-mediated gene silencing in non-human primates. *Nature* 441, 111–114.

RECOMMENDED RESOURCES

www.hapmap.org:
A catalog of human genetic variation in 270 individuals from four different populations.

www.pharmgkb.org/:
A resource for pharmacogenetics and pharmacogenomics.

http://bioinf.itmat.upenn.edu/cvdsnp:
A website for a vascular disease 50K SNP array developed by a consortium of the Institute for Translational Medicine and Therapeutics (University of Pennsylvania) and the Broad Institute (MIT/Harvard).

http://www.broad.mit.edu/tools/data.html:
A website hosted by the Broad Institute with a rich source of genetic information and resources.

CHAPTER

Genomics of Myocardial Infarction

Carlos A. Hubbard and Eric J. Topol

INTRODUCTION

Despite advances in screening and treatment, coronary artery disease (CAD) and myocardial infarction (MI) remain the leading causes of death in the world (Bonow et al., 2002). The global economic implications of MI are staggering, with the 2006 estimated cost of CAD at over $142 billion in the United States alone (Thom et al., 2006). The economic consequences are particularly felt in developing countries which account for 80% of the burden of cardiovascular disease and where resources for prevention and access to treatment are limited (Yusuf et al., 2004). Therefore, a better understanding of the molecular mechanisms that impart risk would have a dramatic economic effect by allowing earlier detection of individuals at highest risk and targeting preventive therapies to individuals most likely to benefit.

Advanced atherosclerotic disease can lead to the formation of large lesions that can encroach upon the lumen of the coronary vasculature and result in ischemic injury and even infarction when myocardial oxygen demand is high. However, the catastrophic consequences of CAD are generally related to acute MI due to the rupture of vulnerable plaques and subsequent occlusive thrombosis. Although there is a high prevalence of atherosclerotic coronary disease worldwide, only a fraction of patients with this disease progress to having an acute MI. Although a number of risk factors have been shown to contribute to the development of cardiovascular disease, a family history of CAD remains one of the most powerful independent predictors of risk, which suggests a significant genetic component to the disease. Several linkage studies suggest that the genes responsible for plaque rupture and thrombosis may differ from the genes responsible for atherosclerotic disease progression, which may account for this discrepancy.

PREDISPOSITION

Genetic Influence on Risk of MI

Data from the Swedish Twin Registry (Lichtenstein et al., 2002) have been used in several studies to establish the moderate but significant effect of genetic factors on the relative risk of death due to CAD. Monozygotic twins have a significantly higher relative risk of death from CAD than dizygotic twins even when outcomes are adjusted for the effects of conventional risk factors associated with CAD (Marenberg et al., 1994; Zdravkovic et al., 2004). Even among siblings who are affected by MI the angiographic location of disease is remarkably concordant (Fischer et al., 2005). The molecular events that lead to plaque rupture and acute MI are likely influenced by a host of intricately related genetic factors involved in a variety of cellular processes such as the inflammatory cascade, extracellular matrix regulation, apoptosis and lipid metabolism. Advances in molecular biology have provided us with an array of tools to begin elucidating the genetic factors underlying these diverse mechanisms.

A significant number of genes involved in the inflammatory cascade are clustered on the short arm of chromosome 6p21 in the major histocompatibility complex region. A review of the available literature regarding polymorphisms of genes (TNF-alpha, lymphotoxin-alpha, HLA-DR, heat shock protein 70-1, hemochromatosis gene and C4) in this region points out the

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

variability in the results to date (Porto et al., 2005). In most cases the strong linkage disequilibrium among these genes appears to point to a complex haplotypic pattern of cardiovascular risk rather than a single genetic variant.

Atherosclerotic plaque instability is a central concept in the current theories of the process of plaque rupture. The matrix metalloproteinase (MMP) gene family has been implicated in the development of unstable plaques. The proteins MMP-1, 2, 3, 7, 8, 9, 12, 13 and 14 are all expressed in atherosclerotic plaques (Faber et al., 2002). Upregulation of MMP-9 by a functional polymorphisms in the promoter region of the gene was shown to correlate with an increase in angiographically significant CAD (Zhang et al., 1999). Also expressed in plaques are the tissue inhibitors of metalloproteinases (TIMPs), which bind and inhibit the activity of MMPs. Upregulation of TIMP-1 and TIMP-2 expression in human saphenous vein cultures resulted in a reduction of neointimal thickening (George et al., 1998a, b). The balance between MMPs and TIMPs affects overall plaque stability such that derangements in the regulation of this process may lead to plaque rupture and MI.

Genome-wide Scans for Linkage

Several genome-wide scans have been performed on diverse ethnic populations in order to identify genes responsible for the development of CAD (Table 24.1). In a Finnish population, loci that correlated with CAD were identified on chromosomes 2q21.1–22 (LOD = 3.2) and Xq23-26 (LOD = 3.5) (Pajukanta et al., 2000). In a northeastern Indian population, a site on chromosome 16p13-pter (LOD = 3.06) was suggestive for CAD (Francke et al., 2001). In an Australian population, a locus on chromosome 2q36-37.3 (LOD = 2.63) was found to significantly correlate with the presence of CAD (Harrap et al., 2002). A whole genome scan of 1406 individuals from 513 German families with a history of early CAD or MI identified a locus for MI on chromosome 14q12.3-13.0 (LOD = 3.9) (Broeckel et al., 2002). The diversity in loci identified by these studies may be due in part to the phenotypic inconsistencies, diversity of the ancestry of the populations, and inadequate statistical power.

The GENECARD study was a genome-wide scan for genetic regions linked to early-onset CAD in over 400 families with a history of early-onset CAD (Hauser et al., 2004). This analysis identified a region on chromosome 3q13 (LOD = 3.5) that was associated with early-onset CAD. Other regions of chromosome 3q have been associated with CAD in three earlier studies (Broeckel et al., 2002; Francke et al., 2001; Harrap et al., 2002). The Diabetes Heart Study also identified linkage of this region of chromosome 3q13 with CAD in type 2 diabetics (Bowden et al., 2006).

A genome-wide scan for susceptibility genes for MI in a population of European-American ancestry from 428 multiplex families with familial premature CAD and MI identified a significant locus for MI on chromosome 1p34-36 (LOD = 11.68) but did not detect a significant locus for CAD (Wang et al., 2004). The chromosome 1p34-36 locus contains the gene for connexin-37 (CX37), which is a gap junction protein expressed in the arterial endothelium and is involved in vascular growth, aging and regeneration after injury. The C allele of the CX37 SNP P319S has previously been shown to be associated with CAD in a Taiwanese population, as well as carotid intima thickening in Swedish men (Boerma et al., 1999; Yeh et al., 2001). The T allele of this same CX37 SNP has also been associated with risk for MI in a case-control study of Japanese men (Yamada et al., 2002). Whether connexin-37 or another gene(s) accounts for the linkage peak remains unresolved.

TABLE 24.1 Genome-wide scans for CAD/MI

Study (ref.), year	Population	No. of families	Mean age (years)	Locus/Loci	Candidate gene
Pajukanta et al. (2000)	Finnish	156	<55	2q21, Xq23	NTD
Francke et al. (2001)	Mauritian	99	47	16p13	NTD
Broeckel et al. (2002)	European	513	52	14q32	NTD
Harrap et al. (2002)	Australian	61	62	2q36	NTD
Ozaki et al. (2002)	Japanese	1133	62.5	6p21	LTA
Wang et al. (2004)	American	428	44	1p34-36	CX37
Hauser et al. (2004); Connelly et al. (2006)	European-American	438	<56	3q13	GATA2
Helgadottir et al. (2004)	Icelandic	296		13q12-13	ALOX5AP
WTCCC (2007); Helgadottir et al. (2007); McPherson et al. (2007); Samani et al. (2007)	European, Icelandic, American			9p21	NTD

CX37 = connexin-37; LTA = lymphotoxin alpha; GATA2 = GATA2 transcription factor; ALOX5AP = arachidonate 5-lipooxygenase-activating protein; NTD = none to date.
Reproduced with permission from Topol (2005) with modification.

Accordingly, studies are targeting the smaller population of individuals with CAD who ultimately suffer MI. Evidence is mounting that some of the genes responsible for progression to MI may differ from the genes responsible for CAD. Although multigenetic interactions are likely responsible for the majority of cardiovascular risk, some exceptions have more recently been described with findings of significant risk being directly attributable to variants in a single gene. The development of the International Haplotype Map is allowing researchers to utilize single nucleotide polymorphisms (SNP) to hone in on the areas of interest identified by linkage studies and identify specific genes that are associated with CAD and MI (Table 24.2).

Association Studies

The GENEQUEST study was the first high-throughput SNP study of individuals with premature MI (Topol et al., 2001). This study examined 72 SNPs of 62 vascular genes in 398 families identified three members of the thrombospondin (TSP) gene family of matricellular proteins, TSP-1, -2 and -4, which were associated with familial premature MI. The findings of significant association of TSP-2 and TSP-4 with premature MI have subsequently been replicated in other studies and by different investigators (Boekholdt et al., 2002; McCarthy et al., 2004; Wessel et al., 2004).

The TSP-1 SNP results in a substitution of serine for asparagine at residue 700 of the protein and although rare (1% frequency),

TABLE 24.2 Specific genes associated with MI risk

Study (ref.), year	Gene	SNP	Estimated Risk
Topol et al. (2001)	TSP-1 TSP-2 TSP-4	N700S 3'UTR region A387P	OR 11.9, $p = 0.041$ OR 0.31, $p = 0.0018$ OR 1.89, $p = 0.002$
Wang et al. (2003)	MEF2A	D15S120	OR/RR/HR not available
Yamada et al. (2002)	CX37 PAI-1 Stromelysin-1	C1019T 4G-668/5G 5A-1171/6A	OR 1.4, $p < 0.002$ OR 1.6, $p < 0.001$ OR 4.7, $p < 0.001$
Ozaki et al. (2002)	LTA	T26N A252G	OR 1.78, $p < 0.001$ OR 1.69, $p < 0.001$
Ozaki et al. (2004)	Galectin-2	C3279T	OR 1.57, $p < 0.001$
Helgadottir et al. (2004)	ALOX5AP	HapA[a]	RR 1.8, $p < 0.001$
Shiffman et al. (2005)	ROS1 Palladin TAS2R50 OR13G1	rs619203 rs12510359 rs1376251 rs1151640	OR 1.23, $p = 0.012$ OR 1.25, $p = 0.0028$ OR 1.28, $p = 0.0018$ OR 1.19, $p = 0.013$
Helgadottir et al. (2006)	LTA4H	HapK[a]	European-Americans: RR 1.31, $p = 0.037$ African-Americans: RR 4.39, $p = 0.008$
Ozaki et al. (2006)	PSMA6	rs1048990[a]	OR 1.36, $p = 0.002$
Cohen et al. (2006)	PCSK9	Nonsense variant: Y142X C679X Missense variant: R46L	HR 0.11, $p = 0.03$ HR 0.5, $p = 0.003$
Connelly et al. (2006)	GATA2	rs2713604[a] rs3803[a]	OR 1.5, $p = 0.011$ OR 0.7, $p = 0.028$
Shiffman et al. (2006)	VAMP8 HNRPUL1	rs1010 rs11881940	OR 1.75, $p = 0.025$ OR 1.92, $p = 0.0043$

TSP = thrombospondin; MEF2A = myocyte-enhancing factor 2A; CX37 = connexin-37; PAI-1 = plasminogen activator inhibitor type-1; LTA = lymphotoxin alpha; ALOX5AP = arachidonate 5-lipooxygenase-activating protein; LTA4H = leukotriene A4 hydrolase; PSMA6 = proteasome subunit type 6; PCSK9 = proprotein convertase subtilisin/kexin type 9 serine protease; HapA = 4 SNP HapA haplotype; HapK = 10 SNP HapK haplotype; GATA2 = GATA2 transcription factor; VAMP8 = vesicle-associated membrane protein 8; HNRPUL1 = heterogeneous nuclear ribonucleoprotein U-like 1; OR = odds ratio; RR = relative risk; HR = hazard ratio.
[a]Indicates validated in more than one cohort.

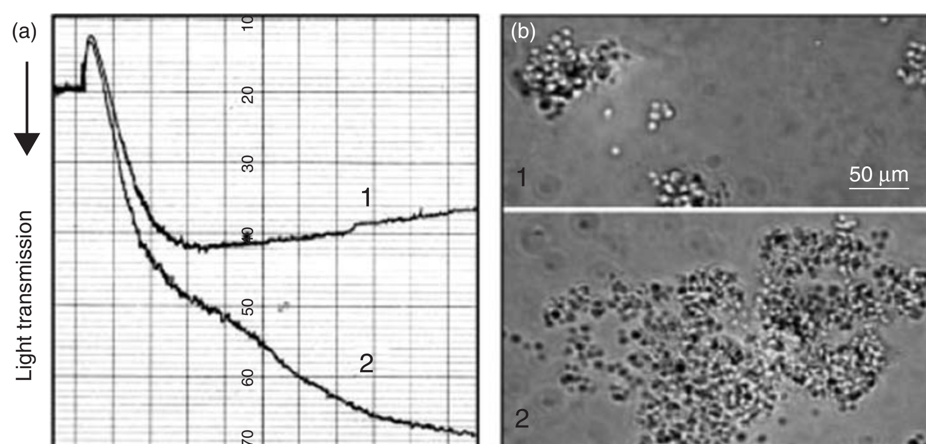


Figure 24.1 TSP-1 variants and platelet aggregation: (a) Platelet aggregation in the absence (curve 1) or presence (curve 2) of TSP-1 variant Asn-700 induced by 20 µM ADP and monitored by decreased light transmission. (b) Aggregates of ADP-stimulated platelets in the absence (panel 1) or presence (panel 2) of purified rTSP-1 at 100 µM. (Reproduced with permission from Narizhneva et al., 2004.)

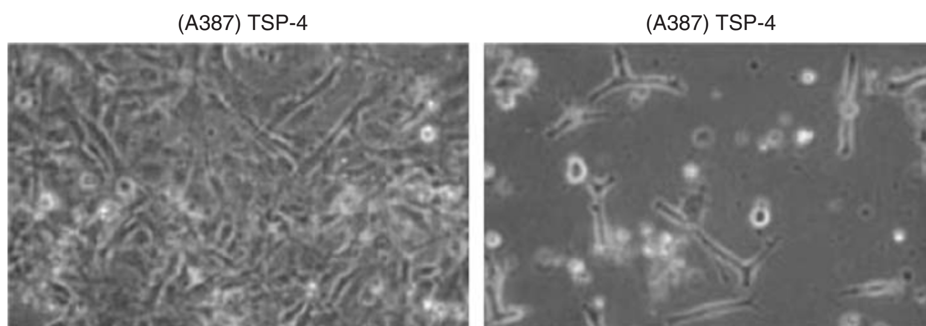


Figure 24.2 Effect of TSP-4 variants on adhesion of HUVECs. Cells were plated on plastic preincubated with purified TSP-4 variants (50,000 cells/well of 24-well plate) and photographed after 24 h of culture. The presence of the P387 variant interferes with HUVEC cell binding and proliferation. (Reproduced with permission from Stenina et al., 2003.)

homozygous individuals had the highest association with MI (OR = 8.66). The TSP-1 SNP has been shown functionally to increase surface expression and platelet aggregation, which may account for its high association with MI (Narizhneva et al., 2004) (Figure 24.1). The TSP-4 SNP is more common with the minor allele present in over 30% of individuals and confers a near twofold risk of MI (OR = 1.89). The TSP-4 SNP results in a substitution of proline for alanine, which has been correlated with gain-of-function, proatherogenic effects by interfering with endothelial cell adhesion and proliferation (Stenina et al., 2003) (Figure 24.2), along with activation of neutrophils (Pluskota et al., 2005).

Another gene that has been identified as a heritable risk for MI is the gene for the transcription factor myocyte-enhancing factor 2A (MEF2A) (Wang et al., 2003). In a large pedigree study of a family with a history of premature CAD and MI, a genome-wide scan identified a locus on chromosome 15q26 that was highly associated with risk (Figure 24.3). An analysis of the candidate genes in this area revealed a 21-bp deletion mutation in exon 11 of the MEF2A gene that imparts an autosomal dominant inheritable risk of MI. This deletion was shown to prevent localization of MEF2A in the nucleus and disrupts transcription. A subsequent study of over 400 nonrelated MI cases and controls found MEF2A nonsynonymous point mutations in exon 7 in MI cases but not controls (Bhagavatula et al., 2004). These point mutations also appear to limit the transcription activity of MEF2A and provide further evidence of the importance of this gene in the development of CAD and risk for MI.

Recently, further analysis of the GENECARD data has identified two SNPs in the 3′ untranslated region of the transcription factor GATA2 gene as being significantly associated with early-onset CAD (Connelly et al., 2006). GATA2 is expressed in hematopoietic stem cells but also in locations susceptible to the development of atherosclerosis such as aortic endothelial cells and smooth muscle cells. The association of the GATA2 SNPs with early-onset CAD was identified first in the familial early-onset CAD cohort of GENECARD

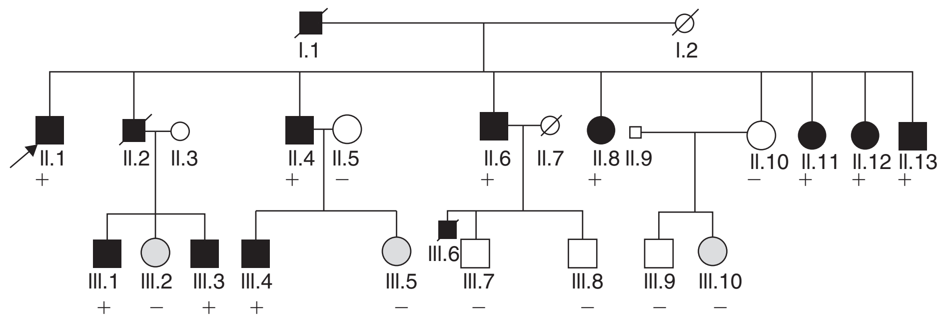


Figure 24.3 MEF2A intragenic deletion cosegregates with CAD in kindred QW1576. The pedigree shows genetic status: +indicates presence of the 21-bp deletion of MEF2A (heterozygous); −indicates the absence of the deletion. Individuals with CAD are indicated by solid squares (males) or circles (females). Individuals under 50 years of age without known CAD are shown in light gray. The proband is indicated by an arrow and deceased individuals are indicated by a slash (/). (Reproduced with permission from Wang et al., 2003.)

and was subsequently validated in a separate nonfamilial young affected CAD case-control cohort. The functional significance of the GATA2 SNPs is unknown, but these findings suggest that GATA2 regulated genes may be significantly involved in early-onset CAD and its progression.

A number of inflammation-mediating molecules, such as the cytokine lymphotoxin-alpha (LTA), have been implicated in the process of plaque rupture. A massive genome-wide SNP study of Japanese patients analyzed over 65,000 SNPs in 13,738 genes and found 2 functional SNPs in the LTA gene that were highly associated with risk of MI (Ozaki et al., 2002) (Figure 24.4). These SNPs were shown to increase expression of vascular cell adhesion molecule and E-selectin which creates a proinflammatory environment. The PROCARDIS Consortium later confirmed the association of a functional allele of LTA (N26 [804A]) with CAD in white Europeans (2004).

LTA protein binds to galectin-2, a member of the galactose-binding lectin family, and this binding is critical for the extracellular expression of LTA. Recently a SNP in *LGALS2* encoding galectin-2 has been shown to be significantly associated with susceptibility to MI by affecting the transcriptional level of galectin-2 leading to altered secretion of LTA (Ozaki et al., 2004). LTA and galectin-2 were shown to be expressed in smooth muscle cells and macrophages in the intima of atherosclerotic plaques but not in normal medial smooth muscle cells. The TT allele of galectin-2 was shown to confer protection while the CC allele confers susceptibility to MI (Ozaki et al., 2004).

A key regulator of the inflammatory cascade, the 26S ubiquitin-proteasome system, has also recently been implicated in susceptibility to MI, further emphasizing the role of inflammation in the pathogenesis of MI. A SNP (exon 1-8C/G) located in the 5′ untranslated region of exon 1 of the proteasome subunit α type 6 (PSMA6) gene was identified in a whole-genome case-control association study of Japanese patients as being significantly associated with MI (OR = 1.36) (Ozaki et al., 2006). The exon 1 -8C/G SNP was shown to alter the transcription levels of the PSMA6 gene both *in vitro* and *in vivo*. The PSMA6

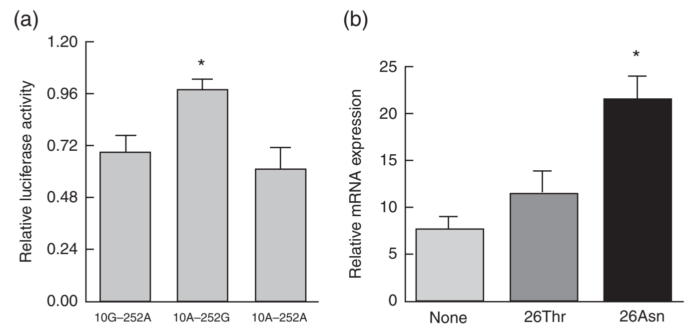


Figure 24.4 Functional SNPs of LTA gene. (a) Transcriptional regulatory activity is affected by the SNP in Intron 1 of LTA (252A → G) as measured by relative luciferase activity (*$p < 0.01$); (b) Induction of adhesion molecules such as VCAM1 is differentially affected by 26Asn-LTA and 26Thr-LTA in HCASMC cultures following treatment with medium only (white) or with 20 ng/ml of 26Thr-LTA (gray) or 26Asn-LTA (black) for 4 h (*$p < 0.01$). (Reproduced with permission from Ozaki et al., 2002.)

gene codes for an α subunit of the 20S proteasome, which is the core particle for the 26S ubiquitin-proteasome system that degrades IκB protein. IκB inhibits the activation of the nuclear factor κB (NF-κB) which regulates the expression of many genes in the inflammatory pathway.

A recent genome-wide association study not limited to candidate genes examined 11,053 SNPs in 6891 genes and found four gene variants that were associated with MI. These gene variants encode a tyrosine kinase (ROS1 [OR 1.75]), the cytoskeletal protein paladin (KIAA0992 [OR 1.40]), and two G protein–coupled receptors (TAS2R50 [OR 1.58] and OR13G1 [OR 1.40]). At this point, the underlying biological mechanism for these associations is unknown and warrants further investigation, which could yield new therapeutic strategies (Shiffman et al., 2005). The same group published a genome-wide

association study for early-onset MI and identified a platelet degranulation gene – VAMP8 – as having an odds ratio of 1.75 in 3 separate cohorts (Shiffman et al., 2006).

Adding to the complexity of the molecular landscape of CAD and MI is the finding of an untranslatable but presumably functional RNA transcript that has been associated with MI in a large case-control GWAS using 52,608 SNPs in a cohort of Japanese patients (Ishii et al., 2006). This functional RNA has been designated the myocardial infarction associated transcript (MIAT), and one SNP in exon 5 was shown to affect transcriptional levels of the gene. The mechanism by which this RNA transcript contributes to susceptibility to MI remains to be identified but examples of other noncoding functional RNA molecules have been demonstrated to influence important transcriptional processes in other models of disease.

The Leukotriene Pathway

More recently, variants of the arachidonate 5-lipoxygenase-activating protein (ALOX5AP) gene have been shown to confer a significant risk for MI. ALOX5AP plays a key role in the inflammatory leukotriene biosynthetic pathway. A genome-wide scan of 296 Icelandic families identified a four-SNP haplotype, called HapA, in the *ALOX5AP* gene that confers a nearly twofold risk of MI and stroke (Helgadottir et al., 2004). This association was further verified in a cohort of British patients with another haplotype, called HapB, of the *ALOX5AP* gene which also conferred a nearly twofold risk of MI. Individuals with these at-risk haplotypes were shown to have a significantly increased neutrophil production of leukotriene B4 (LTB4) supporting the role of this gene in the inflammatory process implicated in atherosclerosis and plaque rupture.

Another gene in the same leukotriene biosynthetic pathway as ALOX5AP has recently been implicated in risk of MI. A study of 1553 Icelandic individuals with history of MI demonstrated that a haplotype, called HapK, of the leukotriene A4 hydrolase (LTA4H) gene was associated with a relative risk of 1.45 ($p = 0.035$) for MI and additional cardiovascular disease (Helgadottir et al., 2006). This association was also confirmed in three separate cohorts from the United States population. The HapK haplotype was also shown to correlate with increased neutrophil production of LTB4. Of further interest, in the United States cohorts tested the relative risk of MI for African-American carriers of the HapK haplotype was much higher (RR = 3.57) although the haplotype was less frequent in this population. This suggests that the LTA4H variant may interact with other gene variants more common in African-Americans, although to date these other gene variants have yet to be determined.

PCSK9 – A Key Player in LDL Cholesterol Receptor Modulation

Genetic variations in the proprotein convertase subtilisin/kexin type 9 serine protease (PCSK9) gene have been shown to reduce LDL cholesterol levels. The results of a massive 15 year longitudinal study of nonsense and missense genetic variants of PCSK9 were recently published (Cohen et al., 2006). The presence of a nonsense mutation (2.6% of the African-American cohort) was associated with a 28% reduction in mean LDL cholesterol and an 88% reduction in risk of cardiovascular events ($p = 0.008$; HR = 0.11) (Figure 24.5). The presence of a missensevariation of PCSK9 was associated with a relatively lower 15% reduction in LDL cholesterol and a 47% reduction in cardiovascular risk ($p = 0.003$; HR = 0.50). Nonsense mutations of PCSK9 were rare (6 of 9537 subjects) in the Caucasian cohort. This study elegantly demonstrates two important concepts. First, ethnic background can be used as a crude marker to identify patients with underlying genetic susceptibilities. Secondly, the benefits conveyed by modification of known conventional risk factors such as LDL cholesterol levels are likely determined by underlying genetic mechanisms and inheritable susceptibilities.

Chromosome 9p21 Identified as a Hotspot for Risk of MI

One of the most exciting developments in the field of cardiovascular genomics is the recent discovery of a strong association of a region of chromosome 9p21 with MI. What makes this discovery so important is that it has been replicated in four separate GWAS spanning multinational cohorts including thousands of cases and controls (Helgadottir et al., 2007; McPherson et al., 2007; Samani et al., 2007; WTCCC, 2007). These studies have shown that over 20% of Caucasians are homozygous for this genetic variation and that it carries a greater than 20–30% risk of MI.

The chromosome 9p21 region of interest contains the genes for two cyclin-dependent kinase inhibitors, *CDKN2A* ($p16^{INK4a}$) and *CDKN2B* ($p15^{INK4b}$). These genes are involved in cell cycle regulation, and *CDKN2B* expression is induced by TGF-β, which may play a role in the pathogenesis of atherosclerosis. Another gene in this region is methylthioadenosine phosphorylase, which is involved in adenine and methionine salvage. Further adding to the excitement surrounding this area is the additional finding of an association of this 9p21 region with type 2 diabetes mellitus, which is a strong conventional risk factor for CAD and MI and suggests a common mechanism of risk for these diseases (Saxena et al., 2007; Scott et al., 2007; Zeggini et al., 2007). However the SNPs which have been associated with CAD and MI are not within these genes, and to date it is not clear what the mechanism is by which these genetic variations might influence susceptibility to CAD and MI.

It is important to note that in each GWAS there were findings of loci of interest that were not replicated in the other studies. This demonstrates the need for cautious interpretation of the findings from a single GWAS and the critical need for replication of findings by independent validation studies. The elucidation of the association between chromosome 9p21 and MI is an elegant example of the degree of scrutiny required for the investigation of other GWAS findings.

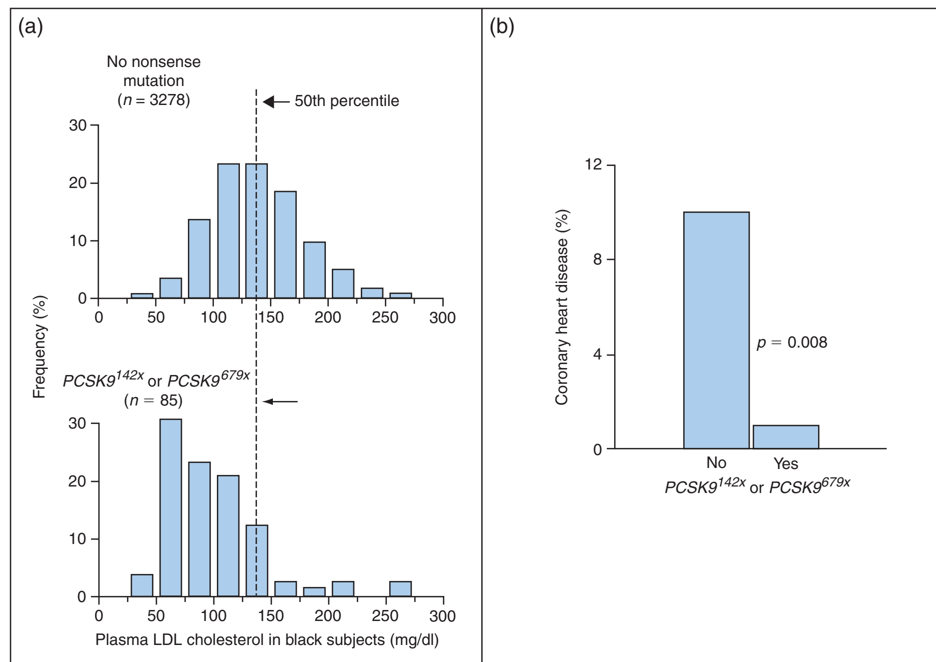


Figure 24.5 Distribution of plasma LDL cholesterol levels (a) and the incidence of coronary heart disease (CHD) (b) among African-American subjects, according to the presence or absence of a $PCSK9^{I42X}$ or $PCSK9^{679X}$ allele. (Reproduced with permission from Cohen et al., 2006.)

Conventional Risk Factors and Environmental Influences

Although genetic factors may play a significant role in the development of CAD and MI, the fact is that 80–90% of patients with CAD have conventional risk factors such as hypertension, hyperlipidemia, diabetes mellitus or cigarette smoking (Khot et al., 2003). Environmental influences such as diet have also been shown to play an important role in the development of CAD as well. Results of the Nurse's Health Study demonstrated *trans*-fat intake was associated with increased risk of CAD, whereas there was an inverse association with polyunsaturated fat intake (Oh et al., 2005).

What remains to be determined is to what extent the effect of these conventional risk factors and environmental influences is dependent upon underlying genetic interactions. It has been shown that increased dietary arachidonic acid significantly enhanced atherogenesis in carriers of variant 5-lipoxygenase genotypes, whereas increased dietary intake of n-3 fatty acids appeared to reduce atherogenesis in this population (Dwyer et al., 2004). As mentioned previously, carriers of nonsense and missense mutation of the PCSK9 gene show significantly reduced levels of LDL cholesterol and enjoy a protective benefit against cardiovascular risk (Cohen et al., 2006). This suggests that susceptibility to conventional risk factors may ultimately depend upon complex genetic and environmental interactions.

SCREENING STRATEGIES

Currently, screening for cardiovascular risk is based on various models that incorporate gender, age, ethnicity and behavioral and disease risk factors. Obviously, risk associated with gender and ethnicity imply a genetic component, but it is likely that the susceptibility to behavioral and disease risk factors is also based on genetic heritability.

Ongoing whole-genome SNP association studies will lay the groundwork for more extensive and definitive identification of SNPs and haplotypes associated with MI. These haplotypes will likely allow us to identify patients at increased risk for CAD and MI before the onset of disease and regardless of exposure to conventional risk factors.

The identification of genetic susceptibility traits will allow for more accurate risk stratification of patients than is achievable with current clinical models. The presence of a particular trait may carry such significant risk as to directly identify patients at risk, as is the case for *BRCA1* gene variant carriers who are at high risk for breast and ovarian cancer. More commonly a genetic trait may be integrated into the risk assessment calculation of more conventional models to improve the accuracy of the risk prediction.

Early identification of patients at risk can prompt risk factor modification and early genetically tailored pharmacological interventions, which could greatly decrease the morbidity and

mortality associated with these diseases. Future models of cardiovascular risk will likely be based on these complex multigenetic and environmental interactions. However, the technical difficulties and financial burden associated with genetic analysis must be overcome before these traits can be economically and efficiently incorporated into every day clinical practice.

Although genetic variation can result in the presence or absence of a particular protein and thereby result in a disease state, it is more common that these variants result in differential expression of a particular protein. For example, a SNP in *LGALS2* has been associated with increased risk of MI (Ozaki et al., 2004). The *LGALS2* SNP encodes for the protein galectin-2 and results in alterations in the transcriptional levels of galectin-2 which binds to LTA. This could result in altered levels of LTA secretion and intravascular inflammation which is thought to be associated with CAD and MI. The expression levels of a number of other proteins, such as C-reactive protein, have also been shown to correlate with risk of MI and it is certain that these expression levels are similarly genetically determined. Findings such as these may lead to the development of protein-based assays, which are faster and more economical than DNA analysis and are more amenable to every day clinical use.

PROGNOSTIC IMPLICATIONS OF MI

Heart Failure and Ventricular Remodeling

Greater than 20% of patients treated for initial MI will develop acute heart failure and the incidence increases with recurrent MI. Post-MI heart failure conveys a three- to sixfold increase in risk of death and is the result of the complex processes of myocardial necrosis and ventricular remodeling. Studies have shown that female gender as well as complex phenotypic traits, such as hypertension and diabetes, is associated with increased risk of post-MI heart failure, which suggests a genetic component to the process. The incidence of post-MI heart failure has been shown to decrease dramatically 8 days after MI (Thomas et al., 2005). Identification of the early molecular mechanisms responsible for the development of post-MI heart failure may yield targets for future therapies.

Risk for Repeat Myocardial Infarction

A recent analysis of atherosclerosis-related gene SNPs in 1586 Japanese patients from the Osaka Acute Coronary Insufficiency Study revealed that G allele carriers at the position 252 of the LTA gene were independently associated with an increased risk of death after acute MI compared to noncarriers who also suffered acute MI as well as recurrent MI when (hazard ratio = 2.46; 95% CI = 1.24–4.86) (Mizuno et al., 2005). This study would suggest that although risk for MI may be a function of multiple genetic polymorphisms, the presence of a select group of polymorphisms may confer an even higher risk for carriers. A better understanding of the functional influences of these polymorphisms as they are discovered will allow for more accurate risk stratification.

Response to Post-MI Medical Therapy

Administration of β-blocker therapy post-MI has been shown to reduce infarct size and mortality, however all patients may not derive equivalent benefit. A recent prospective study of response to β-blocker therapy in patients with two common polymorphisms of the β 2-adrenergic receptor demonstrated a significant difference in 3-year survival which was genotype specific (Lanfear et al., 2005). Similar results have previously been shown for risk of incident coronary events (Heckbert et al., 2003). Genetic markers such as these may allow us to discriminate between treatment responders and nonresponders and should allow for individualized therapies as well as better risk stratification.

PHARMACOGENOMICS OF MI

HMG-CoA Reductase Inhibitors

HMG-CoA reductase inhibitors, commonly referred to as "statins", reliably lower atherogenic lipoproteins and have been shown to decrease the incidence of MI in numerous studies (Davidson, 2005; Schwartz et al., 2005). These drugs have also been shown to have significant anti-inflammatory effect and lower serum levels of inflammatory markers such as C-reactive protein (Ridker et al., 1999). These effects appear to correlate with a reduction in the rate of atherosclerotic plaque progression and possibly even plaque regression (Ridker et al., 2005; Nissen et al., 2005; Nissen et al., 2006).

Studies have also suggested an effect of statin therapy on ventricular remodeling post-MI. Simvastatin was shown to reduce the incidence of heart failure in patients with CAD in a retrospective analysis of the Scandinavian Simvastatin Survival Study (4S) trial (Kjekshus et al., 1997). However, this finding was not replicated with the use of pravastatin in the CARE study (Lewis et al., 2003).

Although overall significant benefit has been demonstrated with the use of statin therapy, considerable interindividual variation exists in the response to statin therapy. It is likely that genetic factors are responsible for this variation in individual response. Multiple genes have been implicated as potential modulators of statin response but few of the findings have been replicated (Kajinami et al., 2004). Identification of genetic haplotypes that correlate with increased response to statin therapy will allow responsive individuals to be targeted for therapy while nonresponders may avoid the side effects of a nonbeneficial and costly therapy.

Antiplatelet Therapies

Aspirin Resistance

Aspirin irreversibly binds cyclooxygenase, which plays a key role in the process of platelet aggregation. The benefit of aspirin antiplatelet therapy for the reduction of all-cause mortality, nonfatal MI and recurrent MI has been well established (Antiplatelet Trialists Collaboration, 2002). However, not all patients appear to receive equal benefit from aspirin therapy and have been

classified as aspirin resistant. Several mechanisms of aspirin resistance have been proposed but one of the most studied mechanisms involves polymorphisms of the IIIa subunit of the platelet glycoprotein IIb–IIIa receptor, which is the final common pathway in platelet aggregation. The PLA2 polymorphism of the glycoprotein IIb–IIIa receptor is present in over 30% of patients with CAD. Heterozygous carriers of the PLA2 allele have also been shown to be aspirin resistant (Cooke et al., 2006; Hanjis et al., 2006). This polymorphism may represent a genetic component for aspirin resistance, and there are likely others that may affect other steps in the platelet activation cascade. Recently a significant association with a synonymous SNP in the platelet gene for the adenosine 5-diphosphate (ADP) receptor P2Y1 was found in an aspirin-resistant population (Jefferson et al., 2005). Identification of patients with gene-mediated aspirin resistance could further reduce the incidence of MI by the use of different antiplatelet therapies in this population.

Clopidogrel Resistance

One of the major advances in the treatment of acute MI has been the use of the ADP receptor antagonist clopidogrel. ADP is released by activated platelets and by binding to the platelet $P2X_1$, $P2Y_1$ and $P2Y_{12}$ receptors greatly accelerates platelet aggregation. Clopidogrel irreversibly binds the $P2Y_{12}$ receptor and blunts ADP-mediated platelet aggregation. However, just as is the case with aspirin therapy, there is significant interpatient variability in the response to clopidogrel. One suggested mechanism for this difference in response is polymorphisms of the $P2Y_{12}$ receptor gene. Although the known mutations of this gene have been shown to result in congenital bleeding disorders, one study has identified a polymorphism in healthy subjects that was associated with increased platelet response to ADP stimulation as well as peripheral artery disease (Fontana et al., 2003). Also, clopidogrel is a prodrug that requires activation by the hepatic cytochrome P3A4 and cytochrome P3A5 enzymes. Numerous SNPs have been identified in these genes, which could affect clopidogrel metabolism and lead to decreased activation. Polymorphisms of the $P2Y_{12}$ receptor and cytochrome P3A genes could account for some of the observed resistance to clopidogrel (Nguyen et al., 2005).

NOVEL AND EMERGING THERAPIES

Drug-Eluting Intracoronary Stents

Currently there is some debate regarding the use of drug-eluting stents for the treatment of acute MI (Saia et al., 2003; Valgimigli et al., 2005); however numerous studies have shown a significant reduction in the rate of in-stent restenosis with the introduction of this technology (Bavry et al., 2005; Hill et al., 2004). These stents are coated with immunosuppressant or cytotoxic agents that reduce the incidence of restenosis. Currently, the two most commonly used agents in the United States are sirolimus and paclitaxel, but a variety of other agents are currently in development.

Sirolimus is a cytostatic macrolide antibiotic which inhibits smooth muscle proliferation by halting cell cycle progression. Smooth muscle proliferation is a key component of the neointimal proliferation process which leads to restenosis. Paclitaxel is a cytotoxic agent which disrupts microtubule assembly and kills proliferating or dividing cells. In addition, these drugs may also inhibit apoptosis of the existing smooth muscle cells and contribute to the stabilization of vulnerable plaque (Faber et al., 2002; Smith et al., 2003). As more is discovered about the molecular mechanisms that promote restenosis, novel therapeutic agents may be developed to more specifically inhibit neointimal proliferation while preserving the cellular mechanisms of healing and plaque stabilization.

Leukotriene Inhibitors

Variants in the *ALOX5AP* gene and the *LTA4H* gene have been shown to be associated with risk of MI (Helgadottir et al., 2004, 2006). Recently, in a randomized trial of 191 patients at risk for MI with *ALOX5AP* or *LTA4H* gene "gain-of-function" haplotypes, an ALOX5AP inhibitor was studied for its effects on common biomarkers associated with MI risk. A 4-week trial of the ALOX5AP inhibitor resulted in a significant and dose-dependent reduction in serum C-reactive protein and amyloid A. Also, the use of the ALOX5AP inhibitor attenuated the generation of leukotriene B4 and myeloperoxidase by activated neutrophils and the urinary metabolite of LTB4. However, there was a significant 9% increase in serum plasma Lp-PLA2 levels with a corresponding 8% increase in LDL cholesterol noted at the highest dose of the ALOX5AP inhibitor. Currently, a clinical trial is underway to determine if the use of this ALOX5AP inhibitor will affect the overall risk of MI in this at-risk patient population (Hakonarson et al., 2005). A study of dietary modification in over 400 patients with ALOX5AP gene promoter polymorphisms demonstrated that increased intake of n-3 fatty acids appeared to reduce atherogenesis in this population (Dwyer et al., 2004). This suggests that in addition to gene targeted pharmacological treatment, gene targeted behavioral modification strategies may yield further reduction in risk of MI.

2009 UPDATE

The number of SNPs associated with CHD and MI continues to grow at a steady pace. Also, SNPs associated with traditional risk factors for CHD such as dyslipidemia are increasingly being identified and replicated by large GWAS (Aulchenko et al., 2009; Kathiresan et al., 2008, 2009; Kooner et al., 2008; Sabatti et al., 2009). Variants associated with increased LDL cholesterol have been shown to convey increased risk of CHD, whereas variants associated with favorable lipid profiles, such as the null mutation variant (R19X) of the apoC-III gene (APOC3), have been shown to be cardioprotective (Pollin et al., 2008; Willer et al., 2008).

The association of SNPs within chromosome 9p21 with CHD and MI were described in 2007 (Helgadottir et al., 2007; McPherson et al., 2007; Samani et al., 2007; Wellcome Trust Case Control Consortium, 2007). Further studies have confirmed these associations across diverse ethnic populations, and a high-risk haplotype has been identified (Abdullah et al., 2008; Anderson et al, 2008; Hiura et al., 2008; Shen et al., 2008; Zhou et al., 2008). These findings have led to enthusiasm that a new susceptibility gene or genetic mechanism for CHD and MI may ultimately be elucidated. The 9p21 risk variants have also been associated with abdominal aortic aneurysm and intracranial aneurysm, suggesting a vascular phenotype that does not necessarily invoke atherosclerosis (Helgadottir et al., 2008); however, the specific role of genes near the 9p21 region in disease susceptibility has not yet been established. Of interest, the high-risk haplotype identified in this region does localize with a large antisense non-coding RNA gene known as ANRIL (Broadbent et al., 2008). It has been postulated that ANRIL may exert its effect via transcriptional regulation of neighboring genes or yet to be identified gene(s) (Hamsten and Eriksson, 2008).

Another allele of interest has been the Trp719Arg polymorphism of the kinesin-like protein 6 (KIF6 gene). Kinesins are motor proteins involved in intracellular transport and their role in MI and CHD are unknown. A genetic analysis of patients enrolled in the CARE and WOSCOPS studies revealed that carriers of the 719Arg allele had a 50% increased risk of CHD or recurrent MI compared to non-carriers (Iakoubova et al., 2008b). This allele was also shown to be associated with increased risk of CHD in patients from the ARIC study and the Women's Health Study (Morrison et al., 2007; Shiffman et al., 2008).

It has been recently shown that the benefit derived from standard statin therapy in reduction of cardiac events in the CARE study and of CHD in the WOSCOPS study was primarily confined to carriers of the KIF6 719Arg allele (Iakoubova et al., 2008b). In addition, the 719Arg carriers were shown to account for the benefits seen from intensive lipid lowering therapy in the PROVE IT – TIMI 22 study (Iakoubova et al., 2008a). Importantly, non-carriers of the 719Arg allele were shown not only to be at lower risk for MI but also do not appear to receive significant benefit from statin therapy. These findings suggest that genetically tailored therapies for heart disease may be approaching, thereby allowing for better allocation of limited resources to responders and limiting potentially harmful side effects in non-responders.

2009 UPDATE REFERENCES

Abdullah, K.G., Li, L., Shen, G.Q., Hu, Y., Yang, Y., MacKinlay, K.G., Topol, E.J. and Wang, Q.K. (2008). Four SNPS on chromosome 9p21 confer risk to premature, familial CAD and MI in an American caucasian population (GeneQuest). *Ann Hum Genet* 72(5), 654–657.

Anderson, J.L., Horne, B.D., Kolek, M.J., Muhlestein, J.B., Mower, C.P., Park, J.J., May, H.T., Camp, N.J. and Carlquist, J.F. (2008). Genetic variation at the 9p21 locus predicts angiographic coronary artery disease prevalence but not extent and has clinical utility. *Am Heart J* 156(6), 1155–1162.

Aulchenko, Y.S., Ripatti, S., Lindqvist, I., Boomsma, D., Heid, I.M., Pramstaller, P.P., Penninx, B.W., Janssens, A.C., Wilson, J.F., Spector, T. et al. (2009). Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet* 41(1), 47–55.

Broadbent, H.M., Peden, J.F., Lorkowski, S., Goel, A., Ongen, H., Green, F., Clarke, R., Collins, R., Franzosi, M.G., Tognoni, G. et al. (2008). Susceptibility to coronary artery disease and diabetes is encoded by distinct, tightly linked, SNPs in the ANRIL locus on chromosome 9. *Hum Mol Genet* 17(6), 806–814.

Hamsten, A. and Eriksson, P. (2008). Identifying the susceptibility genes for coronary artery disease: From hyperbole through doubt to cautious optimism. *J Intern Med* 263(5), 538–552.

Helgadottir, A., Thorleifsson, G., Magnusson, K.P., Grétarsdottir, S., Steinthorsdottir, V., Manolescu, A., Jones, G.T., Rinkel, G.J., Blankensteijn, J.D., Ronkainen, A. et al. (2008). The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm. Nat Genet 40(2), 217–24.

Helgadottir, A., Thorleifsson, G., Manolescu, A., Gretarsdottir, S., Blondal, T., Jonasdottir, A., Jonasdottir, A., Sigurdsson, A., Baker, A., Palsson, A. et al. (2007). A common variant on chromosome 9p21 affects risk of myocardial infarction. *Science* 316 (5843), 1491–1493.

Hiura, Y., Fukushima, Y., Yuno, M., Sawamura, H., Kokubo, Y., Okamura, T., Tomoike, H., Goto, Y., Nonogi, H., Takahashi, R. et al. (2008). Validation of the association of genetic variants on

chromosome 9p21 and 1q41 with myocardial infarction in a Japanese population. *Circ J* 72(8), 1213–1217.

Iakoubova, O.A., Sabatine, M.S., Rowland, C.M., Tong, C.H., Catanese, J.J., Ranade, K., Simonsen, K.L., Kirchgessner, T.G., Cannon, C.P., Devlin, J.J. et al. (2008a). Polymorphism in KIF6 gene and benefit from statins after acute coronary syndromes: results from the PROVE IT-TIMI 22 study. *J Am Coll Cardiol* 51(4), 449–455.

Iakoubova, O.A., Tong, C.H., Rowland, C.M., Kirchgessner, T.G., Young, B.A., Arellano, A.R., Shiffman, D., Sabatine, M.S., Campos, H., Packard, C.J. et al. (2008b). Association of the Trp719Arg polymorphism in kinesin-like protein 6 with myocardial infarction and coronary heart disease in 2 prospective trials: the CARE and WOSCOPS trials. *J Am Coll Cardiol* 51(4), 435–443.

Kathiresan, S., Melander, O., Guiducci, C., Surti, A., Burtt, N.P., Rieder, M.J., Cooper, G.M., Roos, C., Voight, B.F., Havulinna, A.S. et al. (2008). Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet* 40(2), 189–197.

Kathiresan, S., Willer, C.J., Peloso, G.M., Demissie, S., Musunuru, K., Schadt, E.E., Kaplan, L., Bennett, D., Li, Y., Tanaka, T. et al. (2009). Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet* 41(1), 56–65.

Kooner, J.S., Chambers, J.C., Aguilar-Salinas, C.A., Hinds, D.A., Hyde, C.L., Warnes, G.R., Gómez Pérez, F.J., Frazer, K.A., Elliott, P., Scott, J. et al. (2008). Genome-wide scan identifies variation in the MLXIPL associated with plasma triglycerides. *Nat Genet* 40(2), 149–151.

McPherson, R., Pertsemlidis, A., Kavaslar, N., Stewart, A., Roberts, R., Cox, D.R., Hinds, D.A., Pennacchio, L.A., Tybjaerg-Hansen, A., Folsom, A.R. et al. (2007). A common allele on chromosome 9 associated with coronary heart disease. *Science* 316 (5830), 1488–1491.

Morrison, A.C., Bare, L.A., Chambless, L.E., Ellis, S.G., Malloy, M., Kane, J.P., Pankow, J.S., Devlin, J.J., Willerson, J.T. and Boerwinkle, E. (2007). Prediction of coronary heart disease risk using a genetic risk score: The Atherosclerosis risk in communities study. *Am J Epidemiol* 166(1), 28–35.

Pollin, T.I., Damcott, C.M., Shen, H., Ott, S.H., Shelton, J., Horenstein, R.B., Post, W., McLenithan, J.C., Bielak, L.F., Peyser, P.A. et al. (2008). A null mutation in human APOC3 confers a favorable plasma lipid profile and apparent cardioprotection. *Science* 322 (5908), 1702–1705.

Sabatti, C., Service, S.K., Hartikainen, A.L., Pouta, A., Ripatti, S., Brodsky, J., Jones, C.G., Zaitlen, N.A., Varilo, T., Kaakinen, M. et al. (2009). Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat Genet* 41(1), 35–46.

Samani, N.J., Erdmann, J., Hall, A.S., Hengstenberg, C., Mangino, M., Mayer, B., Dixon, R.J., Meitinger, T., Braund, P., Wichmann, H.E. et al. (2007). Genome-wide association analysis of coronary artery disease. *N Engl J Med* 357(5), 443–453.

Shen, G.Q., Li, L., Rao, S., Abdullah, K.G., Ban, J.M., Lee, B.S., Park, J.E. and Wang, Q.K. (2008). Four SNPs on chromosome 9p21 in a South Korean population implicate a genetic locus that confers high cross-race risk for development of coronary artery disease. *Arterioscler Thromb Vasc Biol* 28(2), 360–365.

Shiffman, D., Chasman, D.I., Zee, R.Y., Iakoubova, O.A., Louie, J.Z., Devlin, J.J. and Ridker, P.M. (2008). A kinesin family member 6(KIF6) variant is associated with coronary heart disease in the women's health study. *J Am Coll Cardiol* 51(4), 444–448.

Wellcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447(7145), 661–678.

Willer, C.J., Sanna, S., Jackson, A.U., Scuteri, A., Bonnycastle, L.L., Clarke, R., Heath, S.C., Timpson, N.J., Najjar, S.S., Stringham, H.M. et al. (2008). Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* 40(2), 161–169.

Zhou, L., Zhang, X., He, M., Cheng, L., Chen, Y., Hu, F.B. and Wu, T. (2008). Risk of coronary heart disease in Chinese Han population associations between single nucleotide polymorphisms on chromosome 9p21. *Arterioscler Thromb Vasc Biol* 28(11), 2085–2089.

REFERENCES

Antiplatelet Trialists' Collaboration. (2002). Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. *BMJ*, 324 (7329), 71–86. (2004). A trio family study showing association of the lymphotoxin-alpha N26 (804A) allele with coronary artery disease. *Eur J Hum Genet* 12(9), 770–774.

Bavry, A.A., Kumbhani, D.J., Helton, T.J. and Bhatt, D.L. (2005). What is the risk of stent thrombosis associated with the use of paclitaxel-eluting stents for percutaneous coronary intervention? A meta-analysis. *J Am Coll Cardiol* 45(6), 941–946.

Bhagavatula, M.R., Fan, C., Shen, G.Q., Cassano, J., Plow, E.F., Topol, E.J. and Wang, Q. (2004). Transcription factor MEF2A mutations in patients with coronary artery disease. *Hum Mol Genet* 13(24), 3181–3188.

Boekholdt, S.M., Trip, M.D., Peters, R.J., Engelen, M., Boer, J.M., Feskens, E.J., Zwinderman, A.H., Kastelein, J.J. and Reitsma, P.H. (2002). Thrombospondin-2 polymorphism is associated with a reduced risk of premature myocardial infarction. *Arterioscler Thromb Vasc Biol* 22(12), e24–e27.

Boerma, M., Forsberg, L., Van Zeijl, L., Morgenstern, R., De Faire, U., Lemne, C., Erlinge, D., Thulin, T., Hong, Y. and Cotgreave, I.A. (1999). A genetic polymorphism in connexin 37 as a prognostic marker for atherosclerotic plaque development. *J Intern Med* 246(2), 211–218.

Bonow, R.O., Smaha, L.A., Smith, S.C., Jr, Mensah, G.A. and Lenfant, C. (2002). World Heart Day 2002: The international burden of cardiovascular disease: Responding to the emerging global epidemic. *Circulation* 106(13), 1602–1605.

Bowden, D.W., Rudock, M., Ziegler, J., Lehtinen, A.B., Xu, J., Wagenknecht, L.E., Herrington, D., Rich, S.S., Freedman, B.I., Carr, J.J. et al. (2006). Coincident linkage of type 2 diabetes, metabolic syndrome, and measures of cardiovascular disease in a genome scan of the diabetes heart study. *Diabetes* 55(7), 1985–1994.

Broeckel, U., Hengstenberg, C., Mayer, B., Holmer, S., Martin, L.J., Comuzzie, A.G., Blangero, J., Nurnberg, P., Reis, A., Riegger, G.A. et al. (2002). A comprehensive linkage analysis for myocardial infarction and its related risk factors. *Nat Genet* 30(2), 210–214.

Cohen, J.C., Boerwinkle, E., Mosley, T.H., Jr, and Hobbs, H.H. (2006). Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N Engl J Med* 354(12), 1264–1272.

Connelly, J.J., Wang, T., Cox, J.E., Haynes, C., Wang, L., Shah, S.H., Crosslin, D.R., Hale, A.B., Nelson, S., Crossman, D.C. et al. (2006). GATA2 is associated with familial early-onset coronary artery disease. *PLoS Genet* 2(8), e139.

Cooke, G.E., Liu-Stratton, Y., Ferketich, A.K., Moeschberger, M.L., Frid, D.J., Magorien, R.D., Bray, P.F., Binkley, P.F. and Goldschmidt-Clermont, P.J. (2006). Effect of platelet antigen polymorphism on platelet inhibition by aspirin, clopidogrel, or their combination. *J Am Coll Cardiol* 47(3), 541–546.

Davidson, M.H. (2005). Reducing residual risk for patients on statin therapy: The potential role of combination therapy. *Am J Cardiol* 96(9A), 3K–13K, discussion 34K-35K.

Dwyer, J.H., Allayee, H., Dwyer, K.M., Fan, J., Wu, H., Mar, R., Lusis, A.J. and Mehrabian, M. (2004). Arachidonate 5-lipoxygenase promoter genotype, dietary arachidonic acid, and atherosclerosis. *N Engl J Med* 350(1), 29–37.

Faber, B.C., Heeneman, S., Daemen, M.J. and Cleutjens, K.B. (2002). Genes potentially involved in plaque rupture. *Curr Opin Lipidol* 13(5), 545–552.

Fischer, M., Broeckel, U., Holmer, S., Baessler, A., Hengstenberg, C., Mayer, B., Erdmann, J., Klein, G., Riegger, G., Jacob, H.J. et al. (2005). Distinct heritable patterns of angiographic coronary artery disease in families with myocardial infarction. *Circulation* 111(7), 855–862.

Fontana, P., Dupont, A., Gandrille, S., Bachelot-Loza, C., Reny, J.L., Aiach, M. and Gaussem, P. (2003). Adenosine diphosphate-induced platelet aggregation is associated with P2Y12 gene sequence variations in healthy subjects. *Circulation* 108(8), 989–995.

Francke, S., Manraj, M., Lacquemant, C., Lecoeur, C., Lepretre, F., Passa, P., Hebe, A., Corset, L., Yan, S.L., Lahmidi, S. et al. (2001). A genome-wide scan for coronary heart disease suggests in Indo-Mauritians a susceptibility locus on chromosome 16p13 and replicates linkage with the metabolic syndrome on 3q27. *Hum Mol Genet* 10(24), 2751–2765.

George, S.J., Baker, A.H., Angelini, G.D. and Newby, A.C. (1998a). Gene transfer of tissue inhibitor of metalloproteinase-2 inhibits metalloproteinase activity and neointima formation in human saphenous veins. *Gene Ther* 5(11), 1552–1560.

George, S.J., Johnson, J.L., Angelini, G.D., Newby, A.C. and Baker, A.H. (1998b). Adenovirus-mediated gene transfer of the human TIMP-1 gene inhibits smooth muscle cell migration and neointimal formation in human saphenous vein. *Hum Gene Ther* 9(6), 867–877.

Hakonarson, H., Thorvaldsson, S., Helgadottir, A., Gudbjartsson, D., Zink, F., Andresdottir, M., Manolescu, A., Arnar, D.O., Andersen, K., Sigurdsson, A. et al. (2005). Effects of a 5-lipoxygenase-activating protein inhibitor on biomarkers associated with risk of myocardial infarction: A randomized trial. *JAMA* 293(18), 2245–2256.

Hanjis, C., W. Frishman, H. and Lerner, R.G. (2006). Aspirin resistance: Mechanisms and clinical implications. *Cardiol Rev* 14(1), 18–25.

Harrap, S.B., Zammit, K.S., Wong, Z.Y., Williams, F.M., Bahlo, M., Tonkin, A.M. and Anderson, S.T. (2002). Genome-wide linkage analysis of the acute coronary syndrome suggests a locus on chromosome 2. *Arterioscler Thromb Vasc Biol* 22(5), 874–878.

Hauser, E.R., Crossman, D.C., Granger, C.B., Haines, J.L., Jones, C.J., Mooser, V., McAdam, B., Winkelmann, B.R., Wiseman, A.H., Muhlestein, J.B. et al. (2004). A genomewide scan for early-onset coronary artery disease in 438 families: The GENECARD Study. *Am J Hum Genet* 75(3), 436–447.

Heckbert, S.R., Hindorff, L.A., Edwards, K.L., Psaty, B.M., Lumley, T., Siscovick, D.S., Tang, Z., Durda, J.P., Kronmal, R.A. and Tracy, R.P. (2003). Beta2-adrenergic receptor polymorphisms and risk of incident cardiovascular events in the elderly. *Circulation* 107(15), 2021–2024.

Helgadottir, A., Manolescu, A., Thorleifsson, G., Gretarsdottir, S., Jonsdottir, H., Thorsteinsdottir, U., Samani, N.J., Gudmundsson, G., Grant, S.F., Thorgeirsson, G. et al. (2004). The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke. *Nat Genet* 36(3), 233–239.

Helgadottir, A., Manolescu, A., Helgason, A., Thorleifsson, G., Thorsteinsdottir, U., Gudbjartsson, D.F., Gretarsdottir, S., Magnusson, K.P., Gudmundsson, G., Hicks, A. et al. (2006). A variant of the gene encoding leukotriene A4 hydrolase confers ethnicity-specific risk of myocardial infarction. *Nat Genet* 38(1), 68–74.

Helgadottir, A., Thorleifsson, G., Manolescu, A., Gretarsdottir, S., Blondal, T., Jonasdottir, A., Jonasdottir, A., Sigurdsson, A., Baker, A., Palsson, A. et al. (2007). A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* 316(5830), 1491–1493.

Hill, R.A., Dundar, Y., Bakhai, A., Dickson, R. and Walley, T. (2004). Drug-eluting stents: An early systematic review to inform policy. *Eur Heart J* 25(11), 902–919.

Ishii, N., Ozaki, K., Sato, H., Mizuno, H., Saito, S., Takahashi, A., Miyamoto, Y., Ikegawa, S., Kamatani, N., Hori, M. et al. (2006). Identification of a novel non-coding RNA, MIAT, that confers risk of myocardial infarction. *J Hum Genet* 51(12), 1087–1099.

Jefferson, B.K., Foster, J.H., McCarthy, J.J., Ginsburg, G., Parker, A., Kottke-Marchant, K. and Topol, E.J. (2005). Aspirin resistance and a single gene. *Am J Cardiol* 95(6), 805–808.

Kajinami, K., Takekoshi, N., Brousseau, M.E. and Schaefer, E.J. (2004). Pharmacogenetics of HMG-CoA reductase inhibitors: Exploring the potential for genotype-based individualization of coronary heart disease management. *Atherosclerosis* 177(2), 219–234.

Khot, U.N., Khot, M.B., Bajzer, C.T., Sapp, S.K., Ohman, E.M., Brener, S.J., Ellis, S.G., Lincoff, A.M. and Topol, E.J. (2003). Prevalence of conventional risk factors in patients with coronary heart disease. *JAMA* 290(7), 898–904.

Kjekshus, J., Pedersen, T.R., Olsson, A.G., Faergeman, O. and Pyorala, K. (1997). The effects of simvastatin on the incidence of heart failure in patients with coronary heart disease. *J Card Fail* 3(4), 249–254.

Lanfear, D.E., Jones, P.G., Marsh, S., Cresci, S., McLeod, H.L. and Spertus, J.A. (2005). Beta2-adrenergic receptor genotype and survival among patients receiving beta-blocker therapy after an acute coronary syndrome. *JAMA* 294(12), 1526–1533.

Lewis, E.F., Moye, L.A., Rouleau, J.L., Sacks, F.M., Arnold, J.M., Warnica, J.W., Flaker, G.C., Braunwald, E. and Pfeffer, M.A. (2003). Predictors of late development of heart failure in stable survivors of myocardial infarction: The CARE study. *J Am Coll Cardiol* 42(8), 1446–1453.

Lichtenstein, P., De Faire, U., Floderus, B., Svartengren, M., Svedberg, P. and Pedersen, N.L. (2002). The Swedish Twin Registry: A unique resource for clinical, epidemiological and genetic studies. *J Intern Med* 252(3), 184–205.

Marenberg, M.E., Risch, N., Berkman, L.F., Floderus, B. and de Faire, U. (1994). Genetic susceptibility to death from coronary heart disease in a study of twins. *N Engl J Med* 330(15), 1041–1046.

McCarthy, J.J., Parker, A., Salem, R., Moliterno, D.J., Wang, Q., Plow, E.F., Rao, S., Shen, G., Rogers, W.J., Newby, L.K. et al. (2004). Large scale association analysis for identification of genes underlying premature coronary heart disease: Cumulative perspective from analysis of 111 candidate genes. *J Med Genet* 41(5), 334–341.

Meyer, G.P., Wollert, K.C., Lotz, J., Steffens, J., Lippolt, P., Fichtner, S., Hecker, H., Schaefer, A., Arseniev, L., Hertenstein, B. et al. (2006). Intracoronary bone marrow cell transfer after myocardial infarction: Eighteen months' follow-up data from the randomized, controlled BOOST (BOne marrOw transfer to enhance ST-elevation infarct regeneration) trial. *Circulation* 113(10), 1287–1294.

Mizuno, H., Sato, H., Sakata, Y., Ohnishi, Y., Hishida, E., Kinjo, K., Nakatani, D., Shimizu, M., Kondo, H., Tanaka, T. et al. (2005). Impact of atherosclerosis-related gene polymorphisms on mortality and recurrent events after myocardial infarction. *Atherosclerosis*.

Narizhneva, N.V., Byers-Ward, V.J., Quinn, M.J., Zidar, F.J., Plow, E.F., Topol, E.J. and Byzova, T.V. (2004). Molecular and functional differences induced in thrombospondin-1 by the single nucleotide polymorphism associated with the risk of premature, familial myocardial infarction. *J Biol Chem* 279(20), 21651–21657.

Nguyen, T.A., Diodati, J.G. and Pharand, C. (2005). Resistance to clopidogrel: A review of the evidence. *J Am Coll Cardiol* 45(8), 1157–1164.

Nissen, S.E., Tuzcu, E.M., Schoenhagen, P., Crowe, T., Sasiela, W.J., Tsai, J., Orazem, J., Magorien, R.D., O'Shaughnessy, C. and Ganz, P. (2005). Statin therapy, LDL cholesterol, C-reactive protein, and coronary artery disease. *N Engl J Med* 352(1), 29–38.

Nissen, S.E., Nicholls, S.J., Sipahi, I., Libby, P., Raichlen, J.S., Ballantyne, C.M., Davignon, J., Erbel, R., Fruchart, J.C., Tardif, J.C. et al. (2006). Effect of very high-intensity statin therapy on regression of coronary atherosclerosis: The ASTEROID trial. *JAMA* 295(13), 1556–1565.

Oh, K., Hu, F.B., Manson, J.E., Stampfer, M.J. and Willett, W.C. (2005). Dietary fat intake and risk of coronary heart disease in women: 20 years of follow-up of the nurses' health study. *Am J Epidemiol* 161(7), 672–679.

Ozaki, K., Ohnishi, Y., Iida, A., Sekine, A., Yamada, R., Tsunoda, T., Sato, H., Hori, M., Nakamura, Y. et al. (2002). Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nat Genet* 32(4), 650–654.

Ozaki, K., Inoue, K., Sato, H., Iida, A., Ohnishi, Y., Sekine, A., Sato, H., Odashiro, K., Nobuyoshi, M., Hori, M. et al. (2004). Functional variation in LGALS2 confers risk of myocardial infarction and regulates lymphotoxin-alpha secretion in vitro. *Nature* 429(6987), 72–75.

Ozaki, K., Sato, H., Iida, A., Mizuno, H., Nakamura, T., Miyamoto, Y., Takahashi, A., Tsunoda, T., Ikegawa, S., Kamatani, N. et al. (2006). A functional SNP in PSMA6 confers risk of myocardial infarction in the Japanese population. *Nat Genet* 38(8), 921–925.

Pajukanta, P., Cargill, M., Viitanen, L., Nuotio, I., Kareinen, A., Perola, M., Terwilliger, J.D., Kempas, E., Daly, M., Lilja, H. et al. (2000). Two loci on chromosomes 2 and X for premature coronary heart disease identified in early- and late-settlement populations of Finland. *Am J Hum Genet* 67(6), 1481–1493.

Pluskota, E., Stenina, O.I., Krukovets, I., Szpak, D., Topol, E.J. and Plow, E.F. (2005). Mechanism and effect of thrombospondin-4 polymorphisms on neutrophil function. *Blood* 106(12), 3970–3978.

Porto, I., Leone, A.M., Crea, F. and Andreotti, F. (2005). Inflammation, genetics, and ischemic heart disease: Focus on the major histocompatibility complex (MHC) genes. *Cytokine* 29(5), 187–196.

Ridker, P.M., Rifai, N., Pfeffer, M.A., Sacks, F. and Braunwald, E. (1999). Long-term effects of pravastatin on plasma concentration of C-reactive protein. The Cholesterol and Recurrent Events (CARE) Investigators. *Circulation* 100(3), 230–235.

Ridker, P.M., Cannon, C.P., Morrow, D., Rifai, N., Rose, L.M., McCabe, C.H., Pfeffer, M.A. and Braunwald, E. (2005). C-reactive protein levels and outcomes after statin therapy. *N Engl J Med* 352(1), 20–28.

Saia, F., Lemos, P.A., Lee, C.H., Arampatzis, C.A., Hoye, A., Degertekin, M., Tanabe, K., Sianos, G., Smits, P.C., McFadden, E. et al. (2003). Sirolimus-eluting stent implantation in ST-elevation acute myocardial infarction: A clinical and angiographic study. *Circulation* 108(16), 1927–1929.

Samani, N.J., Erdmann, J., Hall, A.S., Hengstenberg, C., Mangino, M., Mayer, B., Dixon, R.J., Meitinger, T., Braund, P., Wichmann, H.E. et al. (2007). Genomewide association analysis of coronary artery disease. *N Engl J Med* 357(5), 443–453.

Saxena, R., Voight, B.F., Lyssenko, V., Burtt, N.P., de Bakker, P.I., Chen, H., Roix, J.J., Kathiresan, S., Hirschhorn, J.N., Daly, M.J. et al. (2007). Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316(5829), 1331–1336.

Schwartz, G.G. and Olsson, A.G. (2005). The case for intensive statin therapy after acute coronary syndromes. *Am J Cardiol* 96(5A), 45F–53F.

Scott, L.J., Mohlke, K.L., Bonnycastle, L.L., Willer, C.J., Li, Y., Duren, W.L., Erdos, M.R., Stringham, H.M., Chines, P.S., Jackson, A.U. et al. (2007). A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316(5829), 1341–1345.

Shiffman, D., Ellis, S.G., Rowland, C.M., Malloy, M.J., Luke, M.M., Iakoubova, O.A., Pullinger, C.R., Cassano, J., Aouizerat, B.E., Fenwick, R.G. et al. (2005). Identification of four gene variants associated with myocardial infarction. *Am J Hum Genet* 77(4), 596–605.

Shiffman, D., Rowland, C.M., Louie, J.Z., Luke, M.M., Bare, L.A., Bolonick, J.I., Young, B.A., Catanese, J.J., Stiggins, C.F., Pullinger, C.R. et al. (2006). Gene variants of VAMP8 and HNRPUL1 are associated with early-onset myocardial infarction. *Arterioscler Thromb Vasc Biol* 26(7), 1613–1618.

Smith, E.J. and Rothman, M.T. (2003). Antiproliferative coatings for the treatment of coronary heart disease. What are the targets and which are the tools? *J Interv Cardiol* 16(6), 475–483.

Stenina, O.I., Desai, S.Y., Krukovets, I., Kight, K., Janigro, D., Topol, E.J. and Plow, E.F. (2003). Thrombospondin-4 and its variants: Expression and differential effects on endothelial cells. *Circulation* 108(12), 1514–1519.

Thom, T., Haase, N., Rosamond, W., Howard, V.J., Rumsfeld, J., Manolio, T., Zheng, Z.J., Flegal, K., O'Donnell, C., Kittner, S. et al. (2006). Heart disease and stroke statistics – 2006 update: A report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation* 113(6), e85–ee151.

Thomas, K.L. and Velazquez, E.J. (2005). Therapies to prevent heart failure post-myocardial infarction. *Curr Heart Fail Rep* 2(4), 174–182.

Topol, E.J., McCarthy, J., Gabriel, S., Moliterno, D.J., Rogers, W.J., Newby, L.K., Freedman, M., Metivier, J., Cannata, R., O'Donnell, C.J. et al. (2001). Single nucleotide polymorphisms in multiple novel thrombospondin genes may be associated with familial premature myocardial infarction. *Circulation* 104(22), 2641–2644.

Valgimigli, M., Percoco, G., Malagutti, P., Campo, G., Ferrari, F., Barbieri, D., Cicchitelli, G., McFadden, E.P., Merlini, F., Ansani, L.

et al. (2005). Tirofiban and sirolimus-eluting stent vs abciximab and bare-metal stent for acute myocardial infarction: A randomized trial. *JAMA* 293(17), 2109–2117.

Wang, L., Fan, C., Topol, S.E., Topol, E.J. and Wang, Q. (2003). Mutation of MEF2A in an inherited disorder with features of coronary artery disease. *Science* 302(5650), 1578–1581.

Wang, Q., Rao, S., Shen, G.Q., Li, L., Moliterno, D.J., Newby, L.K., Rogers, W.J., Cannata, R., Zirzow, E., Elston, R.C. et al. (2004). Premature myocardial infarction novel susceptibility locus on chromosome 1P34-36 identified by genomewide linkage analysis. *Am J Hum Genet* 74(2), 262–271.

Wellcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447(7145), 661–678.

Wessel, J., Topol, E.J., Ji, M., Meyer, J. and McCarthy, J.J. (2004). Replication of the association between the thrombospondin-4 A387P polymorphism and myocardial infarction. *Am Heart J* 147(5), 905–909.

Yamada, Y., Izawa, H., Ichihara, S., Takatsu, F., Ishihara, H., Hirayama, H., Sone, T., Tanaka, M. and Yokota, M. (2002). Prediction of the risk of myocardial infarction from polymorphisms in candidate genes. *N Engl J Med* 347(24), 1916–1923.

Yeh, H.I., Chou, Y., Liu, H.F., Chang, S.C. and Tsai, C.H. (2001). Connexin37 gene polymorphism and coronary artery disease in Taiwan. *Int J Cardiol* 81(2–3), 251–255.

Yusuf, S., Hawken, S., Ounpuu, S., Dans, T., Avezum, A., Lanas, F., McQueen, M., Budaj, A., Pais, P., Varigos, J. et al. (2004). Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): Case-control study. *Lancet* 364(9438), 937–952.

Zdravkovic, S., Wienke, A., Pedersen, N.L., Marenberg, M.E., Yashin, A.I. and de Faire, U. (2004). Genetic influences on CHD-death and the impact of known risk factors: Comparison of two frailty models. *Behav Genet* 34(6), 585–592.

Zeggini, E., Weedon, M.N., Lindgren, C.M., Frayling, T.M., Elliott, K.S., Lango, H., Timpson, N.J., Perry, J.R., Rayner, N.W., Freathy, R.M. et al. (2007). Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 316(5829), 1336–1341.

Zhang, B., Ye, S., Herrmann, S.M., Eriksson, P., de Maat, M., Evans, A., Arveiler, D., Luc, G., Cambien, F., Hamsten, A. et al. (1999). Functional polymorphism in the regulatory region of gelatinase B gene in relation to severity of coronary atherosclerosis. *Circulation* 99(14), 1788–1794.

RECOMMENDED RESOURCES

Topol, E.J. (2005). Simon Dack Lecture. The genomic basis of myocardial infarction. *J Am Coll Cardiol* 46(8), 1456–1465.

Wang, Q. (2005). Molecular genetics of coronary artery disease. *Curr Opin Cardiol* 20(3), 182–188.

Jefferson, B.K. and Topol, E.J. (2005). Molecular mechanisms of myocardial infarction. *Curr Probl Cardiol* 30(7), 333–374.

The Wellcome Trust Case Control Consortium (WTCCC). (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447(7145), 661–678.

CHAPTER

Acute Coronary Syndromes

L. Kristin Newby

INTRODUCTION

At least 2 million individuals are hospitalized each year with acute coronary syndromes (ACS), and 1 in 5 deaths in the United States is a result of ischemic heart disease (Rosamond et al., 2007). Despite its common occurrence and severe consequences, the exact molecular mechanisms that trigger an ACS are not fully elucidated, and the prognostic tools to identify which individuals are at risk for a first or subsequent ACS event are limited. Although it is clear the there is a genetic component to the occurrence of coronary artery disease (CAD) and acute coronary syndromes, most clinically applicable work thus far has fallen into the use of protein biomarkers measured from peripheral blood for use as diagnostic or prognostic tools. However, few examples exist of the use of these biomarkers to actually guide treatment. With rapid advances in technology, ACS and its treatment are ripe for exploration using genomics and related techniques such as proteomics and metabolomics. This chapter will briefly review what is known of the genetics and genomics of ACS, examine protein-based biomarkers and their combinations as an example of the forerunner of clinical proteomics, and conclude with an examination of the potential applications of "omic" techniques for refinement of diagnosis, risk stratification and management of ACS in the future.

PREDISPOSITION

Clinical Risk Factors and Risk Prediction

The development of CAD and the occurrence of myocardial infarction (MI) are complex processes. Although several clinical characteristics that predispose to these processes are now commonly known, including male sex, cigarette smoking, diabetes, hypertension, hyperlipidemia, and obesity, it is increasingly recognized that these characteristics provide only a superficial understanding of what predisposes to acute coronary syndromes. In addition, the ability to use these characteristics to discriminate individuals at risk from those who are not is at best modest. For example, the most widely used instrument for prediction of the risk for death or MI over a 10-year time horizon among individuals without known coronary disease, the Framingham Risk Score, which considers many of these clinical factors, has a c-index of only 0.69 in men and 0.72 in women (Ohman et al., 2000; Wilson et al., 1998). That is, for pairs of individuals, one who has a death or MI event and one who does not, the Framingham Risk Score will correctly select the one with an event only 69% of the time among men and only 72% of the time among women. Clinical models that assess the risk of a second event after the occurrence of an ACS, such as Global

Copyright © 2010, Elsevier Inc.
All rights reserved.

Registry of Acute Coronary Events (GRACE) model, which estimates the risk for in-hospital and 6-month death or MI (Eagle et al., 2004; Granger et al., 2003), have somewhat better predictive ability. The discriminative capability of the GRACE model (c-index 0.84) was enhanced by the addition of biomarker information (serum creatinine and creatine kinase [CK]-MB) to clinical characteristics. Still, the ability to discriminate which individuals will have an ACS event could be improved further, opening the door for the use of omic technologies, including RNA expression, proteomics, and metabolomics, to better characterize disease state and risk for future events.

Genetic Predisposition

Consensus on the existence of genetic predisposition for CAD is well-established. Family history has been shown repeatedly to be a robust, independent risk factor for CAD (Ciruzzi et al., 1997; Schildkraut et al., 1989; Shea et al., 1984), even after adjustment for shared environmental factors (Schildkraut et al., 1989; Zureik et al., 1999), and the heritability of CAD is particularly strong in early-onset forms, where the relative risk of developing early-onset CAD in a first-degree sibling is between 3.8 and 12.1, depending on the age-of-onset in the proband (Hauser et al., 2004).

Despite this, no single genetic variant has been identified that accounts for a major fraction of the large burden of CAD and cardiovascular events. Further, despite the wealth of published literature implicating genes in CAD, most association studies have been plagued by low strength of association and lack of replication. For summaries of the body of work in the genetics of coronary disease and MI, the reader is referred to excellent reviews in references by Ginsburg et al. (2005) and Chen et al. (2007). Most recently, variants of 2 genes in the 59-lipoxygenase (5-LO) pathway were associated with cardiovascular events in an Icelandic population: 5-LO activator protein (*FLAP*) associated with a twofold increase in MI and stroke (Helgadottir et al., 2004) and a haplotype (*HapK*) spanning the leukotriene A4 hydrolase (*LTA4H*) gene with a relative risk of 1.45 for MI among patients with other vascular disease (Helgadottir et al., 2005). When HapK was validated in cohorts from the United States, there was an even stronger, threefold increase in risk among African Americans.

Two recent, independent case-control genome-wide association studies, which were replicated in multiple independent populations, provided the first evidence of a common genetic variant (risk allele frequency ~45%) located on chromosome 9p21 that is associated with substantial risk of coronary heart disease or MI (Helgadottir 2007; McPherson et al., 2007). Homozygotes for the risk allele comprise 20–25% of the Caucasian population. Of note, this variant, in the region of tumor suppressor genes *CDKN2A* and *CDKN2B*, was associated with MI with an odds ratio of 1.26 for heterozygotes for the risk allele, and 1.64 for homozygotes. Still, these associations are modest and at this point provide little useful information to guide contemporary clinical care.

SCREENING

There are currently no purely genomic tools that have been identified to screen the general population for risk of future ACS, identification of current ACS events or for risk of death or recurrent ischemic events after an ACS. However, early results with expression profiling suggest potential applications for screening. RNA expression profiling using micorarray technology has been shown to accurately classify both the presence and severity of atherosclerotic lesions in aortic tissue (Seo et al., 2004) and to identify genes potentially involved in plaque rupture (Faber et al., 2001), a pathophysiological precursor to the clinical syndrome of ACS. Importantly, since access to tissue is challenging for screening, diagnosis or prognostic testing in ACS, studies have shown that gene expression profiling from circulating monocytes and peripheral blood leukocytes correlates with the extent of carotid vascular disease (Patino et al., 2005) and CAD (Ma and Liew, 2003), suggesting that blood can be used as a reporter tissue for events occurring in the vessel wall.

Unbiased metabolomic profiling of human serum by proton nuclear magnetic resonance (NMR) analysis also has predicted the presence and severity of CAD (Brindle et al., 2002). In fact, in this study, NMR-based analysis of human serum was better at differentiating individuals with one-, two-, and three-vessel CAD than a model of traditional clinical and laboratory risk factors. In addition, in a study of 53 cases with angiographic CAD and 53 controls without angiographic CAD, large-scale pooled plasma proteomics using LC/MS/MS technology identified differential expression of proteins and peptides between groups (Donahue et al., 2006). Of 731 proteins and peptides identified, 95 were differentially expressed in cases and controls. Among these were proteins involved in natural host defense mechanisms, growth, inflammation, and coagulation. These results suggest the potential for development as biomarkers, in addition to their use to further explore the mechanistic underpinnings of CAD and ischemic events. More recently, a study using urine proteomics identified biomarker patterns that correlated with the presence of angiographically severe CAD (Zimmerli et al., 2007). From more than 1000 polypeptides characterized per sample, 15 characterized a unique biosignature for CAD, which when used to predict CAD in a second cohort did so with a sensitivity of 98% and a specificity of 83%. Interestingly, after coronary intervention and increasing physical activity, the pattern reverted towards that of healthy controls, an observation that suggests potential not only for screening and diagnosis, but also for monitoring disease progression.

In summary, although genomic approaches to screening for CAD or quantifying future risk for ACS are not yet ready for clinical application, this is an active area of research with early results that suggest promise for applications to clinical care, development of biomarkers of disease state and risk and use as research tools to enhance understanding of the molecular underpinnings of ACS.

DIAGNOSIS

Conventional Diagnostic Tools

The acute coronary syndromes encompass a spectrum of pathophysiological processes manifest as plaque instability and plaque rupture or erosion, coronary thrombosis with varying degrees of coronary artery and microvascular occlusion and reductions in blood flow leading to myocardial ischemia and in some cases, myocardial necrosis (Fuster et al., 1992a, b). The diagnosis of ACS is largely a clinical one, relying on a patient's description, or history, of the event that led him to seek medical attention and a small number of readily available objective diagnostic tools. The first and most readily available objective assessment of a patient with suspected ACS is the 12-lead electrocardiogram (ECG). Based on the ECG, patients can be stratified into 2 primary groups, those with ST-segment elevation MI (STEMI) and those with non-ST-segment elevation (NSTE) ACS. About two-thirds of patients will have NSTE ACS and 1/3 will have STEMI (Figure 25.1; Morrow et al., 2007). This electrocardiographic distinction is clinically important as it immediately defines a divergence in treatment strategy. Patients with STEMI generally have a totally occluded major coronary artery and benefit from rapid reperfusion therapy with percutaneous revascularization or administration of intravenous fibrinolytic therapy. For an excellent review and current recommendations for the management of STEMI, the reader is referred to the American College of Cardiology/American Heart Association (ACC/AHA) guidelines for management of patients with STEMI (Antman et al., 2004).

In a small case-control study, platelet RNA expression analysis implicated 2 candidate proteins (CD69 and myeloid related protein-14) that differentiated STEMI patients from those with stable CAD (Healy et al., 2006). However, in the context of STEMI, these findings are mostly of interest as a research tool given the central role of the ECG in diagnosis and management of STEMI patients. This work does offer proof of principle and, as discussed subsequently, could have more relevance for application to identification of NSTE ACS patients.

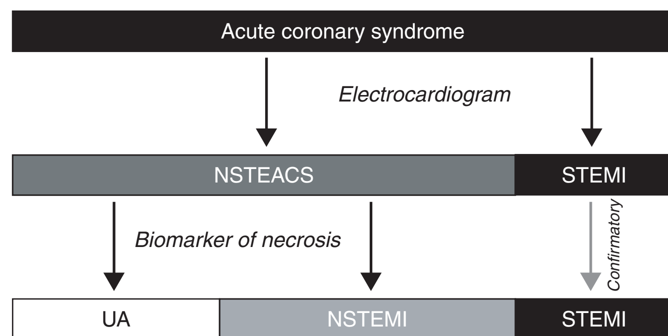


Figure 25.1 Diagnostic flow and distribution of types of acute coronary syndromes.

Patients whose initial ECG does not reveal ST-segment elevation are further subdivided into 2 groups, largely on the basis of measurement of levels of biomarkers released into the blood stream as a result of ischemic myocardial injury and necrosis. Although a number of markers have been used historically (SGOT, LDH, total CK), over the past 2 decades mass assays for the cardiac-specific isoenzyme of creatine kinase (CK-MB) and more recently, cardiac-specific isoforms of troponin T and I have been the recognized diagnostic gold standards for MI. With the publication of the joint ESC/ACC Task Force recommendations for MI redefinition in 2001, troponin, because of its enhanced sensitivity and specificity for myocardial necrosis, became the preferred diagnostic marker of MI in most clinical situations (Alpert et al., 2000). The 2007 revision of this document refined the definition of MI further in an attempt to achieve consensus, particularly on appropriate clinical situations for the use of troponins and/or CK-MB for MI diagnosis and diagnostic parameters of the assays, that would be applicable and practical across medical practice and clinical trials around the world (Thygesen et al., 2007). Along with the National Academy of Clinical Biochemistry (NACB) Practice Guidelines for Laboratory Medicine (Morrow et al., 2007), these documents summarize the base of evidence and recommendations for the use of biochemical markers in the diagnosis of ACS. Key recommendations for the use of biochemical markers to diagnose MI are shown in Table 25.1.

Need for Novel Biomarkers for Diagnosis of ACS

While CK-MB and troponins are useful diagnostic tools for MI, they only become elevated after irreversible myocardial injury has occurred. Ideally, one would identify patients with an acute coronary syndrome early, at a point when therapeutic intervention might mitigate myocardial damage. Further, the group of patients without ECG changes or biomarker elevations on presentation is a particularly challenging group to manage, especially when the clinical presentation is atypical. Over 6 million Americans present to emergency departments each year for evaluation of chest pain syndromes; clearly the vast majority are not ultimately diagnosed with an ACS. Admitting all of these individuals to the hospital is not practical from the perspective of resource use and costs, but the medico-legal risk associated with missing a patient with MI and discharging him from the emergency department is high. Missed MI, which carries a two-fold increased risk of death (Pope et al., 2000), is the leading cause of malpractice claims against emergency department and primary care physicians (Karcz et al., 1996; Rusnak et al., 1989). Therefore, there is a tremendous need for refined diagnostic testing to identify who among this large number of patients with normal ECG and troponin levels at presentation is having an ACS and to identify as early as possible those who need treatment or who can be safely sent home. While this is an area of great diagnostic need, few such tests are commercially available, making this clinical situation poised for the application of omic technologies to identify unique molecular signatures that

distinguish patients with unstable angina or pre-infarction from those with non-cardiac diagnoses.

In current practice, serial testing of multiple necrosis markers with different release kinetics after myocardial injury (CK-MB, troponin, and myoglobin) has been shown to identify MI in more patients in a general emergency department cohort and to do it as much as an hour earlier than conventional single marker testing with troponin or CK-MB (Newby et al., 2001). However, even earlier detection, before necrosis occurs, could be clinically beneficial for triage and treatment. Ischemia modified albumin (IMA) and myeloperoxidase are FDA-approved assays for use as adjuncts to ECG and troponin testing in this situation, but each has limitations to its widespread adoption.

The albumin cobalt binding (ACB) test takes advantage of modification of the *N*-terminus of albumin in the setting of ischemia (Bar-Or et al., 2001), reducing its ability to bind cobalt. The increased free cobalt can then be detected with special assays (Bar-Or et al., 2000). Importantly, in human angioplasty models of myocardial ischemia, IMA becomes positive within a few minutes after the onset of ischemia (Sinha et al., 2003), suggesting that it could be an ideal tool for early differentiation of chest pain patients in the emergency department. Unfortunately, the sensitivity, specificity, and positive predictive value of this test are far too low for use as a diagnostic tool for ischemia or pre-infarction or to guide treatment decisions in those with a positive test. However, it does have excellent negative predictive value, and when combined with negative troponin testing and non-diagnostic ECG can exclude the diagnosis of ischemia with a negative predictive value of over 97% (Peacock et al., 2006).

Myleoperoxidase is an enzyme activated in and released by activated neutrophils and monocytes; it is believed to have a pathophysiological role in ACS and may be a marker of plaque instability (Apple et al., 2005; Malech and Naussef, 1997). A study published in 2003 suggested a potential role for myeloperoxidase in early emergency department diagnosis and risk stratification (Brennan et al., 2003). In this study, among 604 patients presenting with chest pain to the emergency room, myeloperoxidase levels were elevated at baseline among patients who were initially troponin-negative, but subsequently troponin-positive, even when fewer than 3 hours had elapsed since the onset the symptoms. Further, there was a correlation of myeloperoxidase with both short and long-term outcomes. However, leukocyte and monocyte activation and myeloperoxidase release can occur in the setting of a number of disease states (Apple et al., 2005); therefore, like the ACB test, the low specificity of elevated myeloperoxidase levels for ACS precludes its use as a stand alone marker for diagnosis of ischemia. It must be interpreted in the context of clinical findings and other laboratory data.

Omic Technologies in Diagnosis of ACS

One of the most intriguing investigations that foreshadows the potential application of omic technologies to diagnostic test development for use in management of patients with suspected ACS involved metabolomics. In a small case-control study of 36 patients who underwent stress testing with nuclear myocardial perfusion imaging, unbiased metabolomic profiling of blood obtained pre- and post-test using LC/MS technology identified small molecules that changed in abundance differentially between patients with ischemia and those without (Figure 25.2) (Sabatine et al., 2005). After statistical analysis of the output, a score reflecting the presence or absence of change in the 6 most discordantly regulated metabolites predicted ischemia with a

Table 25.1 Class I recommendations for use of biochemical markers for diagnosis of myocardial infarction

Recommendation	Level of evidence
Biomarkers of myocardial necrosis should be measured in all patients who present with symptoms consistent with ACS.	Level of evidence: C
The patient's clinical presentation (history, physical exam) and ECG should be used in conjunction with biomarkers in the diagnostic evaluation of suspected MI.	Level of evidence: C
Cardiac troponin is the preferred marker for the diagnosis of MI. Creatine kinase MB (CK-MB) by mass assay is an acceptable alternative when cardiac troponin is not available.	Level of evidence: A
Blood should be obtained for testing at hospital presentation followed by serial sampling with timing of sampling based on the clinical circumstances. For most patients, blood should be obtained for testing at hospital presentation and at 6–9 h.	Level of evidence: C
In the presence of a clinical history suggestive of ACS, the following are considered indicative of myocardial necrosis consistent with MI a. Maximal concentration of cardiac troponin exceeding the 99th percentile of values (with optimal precision defined by total CV $< 10\%$) for a reference control group on at least 1 occasion during the first 24 h after the clinical event (observation of a rise and/or fall in values is useful in discriminating the timing of injury). b. Maximal concentration of CK-MB exceeding the 99th percentile of values for a sex-specific reference control group on 2 successive samples (values for CK-MB should rise and/or fall).	Level of evidence: C

Figure 25.2 Scatterplot of the statistical significance changes in metabolite levels according to case (ischemia) or control (no ischemia) status. Significant changes ($p < 0.05$) are represented as colored circles; no significant change as black dots. Red indicates and increase in concentration and green a decrease in concentration. The rim of the circle reflects change in the controls and the center reflects changes in the cases.

Figure 25.3 Conceptual framework for clinical proteomics efforts as proposed by the National Heart, Lung and Blood Institute clinical proteomics working group.

c-index of 0.83. Much work remains to translate this finding to practical clinical utility (e.g., validation in unselected emergency department populations with suspected ischemia, development of assays that are practical for use in hospital clinical laboratories and in real time), but the concept highlights the possibilities for application of omic technologies to the early detection of ACS.

The use of proteomics, both candidate-based approaches and unbiased approaches, as a platform for discovery and development of novel biomarkers of cardiovascular disease has been widely espoused (Anderson, 2005a, b; Arab et al., 2006; Granger et al., 2004). Figure 25.3 shows a conceptual framework for such efforts proposed by the NHLBI clinical proteomics working group. Despite this enthusiasm, the use of proteomics to distinguish ACS patients from those without ACS is in its infancy. Recently, however, in a small study of 11 patients with acute MI, 8 with unstable angina and 9 age-matched controls without known coronary

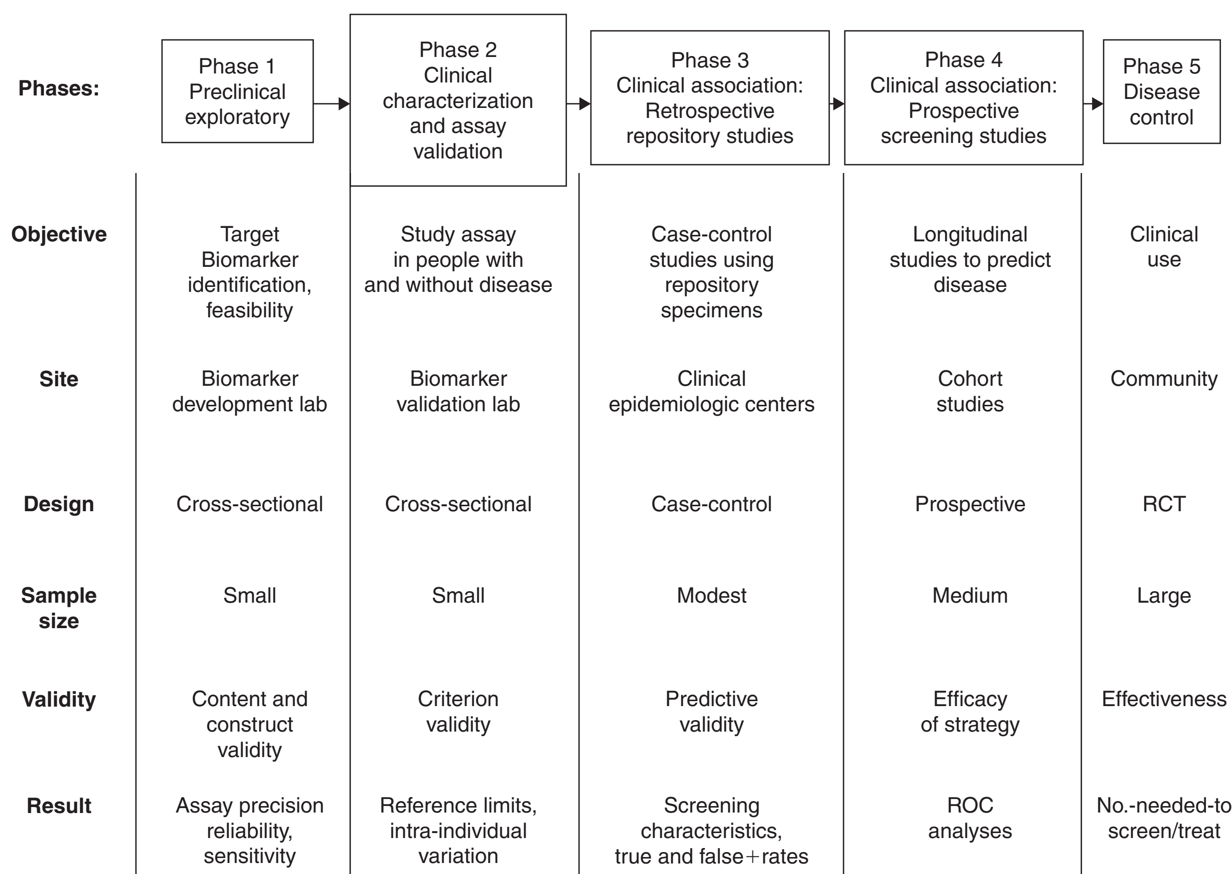


Figure 25.4 A 5-step process that is essential to move from biomarker discovery to clinical utility.

disease, Mateos-Cáceres and colleagues published a seminal report on the used of 2d gel-based proteomics to identify changes in the protein map that discriminated patients with acute MI from those with unstable angina based on changes relative to controls (Mateos-Cáceres et al., 2004). These observations provide support that proteomics may lead to usable clinical tools to distinguish disease state in the clinical setting of suspected ACS. In an excellent review of the development of biomarkers in cardiovascular disease, Vasan has outlined a 5-step process that is essential to move from discovery findings such as this to clinical utility (Figure 25.4; Vasan, 2006).

Finally, as mentioned above, the use of platelet RNA expression profiling (or possibly peripheral blood RNA expression analysis), either directly or to identify candidate protein biomarkers of ACS, may prove to be a useful tool for differentiating difficult to diagnose NSTE ACS presentations from non-ACS presentations or to identify NSTE ACS earlier after presentation.

PROGNOSIS

The goals of "biomarker" testing are multifold. As described in previous sections screening for disease and diagnosis are important applications. However, equally important is the ability to use the information from biomarker testing to discern not only the presence or absence of disease, but more importantly to relate biomarker information to prognosis or downstream meaningful clinical outcomes. Although the concept of using genetic and genomic, proteomic, and metabolomic approaches for risk stratification and to refine prognosis in ACS is appealing, substantial work remains to bring this conceptual construct to clinical utility.

That there is a need for such advances to refine clinical models of risk is undenied, and numerous examples of the development and use of biomarker assessment for prognosis or risk stratification abound in the ACS literature. Although hundreds of individual proteins, in addition to routine laboratory tests, have been identified as predicting clinical outcome, only a few have sufficient evidence to support their use in routine clinical practice. Guidelines for the use of biomarker testing for risk stratification in ACS were recently published by the NACB, and class I recommendations are summarized in Table 25.2 (Morrow et al., 2007). In this regard, there is substantial evidence for use of troponin (I or T), b-type natriuretic peptide (BNP) or *N*-terminal proBNP, and high-sensitivity c-reactive protein (hsCRP) for risk stratification of ACS patients. Each of these biomarkers contributes incremental information to that from easily assessable clinical and laboratory parameters and there are commercially available, high-quality assays for clinical use. Although many

Table 25.2 Class I recommendations for use of biochemical markers for risk stratification in ACS

Recommendation	Level of evidence
Patients with suspected ACS should undergo early risk stratification based on an integrated assessment of symptoms, physical exam findings, ECG findings, and biomarkers.	Level of evidence: C
A cardiac troponin is the preferred marker for risk stratification and, if available, should be measured in all patients with suspected ACS. In patients with a clinical syndrome consistent with ACS, a maximal (peak) concentration exceeding the 99th percentile of values for a reference control group should be considered indicative of increased risk of death and recurrent ischemic events.	Level of evidence: A
Blood should be obtained for testing on hospital presentation followed by serial sampling with timing of sampling based on the clinical circumstances. For most patients, blood should be obtained for testing at hospital presentation and at 6–9 h.	Level of evidence: B

other novel protein markers have been identified (Ginsburg et al., 2005; Morrow et al., 2007), their independent contribution to risk stratification remains to be proved and development of assays for routine use in hospital laboratories remains.

As a forerunner to the application of proteomics (or other multidimensional omic technologies) to risk stratification of ACS patients, many studies have now demonstrated the utility of combining information carried by multiple markers of risk into a common risk prediction model. For example, in one study, a simple score based on measurement of troponin, BNP and hsCRP provided additive, complementary information to that from any marker alone in the context of clinical markers of risk (Sabatine et al., 2002). Additionally, investigators from the CAPTURE trial used a candidate protein approach, to identify six proteins (hsCRP, troponin T, CD40 ligand, myeloperoxidase, vascular endothelial growth factor (VEGF), placental growth factor, and pregnancy-associated plasma protein [PAPP]-A) that together refined prediction of 6-month death or MI compared with use of any single marker in both ACS and general emergency department patients with suspected ACS (Heeschen et al., 2005). In another report (Baldus et al., 2003), five markers each had independent additive value, and in a multivariable model, each protein marker had more predictive information than any clinical variable.

Finally, one of the most important insights into the use of protein biomarkers in ACS risk stratification came from investigators in a substudy of the GUSTO-IV trial. They examined use of markers of multiple processes (necrosis, troponin; inflammation, hsCRP, and neurohormonal activation, NT-proBNP) in the context of clinical characteristics and routine laboratory testing (Westerhout et al., 2006), and made the important observation that the prognostic utility of a given marker depended on the outcome of interest. NT-proBNP, troponin T, and hsCRP were all independently associated with mortality at 1 year, in addition to heart rate, creatinine clearance, and ST-segment depression. However, only troponin T, creatinine clearance and ST-segment depression predicted risk for future MI. In addition, the relative strengths of the association of a biomarker with risk for subsequent events varied according to the time frame in which it was assessed. Thus, protein biomarkers may vary in their utility for short- and long-term risk stratification after ACS, but protein biomarkers consistently comprise several of the most important independent predictors of risk.

2009 UPDATE

The development of genomic biomarkers for use to evaluate clinical risk, speed development, and assess safety and efficacy of therapeutics and guide treatment decisions remains a major focus of activity of the FDA Critical Pathway initiative, including a specific focus in heart disease, and the NIH through multiple avenues including partnership with the FDA and others in the Biomarker Consortium (Biomarkers Consortium, 2009; US Department of Health and Human Services, 2008). However, work to establish a uniform framework for critical evaluation and translation of this work to practical utility in the clinical setting, clinical trials, or drug discovery and development remains an ongoing process.

Although academic activity in biomarker research has remained high over the past year (a simple PubMed search for "biomarkers" in "acute coronary syndromes" yielded nearly 400 citations for the year 2008), there persist the challenges of sample size, determination of independent and incremental contribution to routinely available clinical characteristics or laboratory tests and validation in separate cohorts. Perhaps the most interesting example of new research in ACS that may have near-term practical utility was the finding in 2208 French patients of an association between carrying more than one CYP2C19 loss-of-function alleles and a 2-fold higher 1-year rate of major adverse cardiovascular outcomes after MI (Simon et al., 2009). This association was particularly strong among patients undergoing percutaneous coronary intervention (nearly 4-fold increase in death, MI, or stroke). These findings, if validated in other populations and confirmed prospectively,

could have important implications for clopidogrel dosing or alternate selection of new, more potent antiplatelet agents such as prasugrel based on genetic screening. While limited to smaller populations and less immediately clinically applicable, new examples of the application of metabolomic (Vallejo et al., 2009) and gene expression (Nakayama et al., 2008; Wingrove et al., 2008) profiling in characterization of coronary disease and ACS populations continue to emerge.

Finally, major efforts to establish the large, longitudinal clinical databases and specimen biorepositories that will be necessary to fuel future genomic research and confirm preliminary findings such as these have been initiated and are well underway. For example, the UK Biobank, which has as a goal collecting blood and urine samples and baseline and longitudinal clinical information from 500,000 Britons between the ages of 40 and 69 by the end 2008, had already enrolled half of its target population (UK Biobank, 2009). Thus, there is a continually emerging confluence of scientific interest in and recognition of the need for genomic research in ACS, along with the support and growing resources that should be expected to lead to major advances over the coming years.

2009 UPDATE REFERENCES

Biomarkers Consortium. (2009). Developing & qualifying 'High-Impact' biomarkers to advance science. http://www.biomarkersconsortium.org/index.php?option=com_content&task=view&id=109&Itemid=61 (accessed January 26, 2009)

Nakayama, M., Kudoh, T., Kaikita, K., Yoshimura, M., Oshima, S., Miyamoto, Y., Takeya, M. and Ogawa, H. (2008). Class A macrophage scavenger receptor gene expression levels in peripheral blood mononuclear cells specifically increase in patients with acute coronary syndrome. *Atherosclerosis* 198, 426–433.

Simon, T., Verstuyft, C., Mary-Krause, M., Quteineh, L., Drouet, E., Méneveau, N., Steg, P.G., Ferrières, J., Danchin, N. and Becquemont, L. for the French Registry of Acute ST-Elevation and Non-ST-Elevation Myocardial Infarction (FAST-MI) Investigators (2009). Genetic determinants of response to clopidogrel and cardiovascular events. *N Engl J Med* 360, 363–375.

UK Biobank. (2009). http://www.ukbiobank.ac.uk/about/what.php (accessed January 26, 2009)

US Department of Health and Human Services, Food and Drug Administration (2008). Key FDA critical path activities under way in 2007. http://www.fda.gov/oc/initiatives/criticalpath/report2007.html (accessed January 26, 2009)

Vallejo, M., García, A., Tuñón, J., García-Martínez, D., Angulo, S., Martin-Ventura, J.L., Blanco-Colio, L.M., Almeida, P., Egido, J. and Barbas, C. Plasma fingerprinting with GC-MS in acute coronary syndrome. *Anal Bioanal Chem* 2009 January 28; [Epub ahead of print].

Wingrove, J.A., Daniels, S.E., Sehnert, A.J., Tingley, W., Elashoff, M.R., Rosenberg, S., Buellesfeld, L., Grube, E., Newby, L.K., Ginsburg, G.S. et al. (2008). Correlation of peripheral blood gene expression with the extent of coronary artery stenosis. *Circ Cardiovasc Genet* 1, 31–38.

ACKNOWLEDGEMENTS

Within the past 2 years, Dr Newby has received research grant support from Schering Plough, Millennium Pharmaceuticals, Medicure, Roche Diagnostics, Inverness Medical, and Adolor; and has received consulting honoraria from Johnson and Johnson/Scios, Proctor & Gamble, Novartis, Astra-Zeneca/Atherogenics, Biosite, Roche Diagnostics, and CV Therapeutics.

REFERENCES

Alpert, J.S., Thygesen, K., Antman, E. and Bassand, J.P. (2000). Myocardial infarction redefined--a consensus document of The Joint European Society of Cardiology/American College of Cardiology Committee for the redefinition of myocardial infarction. *J Am Coll Cardiol* 36, 959–969. Erratum in: *J Am Coll Cardiol* 2001; 37, 973.

Anderson, N.L. (2005a). Candidate-based proteomics in the search for biomarkers of cardiovascular disease. *J Physiol* 563, 23–60.

Anderson, N.L. (2005b). The roles of multiple proteomic platforms in a pipeline for new diagnostics. *Mol Cell Proteomics* 4, 1441–1444.

Antman, E.M., Tanasijevic, M.J., Thompson, B., Schactman, M., McCabe, C.H., Cannon, C.P., Fischer, G.A., Fung, A.Y., Thompson, C., Wybenga, D. et al. (1996). Cardiac-specific troponin I levels to predict the risk of mortality in patients with acute coronary syndromes. *N Engl J Med* 335, 1342–1349.

Antman, E.M., Anbe, D.T., Armstrong, P.W., Bates, E.R., Green, L.A., Hand, M., Hochman, J.S., Krumholz, H.M., Kushner, F.G., Lamas, G.A. et al. (2004). ACC/AHA guidelines for the management of patients with ST-elevation myocardial infarction: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Revise the 1999 Guidelines for the Management of Patients with Acute Myocardial Infarction). *Circulation* 110, e82–e292. Erratum in: *Circulation* 2005; 111, 2013–2014. *Circulation* 2007; 115, e411.

Apple, F.S., Wu, A.H., Mair, J., Ravkilde, J., Panteghini, M., Tate, J., Pagani, F., Christenson, R.H., Mockel, M., Danne, O. et al. (2005). For the Committee on Standardization of Markers of Cardiac Damage of the IFCC. Future biomarkers for detection of ischemia and risk stratification in acute coronary syndrome. *Clin Chem* 51, 810–824.

Arab, S., Gramolini, A.O., Ping, P., Kislinger, T., Stanley, B., van Eyk, J., Ouzounian, M., MacLennan, D.H., Emili, A. and Liu, P.P. (2006). Cardiovascular proteomics: Tools to develop novel biomarkers and potential applications. *J Am Coll Cardiol* 48, 1733–1741.

Baldus, S., Heeschen, C., Meinertz, T., Zeiher, A.M., Eiserich, J.P., Münzel, T., Simoons, M.L. and Hamm, C.W. (2003). for the CAPTURE Investigators. Myeloperoxidase serum levels predict risk in patients with acute coronary syndromes. *Circulation* 108, 1440–1445.

Bar-Or, D., Lau, E. and Winkler, J.V. (2000). Novel assay for cobalt-albumin binding and its potential as a marker for myocardial ischemia – a preliminary report. *J Emerg Med* 19, 311–315.

Bar-Or, D., Curtis, G., Rao, N., Bampos, N. and Lau, E. (2001). Characterization of the Co2_ and Ni2_ binding amino-acid residues of the N-terminus of human albumin. *Eur J Biochem* 268, 42–47.

Brennan, M.L., Penn, M.S., Van Lente, F., Nambi, V., Shishehbor, M.H., Aviles, R.J., Goormastic, M., Pepoy, M.L., McErlean, E.S., Topol, E.J. et al. (2003). Prognostic value of myeloperoxidase in patients with chest pain. *N Engl J Med* 349, 1595–1604.

Brindle, J.T., Antti, H., Holmes, E., Tranter, G., Nicholson, J.K., Bethell, H.W., Clarke, S., Schofield, P.M., McKilligin, E., Mosedale, D.E. et al. (2002). Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using ^{1}H-NMR-based metabonomics. *Nat Med* 8, 1439–1444.

Chen, Y., Rollins, J., Paigen, B. and Wang, X. (2007). Genetic and genomic insights into the molecular basis of atherosclerosis. *Cell Metab* 6, 165–179.

Ciruzzi, M., Schargrodsky, H., Rozlosnik, J., Pramparo, P., Delmonte, H., Rudich, V., Piskorz, D., Negri, E., Soifer, S. and La Vecchia, C. (1997). Frequency of family history of acute myocardial infarction in patients with acute myocardial infarction. Argentine FRICAS (Factores de Riesgo Coronario en America del Sur) Investigators. *Am J Cardiol* 80, 122–127.

Donahue, M.P., Rose, K., Hochstrasser, D., Vonderscher, J., Grass, P., Chibout, S.D., Nelson, C.L., Sinnaeve, P., Goldschmidt-Clermont, P.J. and Granger, C.B. (2006). Discovery of proteins related to coronary artery disease using industrial-scale proteomics analysis of pooled plasma. *Am Heart J* 152, 478–485. Erratum in: *Am Heart J* 2006; 152, 1021

Eagle, K.A., Lim, M.J., Dabbous, O.H., Pieper, K.S., Goldberg, R.J., Van de Werf, F., Goodman, S.G., Granger, C.B., Steg, P.G., Gore, J.M. et al. (2004). for the GRACE Investigators. A validated prediction model for all forms of acute coronary syndrome: Estimating the risk of 6-month postdischarge death in an international registry. *JAMA* 291, 2727–2733.

Faber, B.C., Cleutjens, K.B., Niessen, R.L., Aarts, P.L., Boon, W., Greenberg, A.S., Kitslaar, P.J., Tordoir, J.H. and Daemen, M.J. (2001). Identification of genes potentially involved in rupture of human atherosclerotic plaques. *Circ Res* 89, 547–554.

Fuster, V., Badimon, L., Badimon, J.J. and Chesebro, J.H. (1992a). The pathogenesis of coronary artery disease and the acute coronary syndromes (1). *N Engl J Med* 326, 242–250.

Fuster, V., Badimon, L., Badimon, J.J. and Chesebro, J.H. (1992b). The pathogenesis of coronary artery disease and the acute coronary syndromes (2). *N Engl J Med* 326, 310–318.

Ginsburg, G.S., Donahue, M. and Newby, L.K. (2005). Prospects for personalized cardiovascular medicine: The impact of genomics. *J Am Coll Cardiol* 46, 1615–1627.

Granger, C.B., Goldberg, R.J., Dabbous, O., Pieper, K.S., Eagle, K.A., Cannon, C.P., Van De Werf, F., Avezum, A., Goodman, S.G., Flather, M.D. et al. (2003). For the Global Registry of Acute Coronary Events Investigators. Predictors of hospital mortality in the global registry of acute coronary events. *Arch Intern Med* 163, 2345–2353.

Granger, C.B., Van Eyk, J.E., Mockrin, S.C. and Anderson, N.L. (2004). National Heart, Lung, And Blood Institute Clinical Proteomics Working Group. National Heart, Lung, And Blood Institute Clinical Proteomics Working Group report. *Circulation* 109, 1697–1703.

Hauser, E.R., Crossman, D.C., Granger, C.B., Haines, J.L., Jones, C.J., Mooser, V., McAdam, B., Winkelmann, B.R., Wiseman, A.H., Muhlestein, J.B. et al. (2004). A genomewide scan for early-onset coronary artery disease in 438 families: The GENECARD Study. *Am J Hum Genet* 75, 436–447.

Healy, A.M., Pickard, M.D., Pradhan, A.D., Wang, Y., Chen, Z., Croce, K., Sakuma, M., Shi, C., Zago, A.C., Garasic, J. et al. (2006). Platelet expression profiling and clinical validation of myeloid-related protein-14 as a novel determinant of cardiovascular events. *Circulation* 113, 2278–2285.

Heeschen, C., Dimmeler, S., Hamm, C.W., Fichtlscherer, S., Simoons, M.L. and Zeiher, A.M. (2005). For the CAPTURE Investigators. Pregnancy-associated plasma protein-A levels in patients with acute coronary syndromes: Comparison with markers of systemic inflammation, platelet activation, and myocardial necrosis. *J Am Coll Cardiol* 45, 229–237.

Helgadottir, A., Manolescu, A., Thorleifsson, G., Gretarsdottir, S., Jonsdottir, H., Thorsteinsdottir, U., Samani, N.J., Gudmundsson, G., Grant, S.F., Thorgeirsson, G. et al. (2004). The gene encoding 5-liposygenase activating protein confers risk of myocardial infarction and stoke. *Nat Genet* 36, 233–239.

Helgadottir, A., Manolescu, A., Helgason, A., Thorleifsson, G., Thorsteinsdottir, U., Gudbjartsson, D.F., Gretarsdottir, S., Magnusson, K.P., Gudmundsson, G., Hicks, A. et al. (2005). A variant of the gene encoding leukotriene A4 hydrolase confers ethnicity-specific risk of myocardial infarction. *Nat Genet* 38, 68–74.

Karcz, A., Korn, R., Burke, M.C., Caggiano, R., Doyle, M.J., Erdos, M.J., Green, E.D. and Williams, K. (1996). Malpractice claims against emergency physicians in Massachusetts: 1975–1993. *Am J Emerg Med* 14, 341–345.

Ma, J. and Liew, C.C. (2003). Gene profiling identifies secreted protein transcripts from peripheral blood cells in coronary artery disease. *J Mol Cell Cardiol* 35, 993–998.

Malech, H.L. and Naussef, W.M. (1997). Primary inherited defects in neutrophil function: Etiology and treatment. *Semin Hematol* 34, 279–290.

Mateos-Cáceres, P.J., García-Méndez, A., López Farré, A., Macaya, C., Núñez, A. and Gómez, P. (2004). analysis of plasma from patients during an acute coronary syndrome. *J Am Coll Cardiol* 44, 1578–1583.

McPherson, R., Pertsemlidis, A., Kavaslar, N., Stewart, A., Roberts, R., Cox, D.R., Hinds, D.A., Pennacchio, L.A., Tybjaerg-Hansen, A., Folsom, A.R. et al. (2007). A common allele on chromosome 9 associated with coronary heart disease. *Science* 316, 1488–1491.

Morrow, D.A., Cannon, C.P., Jesse, R.L., Newby, L.K., Ravkilde, J., Storrow, A.B., Wu, A.H.B. and Christenson, R.H. (2007). National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines: Clinical characteristics and utilization of biochemical markers in acute coronary syndromes. *Circulation* 115, e356–e375.

Newby, L.K., Storrow, A.B., Gibler, W.B., Garvey, J.L., Tucker, J.F., Kaplan, A.L., Schreiber, D.H., Tuttle, R.H., McNulty, S.E. and Ohman, E.M. (2001). Bedside multimarker testing for risk stratification in chest pain units: The chest pain evaluation by creatine kinase-MB, myoglobin, and troponin I (CHECKMATE) study. *Circulation* 103, 1832–1837.

Ohman, E.M., Granger, C.B., Harrington, R.A. and Lee, K.L. (2000). Risk stratification and therapeutic decision making in acute coronary syndromes. *JAMA* 284, 876–878.

Patino, W.D., Mian, O.Y., Kang, J.G. et al. (2005). Circulating transcriptome reveals markers of atherosclerosis. *Proc Natl Acad Sci U S A* 102, 3423–3428.

Peacock, F., Morris, D.L., Anwaruddin, S., Christenson, R.H., Collinson, P.O., Goodacre, S.W., Januzzi, J.L., Jesse, R.L., Kaski, J.C., Kontos, M.C. et al. (2006). Meta-analysis of ischemia-modified albumin to rule out acute coronary syndromes in the emergency department. *Am Heart J* 152, 253–262.

Pope, J.H., Aufderheide, T.P., Ruthazer, R., Woolard, R.H., Feldman, J.A., Beshansky, J.R., Griffith, J.L. and Selker, H.P. (2000). Missed diagnoses of acute cardiac ischemia in the emergency department. *N Engl J Med* 342, 1163–1170.

Rosamond, W., Flegal, K., Friday, G., Furie, K., Go, A., Greenlund, K., Haase, N., Ho, M., Howard, V., Kissela, B. et al. (2007). For the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics—2007 update: A report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation* 115, e69–e171.

Rusnak, R.A., Stair, T.O., Hansen, K. and Fastow, J.S. (1989). Litigation against the emergency physician: Common features in cases of missed myocardial infarction. *Ann Emerg Med* 18, 1029–1034.

Sabatine, M.S., Morrow, D.A., de Lemos, J.A., Gibson, C.M., Murphy, S.A., Rifai, N., McCabe, C., Antman, E.M., Cannon, C.P. and Braunwald, E. (2002). Multimarker approach to risk stratification in non-ST elevation acute coronary syndromes: Simultaneous assessment of troponin I, C-reactive protein, and B-type natriuretic peptide. *Circulation* 105, 1760–1763.

Sabatine, M.S., Liu, E., Morrow, D.A., Heller, E., McCarroll, R., Wiegand, R., Berriz, G.F., Roth, F.P. and Gerszten, R.E. (2005). Metabolomic identification of novel biomarkers of myocardial ischemia. *Circulation* 112, 3868–3875.

Schildkraut, J.M., Myers, R.H., Cupples, L.A., Kiely, D.K. and Kannel, W.B. (1989). Coronary risk associated with age and sex of parental heart disease in the Framingham Study. *Am J Cardiol* 64, 555–559.

Seo, D., Wang, T., Dressman, H., Herderick, E.E., Iversen, E.S., Dong, C., Vata, K., Milano, C.A., Rigat, F., Pittman, J. et al. (2004). Gene expression phenotypes of atherosclerosis. *Arterioscler Thromb Vasc Biol* 24, 1922–1927.

Shea, S., Ottman, R., Gabrieli, C., Stein, Z. and Nichols, A. (1984). Family history as an independent risk factor for coronary artery disease. *J Am Coll Cardiol* 4, 793–801.

Sinha, M.K., Gaze, D.C., Tippins, J.R., Collinson, P.O. and Kaski, J.C. (2003). Ischemia modified albumin is a sensitive marker of myocardial ischemia after percutaneous coronary intervention. *Circulation* 107, 2403–2405.

Thygesen, K., Alpert, J.S. and White, H.D. (2007). on behalf of the Joint ESC/ACCF/AHA/WHF Task Force for the Redefinition of Myocardial Infarction. Universal definition of myocardial infarction. *Eur Heart J* 28, 2525–2538.

Vasan, R.S. (2006). Biomarkers of cardiovascular disease: Molecular basis and practical considerations. *Circulation* 113, 2335–2362.

Westerhout, C.M., Fu, Y., Lauer, M.S., James, S., Armstrong, P.W., Al-Hattab, E., Califf, R.M., Simoons, M.L., Wallentin, L. and Boersma, E. (2006). for the GUSTO-IV ACS Trial Investigators. Short- and long-term risk stratification in acute coronary syndromes: The added value of quantitative ST-segment depression and multiple biomarkers. *J Am Coll Cardiol* 48, 939–947.

Wilson, P.W., D'Agostino, R.B., Levy, D., Belanger, A.M., Silbershatz, H. and Kannel, W.B. (1998). Prediction of coronary heart disease using risk factor categories. *Circulation* 97, 1837–1847.

Zimmerli, L.U., Schiffer, E., Zürbig, P., Good, D.M., Kellmann, M., Mouls, L., Pitt, A.R., Coon, J.J., Schmieder, R.E., Peter, K. et al. (2007 19). Urinary proteomic biomarkers in coronary artery disease. *Mol Cell Proteomics*. [Epub ahead of print]

Zureik, M., Touboul, P.J., Bonithon-Kopp, C., Courbon, D., Ruelland, I. and Ducimetiere, P. (1999). Differential, association of common carotid intima-media thickness and carotid atherosclerotic plaques with parental history of premature death from coronary heart disease: The EVA study. *Arterioscler Thromb Vasc Biol* 19, 366–371.

RECOMMENDED RESOURCES

The following 3 articles represent state of the art summaries and recommendations for biomarker testing in the diagnosis of myocardial infarction and prognostication in ACS patients:

Thygesen, K., Alpert, J.S. and White, H.D. (2007). On behalf of the Joint ESC/ACCF/AHA/WHF Task Force for the Redefinition of Myocardial Infarction. Universal definition of myocardial infarction. *Eur Heart J* 28, 2525–2538.

Morrow, D.A., Cannon, C.P., Jesse, R.L., Newby, L.K., Ravkilde, J., Storrow, A.B., Wu, A.H.B. and Christenson, R.H. (2007). National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines: Clinical characteristics and utilization of biochemical markers in acute coronary syndromes. *Circulation* 115, e356–e375.

Apple, F.S., Jesse, R.L., Newby, L.K., Wu, A.H.B. and Christenson R.H. (2007). National Academy of Clinical Biochemistry and IFCC Committee for Standardization of Markers of Cardiac Damage Laboratory Medicine. Practice Guidelines: Analytical issues for biomarkers of acute coronary syndromes. *Circulation* 115, e352–e355.

This paper presents an excellent perspective on the potential for application of genomics to personalized cardiovascular care:

Ginsburg, G.S., Donahue, M. and Newby, L.K. (2005). Prospects for personalized cardiovascular medicine: The impact of genomics. *J Am Coll Cardiol* 46, 1615–1627.

These 2 papers provide expert opinion and recommendations for the application of proteomics to cardiovascular medicine and biomarker development:

Granger, C.B., Van Eyk, J.E., Mockrin, S.C. and Anderson N.L. (2004). National Heart, Lung, and Blood Institute Clinical Proteomics Working Group. National Heart, Lung, and Blood Institute Clinical Proteomics Working Group Report. *Circulation* 109, 1697–1703.

Vasan, R.S. (2006). Biomarkers of cardiovascular disease: Molecular basis and practical considerations. *Circulation* 113, 2335–2362.

CHAPTER

Heart Failure in the Era of Genomic Medicine

Ivor J. Benjamin and Jeetendra Patel

INTRODUCTION

Heart failure has been conveniently subdivided according to abnormalities in the cardiac cycle: namely, systolic heart failure (SHF) and diastolic heart failure (DHF). SHF is associated with decreased cardiac output and ventricular contractility, termed systolic dysfunction, and is attributed to a loss of ventricular muscle cells. Dilated cardiomyopathy (DCM) is characterized by impaired systolic function and myocardial remodeling and enlargement of one or both ventricles. Idiopathic dilated cardiomyopathy (IDCM) refers to primary myocardial disease in the absence of coronary, valvular or systemic disease. The ventricular remodeling of DHF, however, is characterized by normal chamber size with impaired ventricular filling from abnormal myocardial stiffness during the relaxation phase. More recently, the clinical syndrome of heart failure with preserved ejection fraction (HFPEF) – left ventricular ejection fraction $> 50\%$ – has been recognized in several cross-sectional studies (Bhatia et al., 2006; Owan et al., 2006).

PREDISPOSITION (GENETIC AND NON-GENETIC)

In western societies, ischemic heart disease (~60%) and hypertension are the most common causes of ventricular systolic dysfunction but there is now irrefutable evidence for genetic defects whose onset and progression occur in adulthood (e.g., familial cardiomyopathy) (Benjamin and Schneider, 2005; Morita et al., 2005). Beginning in the late 1950s, distinct alterations in the size and geometry of the left ventricle, termed "ventricular remodeling," were being recognized, but the ensuing debate for over three decades was primarily focused on morphological classifications. Hypertrophic cardiomyopathy (HCM) is characterized by predominant and marked thickening of the left circumferential ventricular wall (i.e., hypertrophy), small LV cavity size and hypercontractility. Such patients including young athletes were prone to sudden cardiac death attributed pathophysiologically to subaortic stenosis and cavity obliteration triggering inadequate cardiac output and lethal arrhythmias. In contrast, dilatation of left ventricular cavity and reduced systolic function are the hallmarks of dilated cardiomyopathy (DCM). In 1991, the Seidmans' laboratory at Harvard Medical School reported for the first time that mutations in the gene encoding the β-myosin heavy chain, a major structural and contractile protein, was the genetic basis for familial HCM associated with sudden death, ushering in the present era of cardiovascular genomic medicine. This seminal discovery permanently shifted the paradigm from the morphological to the molecular, enabling basic insights into disease pathogenesis to be viewed from how single gene defects orchestrate profound alterations at the biochemical, metabolic, hemodynamic, and physiological levels. In parallel, genetically engineered animals models became the state-of-the art for establishing causality

Copyright © 2010, Elsevier Inc.
All rights reserved.

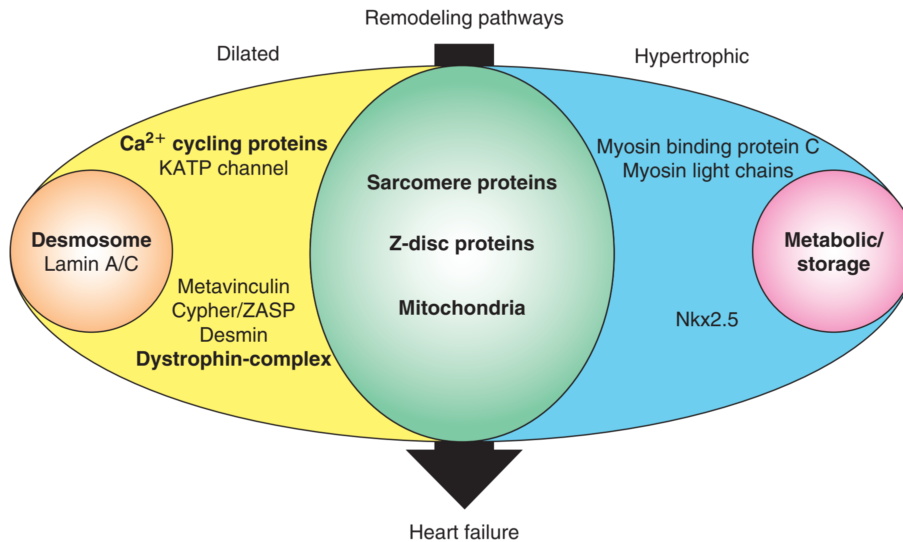


Figure 26.1 Human gene mutations can cause cardiac hypertrophy (blue), dilation (yellow), or both (green). In addition to these two patterns of remodeling, particular gene defects produce hypertrophic remodeling with glycogen accumulation (pink) or dilated remodeling with fibrofatty degeneration of the myocardium (orange). Sarcomere proteins denote β-myosin heavy chain, cardiac troponin T, cardiac troponin I, α-tropomyosin, cardiac actin, and titin. Metabolic/storage proteins denote AMPactivated protein kinase γ subunit, LAMP2, lysosomal acid α 1,4-glucosidase, and lysosomal hydrolase α-galactosidase A. Z-disc proteins denote MLP and telethonin. Dystrophin-complex proteins denote δ-sarcoglycan, β-sarcoglycan, and dystrophin. Ca^{2+} cycling proteins denote PLN and RyR2. Desmosome proteins denote plakoglobin, desmoplakin, and plakophilin-2. (Reprinted with permission from Morita et al., 2005.)

and, ultimately, a basis to test proof-of-concept leading to disease prevention.

Many more single genetic defects are routinely being linked to familial heart failure (Figure 26.1) but an important future challenge is to establish how inherited and acquired factors conspire to drive the growing epidemic of heart failure.

Severe occlusive coronary disease is the substrate for acute coronary syndromes, myocardial infarctions and subsequent pump failure as shown in Figure 26.2. The high prevalence of heart failure in African-Americans with hypertension underscores potential gene–environment interactions in selected populations. Infectious etiologies (e.g., rheumatic heart disease) are declining but valvular heart disease from iatrogenic causes (e.g., diet pills, toxins) remains an important risk factor (Figure 26.2). Viruses (e.g., Coxsackie's B3, parvovirus) are the major suspected culprits for idiopathic dilated cardiomyopathy (IDCM) in which the postviral sequalae of inflammation and apoptosis trigger ventricular remodeling and dilation (Liu and Mason, 2001). IDCM accounts for 30% of cases of DCM. Heart failure on presentation in the peri-partum or post-partum period has a variable clinical course from severe pump failure to complete recovery. The most common cause of right ventricular heart failure (RVHF) is left ventricular systolic dysfunction. In addition, RVHF is associated with congenital heart disease (e.g., tetralogy of fallot), primary pulmonary hypertension, and arrhythmogenic right ventricular dysphasia and right ventricular infarction. Stress cardiomyopathy is a rare reversible form of left ventricular dysfunction associated clinically with emotional stress, angiographically with "apical ballooning," and pathophysiologically with excess sympathetic activation (Wittstein et al., 2005). This entity remains a diagnosis of exclusion, which mimics ST segment elevation MI (STEMI) on presentation, has a much more favorable clinical outcome than STEMI. Lastly, thyrotoxicosis, Paget's disease and severe chronic anemia are rare causes of high output heart failure. Individuals afflicted with HFPEF are more commonly older age, female gender and have a history of hypertension and atrial fibrillation.

SCREENING

The New York Heart Association (NYHA) functional classification scheme, an older but widely used screening tool, assesses the severity of functional limitations of individuals afflicted with heart failure. The four classes of the NYHA classification are linked to increasing severity of signs and symptoms and correlate well with prognosis. This classification scheme, however, has important limitations since diverse pathophysiological processes leading to symptomatic heart failure are overlooked (Dunselman et al., 1988). Accordingly, the American College of Cardiology and American Heart Association (ACC/AHA) Classification of Chronic Heart Failure was developed to account for the multiple stages and predisposition conditions associated with the clinical syndrome. Designed to encompass emerging scientific evidence, an expert panel periodically assembles these updates, which are the most widely used and authoritative sources on the evaluation, management, performance measures, and outcomes on heart failure (Bonow et al., 2005; Hunt et al., 2005; Radford et al., 2005). In turn, these guidelines incorporating preclinical

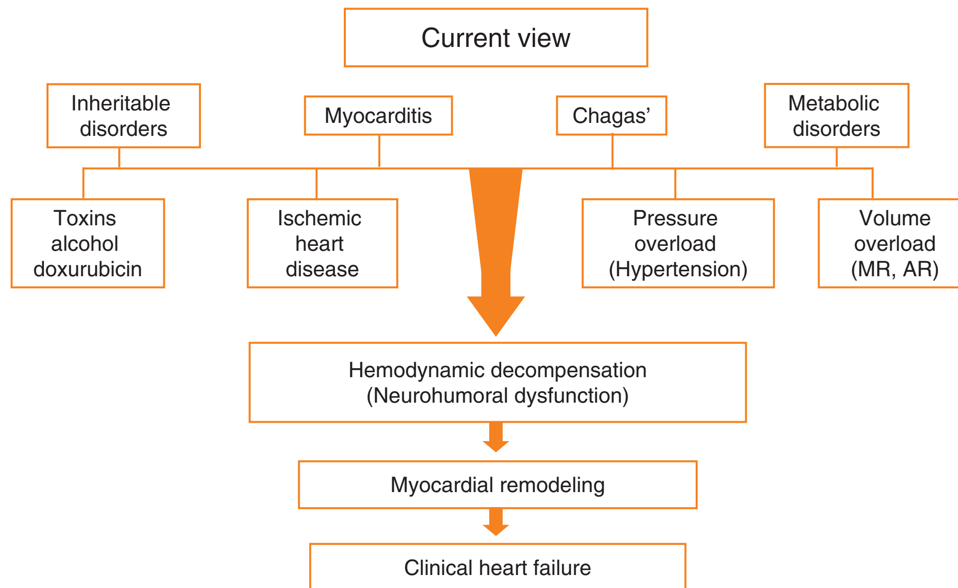


Figure 26.2 The schematic diagram illustrates the different etiologies and multiple compensatory and adaptive pathways implicated in the clinical syndrome of heart failure.

stages, risk factors, pathophysiologic stages, and clinical recognition of heart failure are further subdivided into four stages.

Stage A patients are at high risk for developing heart failure, but have had neither symptoms nor evidence of structural cardiac abnormalities. Major risk factors include hypertension, diabetes mellitus, coronary artery disease and family history of cardiomyopathy. In selected patients, the administration of angio-tensin converting enzyme (ACE) inhibitor is recommended to prevent adverse ventricular remodeling.

Stage B patients have structural abnormalities from previous myocardial infarction, LV dysfunction or valvular heart disease but have remained asymptomatic. Both ACE inhibitors and beta-blockers are recommended.

Stage C patients have evidence for structural abnormalities along with current or previous symptoms of dyspnea, fatigue and impaired exercise tolerance. In addition to ACE inhibitors and beta-blockers, optimal medical regimen may include diuretics, digoxin, and aldosterone antagonists.

Stage D patients have end-stage symptoms of heart failure that are refractory to standard maximal medical therapy. Such patients are candidates for left ventricular assist devices (LVADs) and other sophisticated maneuvers for myocardial salvage or end-of-life care.

PATHOPHYSIOLOGY

Neurohumoral Mechanisms

Low-cardiac output and systemic hypoperfusion elicit a cascade of compensatory mechanisms but predominantly activation of the neurohumoral pathway for augmenting fluid retention (Figure 26.2). Sympathetic nervous system activation increases heart rate and peripheral vasoconstriction from the release of catecholamines, triggering increased afterload and myocardial oxygen consumption. Catecholamines also increase renin secretion, cell death, fibrosis, and myocardial irritability, underlying substrates for lethal arrhythmias and sudden death. In contrast, natriuretic peptides released from specialized cells in the atria exert hormonal actions in distant vascular beds, stimulating vasodilation, and diuresis. Afterload reducing agents and beta-adrenergic blockers have significantly reduced the morbidity and mortality while improving the survival of patients with heart failure. Likewise, antagonists of aldosterone, which promotes salt and water retention, have proven clinical benefits.

Myocardial Remodeling

Left ventricular dysfunction and systolic heart failure secondary to myocardial infarction or ischemia are the prerequisites of low-ejection fraction and elevated pulmonary pressures with congestion. Acquired or inherited conditions that either decrease cardiomyocyte viability and/or increase cell death will ultimately trigger pump failure and symptomatic heart failure. Given the heart's limited capacity for regeneration, terminally differentiated ventricular cardiomyocytes may undergo hypertrophy in response to increase metabolic and homodynamic demands. Activation of the "fetal gene program" orchestrates transcriptional upregulation of genes encoding contractile and cytoskeletal proteins – the prerequisite for compensatory hypertrophy. Recruitment of such adaptive mechanisms provides a variable but stable and asymptomatic interval – perhaps lasting years – before cardiac decompensation. The ensuing ventricular dilatation is a pathologic form of adaptation, termed "ventricular remodeling," affecting intrinsic cardiac mass, the extracellular matrix, collagen deposition and fibrosis as shown in Figures 26.2 and 26.3. Whereas low levels of reactive oxygen species (ROS) serve as stress signals in

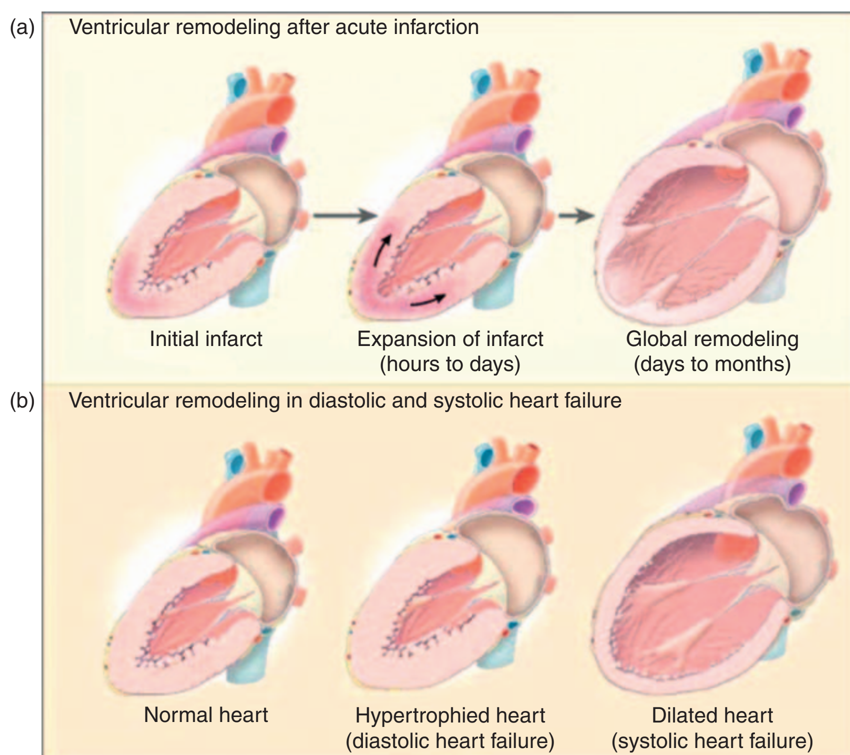


Figure 26.3 Ventricular remodeling after infarction (a) and in diastolic heart failure (b). (Adapted from Jessup et al. *N Engl J Med* 348, 2007–2018.)

redox-dependent regulation, elevated levels of ROS caused by mitochondrial dysfunction may alter myocardial energetics, cardiac metabolism, and trigger the release of cytochrome *c*, thereby activating cell survival/death pathways. Endothelial dysfunction gives rise to the aberrant release of nitric oxide, a potent vasodilator, and/or reactivity with ROS to form peroxynitrite, which causes oxidative damage and cellular injury. Progressive remodeling, in attempts to maintain systolic function and homeostasis (Stage B), leads to valvular regurgitation from inadequate apposition of the mitral leaflets, increasing myocardial stress and, ultimately, decompensated heart failure (Stage C and Stage D).

Mechanisms of Cell Death in Heart Failure

Progressive loss of cardiomyocytes from either necrosis or apoptosis with diverse pathogenic states contributes to the pathogenesis of heart failure as shown in Table 26.1 and Figure 26.4 (Liew and Dzau, 2004; Wencker et al., 2003). Apoptosis or programmed cell death is activated by signaling cascades, via either the extrinsic or intrinsic cell survival/death pathways (Danial and Korsmeyer, 2004). Ligands such as TNF-α, which bind to cognate receptors at the plasma membrane, mediate cell death through the extrinsic pathway, whereas the Bcl-2 family – consisting of both pro- and anti-apoptotic proteins – regulates the intrinsic pathway. Mitochondria play a central role in cell survival/death principally from the initiation of stress signals (e.g., ROS) and release of mitochondrial cytochrome c, which initiates complex formation and the activation of apoptotic proteases (e.g., caspase-9) (Danial and Korsmeyer, 2004). The role of apoptosis in chronic heart failure, which ranges between 80 and 250 myocytes per 100,000 nuclei in failing human hearts was elegantly validated by Wencker and coworkers using transgenic mice harboring a fusion protein FKBP fused with a conditionally active caspase (Wencker et al., 2003). In contrast, low-level inhibition of apoptosis prevented DCM and death, suggesting possible therapeutic strategies for combating heart failure.

DIAGNOSIS

Newly diagnosed patients with heart failure most commonly seek medical attention for either gradual or abrupt onset of the classical signs and symptoms with pulmonary congestion. The clinical spectrum varies widely but dyspnea on exertion, peripheral edema, orthopnea, and paroxysmal nocturnal dyspnea are not uncommon. Exertional chest pain or angina at rest requires an immediate evaluation to determine if biochemical evidence of myocardial damage demands more aggressive management for acute coronary syndromes. Elevated jugular venous distension from right ventricular failure, ascites, and cachexia are more ominous signs for low-cardiac output and decompensation, requiring urgent attention, preferably, from a provider who specializes in heart failure management.

Routine diagnostic studies include an electrocardiogram, chest radiograph, and B-type natriuretic peptide, the latter having the best predictive value for distinguishing between CHF and

TABLE 26.1 The genetic basis of cardiomyopathies

Symbol	Chromosome	Gene product	Cardiomyopathy type
ACTC	15q11–14	Cardiac muscle α-actin	Hypertrophic and dilated
ABCC9	12p12.1	Member 9 of the superfamify C of ATP-binding cassette (ABC) transporters	Dilated
CSRP3, MLP	11p15.1	Cysteine- and glycine-rich protein 3	Dilated
DES	2q35	Desmin	Dilated
DSP	6p24	Desmopfakin	Dilated
LMNA	1q21.2–21,3	Lamin A/C	Dilated
VCL	10q22.1–q23	Metavlnculin	Dilated
MYBPC3	11p11.2	Cardiac myosin-binding protein C	Hypertrophic and dilated
MYH6	14q12	Cardiac muscle a-isoform of myosin heavy chain (heavy polypeptide 6)	Hypertrophic
MYH7	14q12	Cardiac muscle a-isoform of myosin heavy chain (heavy polypeptide 7)	Hypertrophic and dilated
MYL2	12q23–24.3	Myosin regulatory light chain associated with cardiac myosin-β(or slow) heavy chain	Hypertrophic
MYL3	3p21.2–21.3	Myosin light chain 3	Hypertrophic
PLN	6q22.1	Phospholamban	Dilated
PRKAG2	7q35–36	γ2 non-catalytic subunit of AMP-activated protein kinase	Hypertrophic
SGCB	4q12	β-Sarcoglycan (43kDa dystrophin-associated glycoprotein)	Dilated
SGCD	5q33–34	δ-Sarcoglycan (35kDa dystroph in-associated glycoprotein)	Dilated
TAZ, G4.5	Xq28	Tafazzin	Dilated
TTN	2q31	Titin	Hypertrophic and dilated
TCAP	17q12	Titin-cap	Dilated
TPM1	15q22.1	Tropomyosin 1 (α)	Hypertrophic
TNNI3	19q13.4	Troponin I, a subunit of the troponin complex of the thin filaments of striated muscle	Hypertrophic
TNNT2	1q32	Cardiac isoform of troponin T2, tropomyosin-binding subunit of the troponin complex	Hypertrophic and dilated

Adapted from Dzau et al., *Nat Rev Genet* 5, 811–825 (2004).

non-CHF patients (Maisel and McCullough, 2003). Noninvasive echocardiography is the most commonly used diagnostic tool for the assessment and follow-up of patients with heart failure with or without preserved ejection fraction. Coronary angiography should be performed to exclude reversible causes for left ventricular dysfunction or to guide prompt revascularization. If the coronary vessels are widely patent in the setting of global dysfunction, then endomyocardial biopsy should be considered to assess for reversible causes including viral myocarditis (Liu and Mason, 2001). Equilibrium radionucleotide angiography (ERNA) is another noninvasive diagnostic study that assesses both left and right ventricular systolic function. Screening tools such as contrast computer tomographic angiography and magnetic resonance imaging (MRI) are gaining attention as emerging technologies with equivalent sensitivity and specificity as the invasive angiogram for coronary arteriography. MRI may also uncover unsuspected infiltrative cardiomyopathy, arrhythmogenic right ventricular dysplasia, and is superior for the assessment of myocardial viability before revascularization.

PROGNOSIS

Over 5 million Americans or 1.5% of the US population have chronic heart failure, and there is a similar prevalence at risk of undiagnosed left ventricular dysfunction (Braunwald and

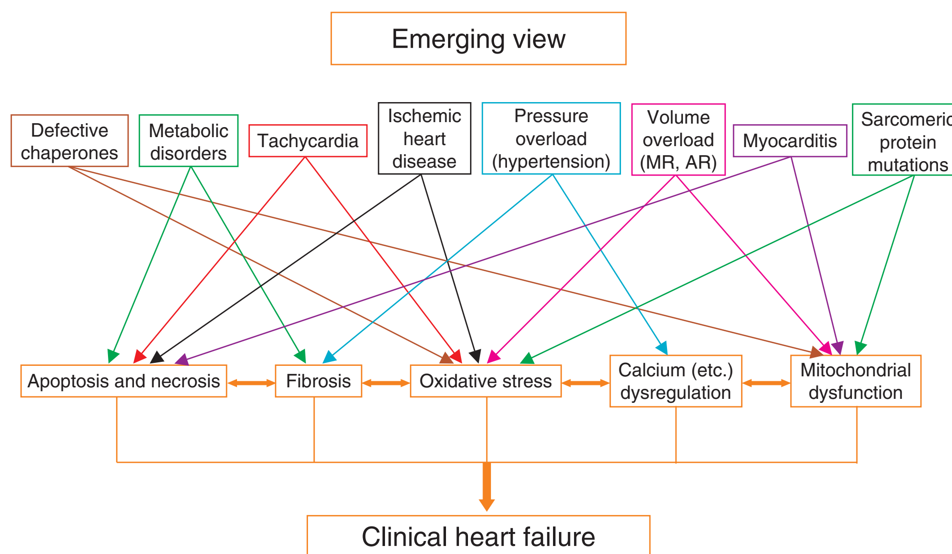


Figure 26.4 At the cellular and molecular levels, crosstalk among pathways related to oxidative stress and calcium dysregulation, for example, may contribute to secondary apoptosis and necrosis. In spite of considerable insights about mechanisms, current therapies focus on reversing neurohumoral imbalances but rarely on underlying mechanisms.

Bristow, 2000). With over 550,000 new cases of heart failure, the disproportionate health and economic burden exceeds 24 billion dollars annually (DiBianco, 2003). Soon, heart failure will become the number one cause of death worldwide, eclipsing infectious disease (Bleumink et al., 2004). Whereas new pharmacologic management and revascularization techniques continue to improve the survival after acute myocardial infarction, the prevalence of chronic heart failure appears to be increasing as the population ages (Braunwald and Bristow, 2000). Notwithstanding, heart failure accounts for 20% of all hospital admissions in patients older than 65, and the hospitalization rate has increased by 159% in the past decade (Jessup and Brozena, 2003). Available treatments for heart failure have only modestly improved the morbidity and mortality (Jessup and Brozena, 2003), and, for patients with advanced heart failure, the prognosis still remains grim with 1-year mortality rates between 20% and 45%, overshadowing the worse forms of some cancers (Jessup and Brozena, 2003).

PHARMACOGENOMICS

Substantial phenotypic heterogeneity in heart failure and the variability of responses among individuals taking pharmacologic agents have been attributed to common polymorphisms in the genome (Liggett, 2001). A surmountable hurdle, however, is the robustness of the association between putative genetic markers and therapeutic response. Recent lessons from studies of human heart failure illustrate handsomely both the enormous potential and challenges for pharmacogenomics – a maturing discipline in which an individual's genetic determinants are used to predict drug response and outcomes (Evans and Relling, 1999; Liggett, 2001). Liggett and coworkers have demonstrated that non-synonymous single-nucleotide polymorphisms of the β_1-adrenergic receptor (β_1-AR), a member of the seven membrane-spanning receptor superfamily, alters the therapeutic response to β-blockers during heart failure (Liggett et al., 2006). Stimulatory effects between β_1-AR and heterotrimeric G proteins G_s mediate both beneficial and deleterious signal transduction pathways during the onset and progression of heart failure. Because β_1-AR is the major subtype in cardiac myocytes, increased catecholamines exert potent cardiomyopathic effects, cardiac remodeling and abrogation of gene expression, which are antagonized by β_1-AR blockers resulting in improved outcomes. A single nucleotide variation at nucleotide 1165 in the gene encoding β_1-AR results in either Arg or Gly at position 389 residue (Liggett et al., 2006). In response to inotropic stimulation, human trabeculae muscle with the β_1-Arg-389 residue from either nonfailing or failing hearts exhibited significantly greater contractility than β_1-Gly-389 polymorphism.

As shown in the β-Blocker Evaluation of Survival Trial (BEST), which evaluated the β-blocker bucindolol for the treatment of Class III/IV, insights into the mechanisms for pharmacogenomic phenotypes involving the Arg/Gly polymorphism of the β_1-AR owe much credit to the DNA Study Group (Feldman et al., 2005; Liggett et al., 2006). The foresight of BEST investigators to recognize the power of genetic haplotyping underscores the importance for all future well-designed human trials to include contingencies for pharmacogenomics in an era of genomic medicine. Notwithstanding the success gleaned from a highly penetrant single-gene trait such as Arg/Gly polymorphism of the β-AR, future advances will require undertaking the more formidable challenges related to multigene traits that influence drug metabolism and response for therapeutic individualization (Evans and McLeod, 2003).

The recent African-American Heart Failure (A-HeFT) enrolled 1050 black patients with NYHA class III or IV to receive a fixed dose of two well-established medications, isosorbide

dinitrate and hydralazine, in a placebo controlled ran-domized multicenter trial. Combination nitrates and hydralazine, termed BiDil, when added to standard therapy was efficacious improving survival in blacks. But the implications of this high-profile study have drawn considerable scientific and ethical scrutiny owing to the marketing strategy of this therapy, which under the proprietary label, was advanced as a novel approach for race-based management. Because physical and genetic traits are not interchangeable, A-HeFT *per se* might prove to be poor surrogate for studies of pharmacogenetics since neither BiDil's efficacy in other racial and ethnic groups nor genetic markers for predicting the response of blacks to BiDil were ever tested.

In contrast, polymorphisms in the ACE pathway have been extensively studied, especially the ACE DD polymorphism, which had significantly higher death and need for transplant compared to II and ID genotypes (McNamara et al., 2001). With concurrent β-blocker treatment, patients with ACE DD polymorphism showed improved survival but benefited with a higher ACE dosage (McNamara et al., 2004), supporting the clinical utility of genetic information in clinical management. For at-risk populations, the pace for moving bidirectional bench ⇔ bedside and bedside ⇔ community-based practices should accelerate using evidence-based strategies emerging from disciplines such as health outcomes.

MONITORING

Genomic Profiling

Heart failure encompasses dynamic processes in which the activation or deactivation of distinct pathways at different stages in the pathogenesis suggest opportunities for intervention and even prevention before irreversible decompensation. A fundamental question, therefore, is how to develop improved diagnostic and prognostic indices that may guide improvements in treatment and outcomes for heart failure. A major goal of microarray-based analyses is to identify genes whose similar patterns of expression accurately represent the disease state or biological process. Such information, however, is often insufficient to identify the causal mechanisms but provides a comprehensive picture of the underlying process, which can predict responses to therapy or disease stage.

Both unsupervised and supervised approaches are applied to determine if previously unrecognized or unexpected patterns of expression exist in the datasets. Hierarchical clustering, for example, is an unsupervised approach that may be used after gene expression profiling to identify interdependent pathways before the onset of overt heart failure. Identification and validation of genes or novel pathways that are activated earliest may improve early detection and, ultimately, will be essential for designing therapies that prevent the natural history and progression of disease. If individual genes have different predictive power, then a "weighted voting scheme," based on the levels of gene expression, can be designed and tested before widespread application. Considerable caution in data interpretation is warranted, however, as comparisons from different laboratories may be skewed considerably by patient selection, different treatments, and clinical stages.

While many technical details and microarrays are beyond the scope of this chapter, considerable attention is being paid to standardization of data collection, normalization, and data reporting. Most peer-reviewed journals have instituted policies for primary sequence databases to be deposited before publication using the "minimal information about a microarray experiment" or MIAME standards (Ball et al., 2002; Ball et al., 2004; Brazma et al., 2001). We predict that meta analyses of large datasets hold particular promise for finding universal themes for classification, clinical staging, and predicting outcomes from gene expression profiles.

Transcriptional Profiling of Heart Failure

Neurohumoral, hemodynamic and environmental factors participate in remodeling the failing heart, but genetic, molecular, and cellular events are inscribed at the transcriptional level. Signaling pathways and biological processes implicated in the hypertrophic response of the heart are shown in Figure 26.4. Early genetic markers of cardiac hypertrophy include transcriptional reprogramming of genes encoding contractile proteins, oncogenes, neurohumoral factors, and transcription factors have been identified. Genes encoding proto-oncogenes c-jun, c-fos, c-myc, skeletal α-actin and ANF are also activated in response to hypertrophic stimuli. (Izumo et al., 1988; Komuro et al., 1988; Mulvagh et al., 1987; Schwartz et al., 1986). In angiotensin II receptor type 1alpha knockout mice, cardiomyocytes were capable of evoking increased protein synthesis and mitogen-activated protein kinase (MAPK) activation when stretched, strengthening the primary role of mechanical stretch in maintaining the hypertrophic phenotype. The mechanisms by which mechanical stress is converted into biological response are yet to be fully elucidated. High-density oligonucleotide arrays have also identified multiple genes, representing diverse biological process (e.g., myocardial structure, myocardial assembly and degradation, metabolism, protein synthesis, and stress response), which were differentially expressed in nonfailing and failing human hearts (Yang et al., 2000). Other larger studies of human heart failure have confirmed the role for MAPKs, mechanical stress and neurohumoral pathways in heart failure (Kudoh et al., 1998). Likewise, genetic pathways identified during acute and chronic pressure overload reflect differential gene expression during distinct phases may represent potential target for therapy. A substantial limitation, however, remains that lack of reproducibility and reliability in the sample sets owing to selection bias and differences in etiology, age, sex, mode of onset, treatment regimens and clinical course.

End-stage heart failure is associated with an increased activity and alterations of multiple gene products including the extracellular matrix/cytoskeletal (e.g., collagen types I and III, fibromodulin, fibronectin, and connexin 43) (Tan et al., 2002). When gene expression profiling was applied in a transgenic model of tumor necrosis factor-α overexpression, a large number of immune response-related genes were over-expressed, along with a IgG deposition in myocardium, supporting activation of immune system and inflammatory mechanisms in the development and

progression of heart failure (Feldman and McTiernan, 2004; Kubota et al., 1997).

Gene expression profiles of heart failure caused by alcoholic cardiomyopathy and familial cardiomyopathy suggest that the onset and disease progression may involve different genetic determinants. A provocative study by Blaxall and coworkers has reported that genomic profiling in a murine model of heart failure reverted to the normal phenotype after rescue by expression β-adrenergic receptor kinase and suggested mild and advanced heart failure maybe similar in mice and humans (Blaxall et al., 2003). As previously mentioned, the conclusions must be interpreted cautiously owing to the complexities of heart failure related to the imprecision associated with genetic, physiologic, and clinical phenotypes.

A Case for Biologic Reclassification of Heart Failure

Gene expression profiling has significantly improved the diagnostic classification of specific conditions (e.g., breast cancer, chronic myelogenous leukemia) but remains a formidable challenge for deciphering meaningful insights about the biological mechanisms underlying disease pathogenesis (Quackenbush, 2006). Among inheritable forms of cardiovascular diseases, recent advances of single-gene disorders have fundamentally altered our understanding about the cellular processes, metabolic alterations and transcriptional reprogramming of the diseased heart (Seidman and Seidman, 2001). Much like the success seen for tumor classification and other improvements in cancer therapeutics (Bell, 2004; Quackenbush, 2006), and beyond the availability of genetic tests for disease-causing mutations of cardiomyopathy (Morita et al., 2005), the development of genomic tools that are causally linked to disease pathogenesis, termed a "molecular signature," will likely accelerate progress for early detection, targeted therapy and disease monitoring of inheritable heart failure (Bell, 2004). We suggest that the opportunities exist for microarray-based profiling, proteomics, metabolomics, and genome-wide technologies to propel the transition from clinicopathologic to clinico-genomic classifications for heart failure.

Different gene profiles for failing and nonfailing hearts have already permitted differentiation among heart failure with different etiologies, as shown recently by Donahue and colleagues in Table 26.2 (Donahue et al., 2006). Considerable discordance, however, exists as our ability to diagnose heart failure using genomic profile lags substantially behind clinical management. Important obstacles remain, such as limitations in procuring tissue samples needed for genomic profiles and their transition from use as research tools into the realm of clinical diagnostics.

Biomarkers Versus Biosignatures for Heart Failure

Considerable biological heterogeneity of heart failure demands more robust tools to guide clinical outcomes. Much recent attention has focused on biological markers, or biomarkers, which objectively measure and evaluate normal biological processes, pathologic process, or pharmacological response to therapeutic intervention (Vasan, 2006). Current enthusiasm for biomarker strategies, however, has also brought confusion and ambiguity for applications in clinical practice. Too often, highly fragmented information obtained from patients at different clinical stages precludes meaningful analysis and extrapolation to broader subclasses.

Accordingly, we propose that an integrative approach – that encompasses our ability to predict the onset, rate of progression, and response to therapy and/or clinical outcome with reproducibility and reliability may circumvent such limitation of biomarkers – requires the development of a molecular signature or "biosignature." In order to circumvent existing limitations of biomarkers, proof-of-concept for such biosignatures, however, may require tissue sampling and serial analysis for identification and validation (Figure 26.5). Among eight individuals with IDCM but with similar clinical characteristics for chronic heart failure at baseline, Lowes and colleagues reported that serial sampling was superior to cross-sectional gene expression profiling since there was less variance in the differences on gene chip analysis of endomyocardial biopsies (EMB) from the same patient than among the different subjects with similar phenotypes (Lowes et al., 2006). Because these biological processes may precede the transition into heart failure and premature death, future work should address the intriguing possibility that distinct metabolic pathways might be linked to novel molecular signatures in disease pathogenesis (Figure 26.5).

Protein aggregation cardiomyopathy (PAC) (also termed desmin-related myopathy – DRM) is a multi-system disease, caused by the missense R120G mutation in the gene encoding the human small HSP αB-crystallin (hR120GCryAB). To understand the pathogenic mechanisms by which hR120GCryAB expression causes cardiotoxicity and heart failure, our group has shown in recent genetic studies in mice that selective hR120GCryAB expression in the heart induces a novel toxic gain-of-function mechanism involving reductive stress, apparently emanating from increased activity of glucose 6-phosphate dehydrogenase (G6PD) (Rajasekaran et al., 2007). Reductive stress refers to an abnormal increase in the amounts of reducing equivalents (e.g., glutathione, NADPH), which has been demonstrated in lower eukaryotes (Simons et al., 1995; Trotter and Grant, 2002) but has not been commonly shown in the mammals and/or in disease states (Chance et al., 1979; Gores et al., 1989).

Such genetic evidence, that dysregulation of G6PD activity is a causal mechanism for R120GCryAB cardiomyopathy, forms the rationale for ideas related to metabolic and genetic pathways that might codify biosignatures. What metabolic changes occur before the onset of detectable myopathic or pathologic alterations, and how does such imbalance contribute to cardiomyopathy and heart failure? Does reductive stress exert direct or indirect consequences on mitochondrial (dys)function? Applications in redox proteomics and multiplex protein markers are presently being pursued to determine if glutathionylation, for example, of key components in mitochondrial and other metabolic pathways are causally linked to disease pathogenesis.

Molecular Diagnosis of Allograft Rejection

Although peripheral blood mononuclear cells (PBMCs) are abundant and highly accessible sources of genomic material, a potential for diagnostic inaccuracy and therapeutic failure exists

Table 26.2 Discovery projects in heart failure

Comparison	Subjects	Platform	Finding
Failing versus non-failing	2 cases (1 ICM and 1 DCM) 2 control cases	Affymetrix Hu 6800	Alterations of expression of cytoskeletal and myofibrillar genes, genes encoding stress proteins, and genes involved in metabolism, protein synthesis, and protein degradation
Failing versus non-failing	7 cases (DCM) 5 control cases	Cardiochip (custom array)	Upregulation of genes for atrial natriuretic peptide, sarcomeric and cytoskeletal proteins, stress proteins, and transcription/translation regulators Down-regulation of genes regulating calcium signaling pathways
Failing versus non-failing	8 cases (DCM) 7 control cases	Affymetrix Hu 6800	103 differentially expressed genes with most prominent being atrial natriuretic factor and brain natriuretic peptide
Failing versus non-failing	10 cases (DCM) 4 control cases	Custom arrays	364 differentially expressed genes Up-regulation being most prominent in genes for energy pathways, muscle contraction, electron transport, and intracellular signaling Down-regulation was most prominent in genes for cell cycle control
Failing versus non-failing	9 cases (5 ICM and 4 DCM) 1 control case	Affymetrix HG-U95A	95 differentially expressed genes with notable up-regulation of atrial natriuretic peptide and brain natriuretic peptide Prominent pathways up-regulated include cell signaling and muscle contraction
Failing versus non-failing	6 cases (DCM) 5 control cases	Affymetrix HG-U133A	165 differentially expressed genes, the most prominent being structural and metabolic genes
Failing versus non-failing	5 cases (DCM) 5 control cases	Custom array for apoptotic pathways	Differentially expressed genes in apoptotic pathways
Pre- and post-left ventricular assist device	6 cases (3 DCM and 3 ICM)	Affymetrix Hu 6800	530 differentially expressed genes (295 up and 235 down) with prominent changes in genes for metabolism
Pre- and post-left ventricular assist device	7 cases (DCM)	Affymetrix HG-U133A	179 differentially expressed genes (130 up and 49 down There was prominent up-regulation in nitric oxide pathways and down-regulation of inflammatory genes
Pre- and post-left ventricular assist device	19 cases (8 DCM and 11 ICM)	Affymetrix HG-U133A	107 differentially regulated genes (85 up and 22 down) Prominent was the up-regulation of genes regulating vascular networks and down-regulation of genes regulating myocyte hypertrophy
HCM and DCM versus non-failing	3 DCM 2 HCM 3 control cases	Cardiochip (custom array)	Multiple genes and pathways up- and down-regulated some common to DCM and HCM some distinct to each

DCM = dilated cardiomyopathy; HCM = hypertrophic cardiomyopathy; ICM = ischemic cardiomyopathy.

if there is discordance between the information in PBMCs and underlying condition in the diseased tissues. Significant progress has been made for patients after cardiac transplantation, which could change existing paradigms for clinical decision-making and management of allograft rejection. Standard protocols after heart transplantation requires patients to undergo serial EMB as a means to monitor for rejection and to guide immunosuppressive therapy. Such surveillance maneuvers are invasive, expensive and carry considerable risks such as perforation of the ventricular wall and hemopericardium. Analysis of the histological data by expert pathologists is subject to inter-observer variability and the diagnosis of acute rejection has been controversial (Nielsen et al., 1993; Winters and McManus, 1996). Horwitz and colleagues were among the first to demonstrate that gene expression

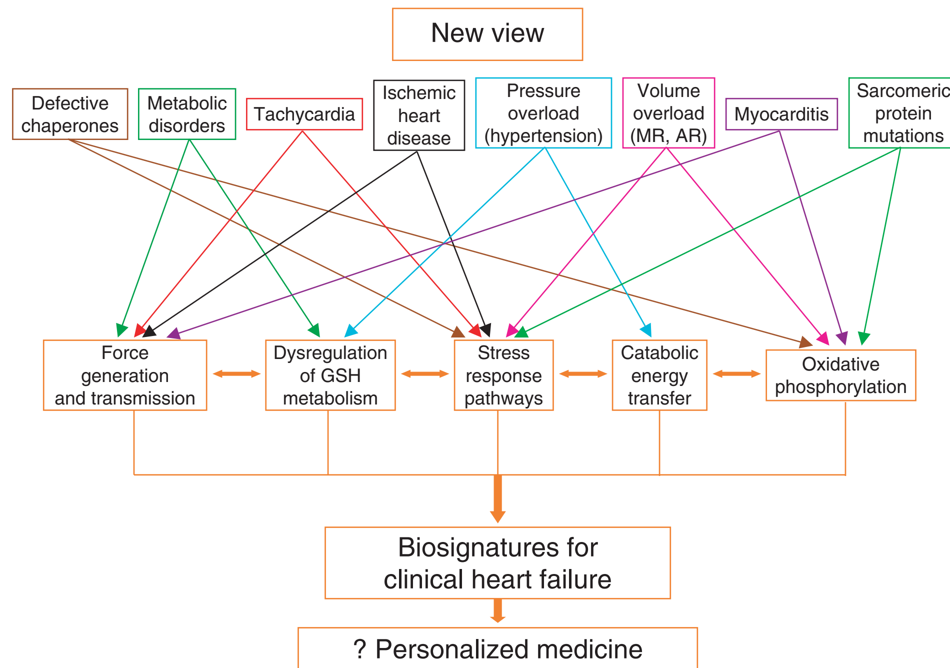


Figure 26.5 New diagnostic approaches, based on information that integrates genes and molecular pathways at the onset, progression and end stages are needed to improve heart failure classification. "Biosignatures" for heart failure – developed from microarray analysis technologies, proteomics and genomic technologies – are proposed here to integrate the biological processes and molecular mechanisms for rationale drug design and treatment in the post genomic era of personalized medicine.

profiles of PBMCs might provide an alternative approach of the diagnosis of allograft rejection (Horwitz et al., 2004). Patients who subsequently developed acute rejection had a distinct genomic profile compared with patients without any rejection and, after treatment for rejection, the majority (98%) of differentially expressed genes returned to baseline.

The CARGO (Cardiac Allograft Rejection Gene Expression Observational) study prospectively investigated gene expression analysis from PMBCs as a diagnostic tool to predict transplant rejection (Mehra, 2005). From the core group of 11 genes associated with immune response pathways, which were identified by quantitative real-time polymerase chain reaction (QT-PCR) and assigned weighted scores, CARGO investigators were able to predict rejection with a sensitivity and specificity of 80% and 60%, respectively (Deng et al., 2006). Owing to reduced sensitivity and specificity immediately after transplantation, the test may also be unreliable for the diagnosis of low/intermediate grade rejection. Now commercially available (AlloMap®), this landmark study provides proof-of-concept that gene expression profiling in PBMCs are predictive for acute rejection pathways in cardiac transplant patients. One important implication is that genomic profiling of specific targets expressed in peripheral blood will increasingly be used as a sensitive marker for transplant rejection but direct evidence that such monitoring should guide therapeutic management awaits further independent validation.

2009 UPDATE

Identification and characterization of the pathways and mechanisms by which common genetic variants (polymorphisms) contribute to the initiation and propagation of life-threatening conditions remain the foundations for personalized medicine in cardiology. Among established targets of neurohormonal antagonists associated with mortality benefits for heart failure, a pharmacogenomic interaction is suggested where common sequence variations of genes encoding neurohormonal signaling alter an individual's therapeutic response (Dorn, 2007). Epigenetic alterations are now being considered in light of such effects on the development of cardiac hypertrophy and heart failure. For example, recent analysis of ventricular tissue from both experimental and human heart failure has revealed epigenetic changes such as trimethylation of histone H3 at lysine-4

or lysine-9, which were present between the regions of genes encoding signal transduction pathways (Kaneda et al., 2009).

Autocrine and paracrine factors elaborated by the heart serve as key sensors that self-regulate cardiomyocyte contraction, growth, and programmed death (Dorn, 2007). For the irreversibly damaged myocardium, the clinical trials of adult stem cell therapy have suggested potential benefit but there is less confidence in the current approaches being used for regeneration of cardiac tissue (Mayorga et al., 2009). A recent catalog of proteotypic peptides in heart human has been developed to serve as a public reference to guide future targeted mass spectrometry experiments monitoring protein biomarkers (Edwards et al., 2008; Kline et al., 2008). The discovery of human gene polymorphisms and/or gene variant expression holds particular promise as therapeutic targets for ameliorating adaptive and maladaptive mechanisms related to stress-response pathways.

2009 UPDATE REFERENCES

Dorn, G.W. (2007). The Gordon Wilson lecture: Neurohormonal signaling pathways that link cardiac growth and death. *Trans Am Clin Climatol* 118, 137–152.

Edwards, A.V., White, M.Y. and Cordwell, S.J. (2008). The role of proteomics in clinical cardiovascular biomarker discovery. *Mol Cell Proteomics* **7**(10), 1824–1837.

Kaneda, R., Takada, S., Yamashita, Y., Choi, Y.L., Nonaka-Sarukawa, M., Soda, M., Misawa, Y., Isomura, T., Shimada, K. and Mano, H. (2009). Genome-wide histone methylation profile for heart failure. *Genes Cells* 14(1), 69–77.

Kline, K.G., Frewen, B., Bristow, M.R., Maccoss, M.J. and Wu, C.C. (2008). High quality catalog of proteotypic peptides from human heart. *J Proteome Res* **7**(11), 5055–5061.

Mayorga, M., Finan, A. and Penn, M. Pre-transplantation specification of stem cells to cardiac lineage for regeneration of cardiac tissue. *Stem Cell Rev* 2009; [Epub ahead of print].

ACKNOWLEDGEMENTS

An award from NHLBI (5R01 HL63874) and Christi T. Smith Foundation provided support for this work. Krista Boeger and Lori Kaumans provided excellent editorial assistance during preparation of this manuscript.

REFERENCES

Ball, C.A., Brazma, A., Causton, H., Chervitz, S., Edgar, R., Hingamp, P., Matese, J.C., Parkinson, H., Quackenbush, J., Ringwald, M. et al. (2004). Submission of microarray data to public repositories. *PLoS Biol* 2, E317.

Ball, C.A., Sherlock, G., Parkinson, H., Rocca-Sera, P., Brooksbank, C., Causton, H.C., Cavalieri, D., Gaasterland, T., Hingamp, P., Holstege, F. et al. (2002). Standards for microarray data. *Science* 298, 539.

Bell, J. (2004). Predicting disease using genomics. *Nature* 429, 453–456.

Benjamin, I.J. and Schneider, M.D. (2005). Learning from failure: congestive heart failure in the postgenomic age. *J Clin Invest* 115, 495–499.

Bhatia, R.S., Tu, J.V., Lee, D.S., Austin, P.C., Fang, J., Haouzi, A., Gong, Y. and Liu, P.P. (2006). Outcome of heart failure with preserved ejection fraction in a population-based study. *New Engl J Med* 355, 260–269.

Blaxall, B.C., Spang, R., Rockman, H.A. and Koch, W.J. (2003). Differential myocardial gene expression in the development and rescue of murine heart failure. *Physiol Genomics* 15, 105–114.

Bleumink, G.S., Knetsch, A.M., Sturkenboom, M.C., Straus, S.M., Hofman, A., Deckers, J.W., Witteman, J.C. and Stricker, B.H. (2004). Quantifying the heart failure epidemic: prevalence, incidence rate, lifetime risk and prognosis of heart failure. The Rotterdam Study. *Eur Heart J* 25, 1614–1619.

Bonow, R.O., Bennett, S., Casey, D.E., Jr., Ganiats, T.G., Hlatky, M.A., Konstam, M.A., Lambrew, C.T., Normand, S.L., Pina, I.L., Radford, M.J. et al. (2005). ACC/AHA clinical performance measures for adults with chronic heart failure: a report of the American College of Cardiology/American Heart Association Task Force on Performance Measures (Writing Committee to Develop Heart Failure Clinical Performance Measures): endorsed by the Heart Failure Society of America. *Circulation* 112, 1853–1887.

Braunwald, E. and Bristow, M.R. (2000). Congestive heart failure: fifty years of progress. *Circulation* 102, IV14–IV23.

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C. et al. (2001). Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Gen* 29, 365–371.

Danial, N.N. and Korsmeyer, S.J. (2004). Cell death: critical control points. *Cell* 116, 205–219.

Deng, M.C., Eisen, H.J., Mehra, M.R., Billingham, M., Marboe, C.C., Berry, G., Kobashigawa, J., Johnson, F.L., Starling, R.C., Murali, S. et al. (2006). Noninvasive discrimination of rejection in cardiac allograft recipients using gene expression profiling. *Am J Transplant* 6, 150–160.

DiBianco, R. (2003). Update on therapy for heart failure. *Am J Med* 115, 480–488.

Dunselman, P.H., Kuntze, C.E., van Bruggen, A., Beekhuis, H., Piers, B., Scaf, A.H., Wesseling, H. and Lie, K.I. (1988). Value of New York Heart Association classification, radionuclide ventriculography,

and cardiopulmonary exercise tests for selection of patients for congestive heart failure studies. *Am Heart J* 116, 1475–1482.

Evans, W.E. and McLeod, H.L. (2003). Pharmacogenomics – drug disposition, drug targets, and side effects. *New Eng J Med* 348, 538–549.

Evans, W.E. and Relling, M.V. (1999). Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* 286, 487–491.

Feldman, A.M. and McTiernan, C. (2004). Is there any future for tumor necrosis factor antagonists in chronic heart failure? *Am J Cardiovasc Drugs* 4, 11–19.

Feldman, D.S., Carnes, C.A., Abraham, W.T. and Bristow, M.R. (2005). Mechanisms of disease: beta-adrenergic receptors – alterations in signal transduction and pharmacogenomics in heart failure. *Nat Clin Pract* 2, 475–483.

Horwitz, P.A., Tsai, E.J., Putt, M.E., Gilmore, J.M., Lepore, J.J., Parmacek, M.S., Kao, A.C., Desai, S.S., Goldberg, L.R., Brozena, S.C. et al. (2004). Detection of cardiac allograft rejection and response to immunosuppressive therapy with peripheral blood gene expression. *Circulation* 110, 3815–3821.

Hunt, S.A., Abraham, W.T., Chin, M.H., Feldman, A.M., Francis, G.S., Ganiats, T.G., Jessup, M., Konstam, M.A., Mancini, D.M., Michl, K. et al. (2005). ACC/AHA 2005 Guideline Update for the Diagnosis and Management of Chronic Heart Failure in the Adult: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Update the 2001 Guidelines for the Evaluation and Management of Heart Failure): developed in collaboration with the American College of Chest Physicians and the International Society for Heart and Lung Transplantation: endorsed by the Heart Rhythm Society. *Circulation* 112, e154–e235.

Izumo, S., Nadal-Ginard, B. and Mahdavi, V. (1988). Protooncogene induction and reprogramming of cardiac gene expression produced by pressure overload. *Proc Natl Acad Sci USA* 85, 339–343.

Jessup, M. and Brozena, S. (2003). Heart failure. *New Engl J Med* 348, 2007–2018.

Komuro, I., Kurabayashi, M., Takaku, F. and Yazaki, Y. (1988). Expression of cellular oncogenes in the myocardium during the developmental stage and pressure-overloaded hypertrophy of the rat heart. *Circ Res* 62, 1075–1079.

Kubota, T., McTiernan, C.F., Frye, C.S., Slawson, S.E., Lemster, B.H., Koretsky, A.P., Demetris, A.J. and Feldman, A.M. (1997). Dilated cardiomyopathy in transgenic mice with cardiac-specific overexpression of tumor necrosis factor-alpha. *Cir Res* 81, 627–635.

Kudoh, S., Komuro, I., Hiroi, Y., Zou, Y., Harada, K., Sugaya, T., Takekoshi, N., Murakami, K., Kadowaki, T. and Yazaki, Y. (1998). Mechanical stretch induces hypertrophic responses in cardiac myocytes of angiotensin II type 1a receptor knockout mice. *J Biol Chem* 273, 24037–24043.

Liew, C.C. and Dzau, V.J. (2004). Molecular genetics and genomics of heart failure. *Nat Rev Genet* 5, 811–825.

Liggett, S.B. (2001). Pharmacogenetic applications of the Human Genome project. *Nat Med* 7, 281–283.

Liggett, S.B., Mialet-Perez, J., Thaneemit-Chen, S., Weber, S.A., Greene, S.M., Hodne, D., Nelson, B., Morrison, J., Domanski, M.J., Wagoner, L.E. et al. (2006). A polymorphism within a conserved beta(1)-adrenergic receptor motif alters cardiac function and beta-blocker response in human heart failure. *Proc Natl Acad Sci USA* 103, 11288–11293.

Liu, P.P. and Mason, J.W. (2001). Advances in the understanding of myocarditis. *Circulation* 104, 1076–1082.

Lowes, B.D., Zolty, R., Minobe, W.A., Robertson, A.D., Leach, S., Hunter, L. and Bristow, M.R. (2006). Serial gene expression profiling in the intact human heart. *J Heart Lung Transplant* 25, 579–588.

Maisel, A.S. and McCullough, P.A. (2003). Cardiac natriuretic peptides: a proteomic window to cardiac function and clinical management. *Rev Cardiovas Med* 4(Suppl 4), S3–S12.

Mehra, M.R. (2005). The emergence of genomic and proteomic biomarkers in heart transplantation. *J Heart Lung Transplant* 24, S213–S218.

Morita, H., Seidman, J. and Seidman, C.E. (2005). Genetic causes of human heart failure. *J Clin Invest* 115, 518–526.

Mulvagh, S.L., Michael, L.H., Perryman, M.B., Roberts, R. and Schneider, M.D. (1987). A hemodynamic load in vivo induces cardiac expression of the cellular oncogene, c-myc. *Biochem Biophys Res Commun* 147, 627–636.

Nielsen, H., Sorensen, F.B., Nielsen, B., Bagger, J.P., Thayssen, P. and Baandrup, U. (1993). Reproducibility of the acute rejection diagnosis in human cardiac allografts. The Stanford Classification and the International Grading System. *J Heart Lung Transplant* 12, 239–243.

Owan, T.E., Hodge, D.O., Herges, R.M., Jacobsen, S.J., Roger, V.L. and Redfield, M.M. (2006). Trends in prevalence and outcome of heart failure with preserved ejection fraction. *New Engl J Med* 355, 251–259.

Park, J., Choe, S.S., Choi, A.H., Kim, K.H., Yoon, M.J., Suganami, T., Ogawa, Y. and Kim, J.B. (2006). Increase in glucose-6-phosphate dehydrogenase in adipocytes stimulates oxidative stress and inflammatory signals. *Diabetes* 55, 2939–2949.

Quackenbush, J. (2006). Microarray analysis and tumor classification. *New Engl J Med* 354, 2463–2472.

Radford, M.J., Arnold, J.M., Bennett, S.J., Cinquegrani, M.P., Cleland, J.G., Havranek, E.P., Heidenreich, P.A., Rutherford, J.D., Spertus, J.A., Stevenson, L.W. et al. (2005). ACC/AHA key data elements and definitions for measuring the clinical management and outcomes of patients with chronic heart failure: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Data Standards (Writing Committee to Develop Heart Failure Clinical Data Standards): developed in collaboration with the American College of Chest Physicians and the International Society for Heart and Lung Transplantation: endorsed by the Heart Failure Society of America. *Circulation* 112, 1888–1916.

Schwartz, K., de la Bastie, D., Bouveret, P., Oliviero, P., Alonso, S. and Buckingham, M. (1986). Alpha-skeletal muscle actin mRNA's accumulate in hypertrophied adult rat hearts. *Cir Res* 59, 551–555.

Seidman, J.G. and Seidman, C. (2001). The genetic basis for cardiomyopathy: from mutation identification to mechanistic paradigms. *Cell* 104, 557–567.

Tan, F.L., Moravec, C.S., Li, J., Apperson-Hansen, C., McCarthy, P.M., Young, J.B. and Bond, M. (2002). The gene expression fingerprint of human heart failure. *Proc Natl Acad Sci USA* 99, 11387–11392.

The International HapMap Consortium (2005). A haplotype map of the human genome. *Nature* 437, 1299–1320.

Vasan, R.S. (2006). Biomarkers of cardiovascular disease: molecular basis and practical considerations. *Circulation* 113, 2335–2362.

Wencker, D., Chandra, M., Nguyen, K., Miao, W., Garantziotis, S., Factor, S.M., Shirani, J., Armstrong, R.C. and Kitsis, R.N. (2003). A mechanistic role for cardiac myocyte apoptosis in heart failure. *J Clin Invest* 111, 1497–1504.

Winters, G.L. and McManus, B.M. (1996). Consistencies and controversies in the application of the International Society for Heart and Lung Transplantation working formulation for heart transplant biopsy specimens. Rapamycin Cardiac Rejection Treatment Trial Pathologists. *J Heart Lung Transplant* 15, 728–735.

Wittstein, I.S., Thiemann, D.R., Lima, J.A., Baughman, K.L., Schulman, S.P., Gerstenblith, G., Wu, K.C., Rade, J.J., Bivalacqua, T.J. and Champion, H.C. (2005). Neurohumoral features of myocardial stunning due to sudden emotional stress. *New Engl J Med* 352, 539–548.

Yang, J., Moravec, C.S., Sussman, M.A., DiPaola, N.R., Fu, D., Hawthorn, L., Mitchell, C.A., Young, J.B., Francis, G.S., McCarthy, P.M. et al. (2000). Decreased SLIM1 expression and increased gelsolin expression in failing human hearts measured by high-density oligonucleotide arrays. *Circulation* 102, 3046–3052.

CHAPTER

Genomic Assessment of Cardiac Transplant Rejection

Michael Pham, Mario C. Deng, Jay Wohlgemuth and Thomas Quertermous

INTRODUCTION

The development of highly parallel assessment of genome-wide patterns of gene expression and the development of statistical and bioinformatics algorithms for evaluation of these large datasets have provided great insight into the molecular basis of physiology and disease. While less widely appreciated, it was apparent early on that large gene expression databases, if collected in a clinically and epidemiologically appropriate fashion, could provide the basis for diagnosis and monitoring tools capable of informing on human disease. Attention has focused on tissue samples obtained through biopsy for cancer profiling and monitoring and on gene expression in circulating leukocytes to gain information about the broad array of immune-associated diseases. Although technically demanding with respect to reproducibility of measuring changes in gene expression and appropriate handling of data, gene expression measurements have nonetheless provided the basis for a small number of clinically useful and recognized tests that are now available to physicians for disease monitoring, augmenting more classical laboratory data to guide patient management decisions.

CARDIAC ALLOTRANSPLANTATION AS A DEFINITIVE THERAPY FOR END-STAGE HEART FAILURE

Heart failure (HF) is a leading cause of morbidity and mortality in the United States. In 2001, approximately five million Americans were affected by this condition, and over half a million new cases are diagnosed annually. The disease accounts for the number one cause of hospitalization for individuals aged 65 or greater. In patients with end-stage HF and severe functional limitation, 1-year mortality approaches 75%, rivaling the effects of most cancers (Rose et al., 2001). Despite major advances in the treatment of this disease over the past decade, a sizable number of patients with terminal or progressive myocardial dysfunction are fated to die or to be severely limited by symptoms. In these patients, biological replacement of the heart with human organs (allotransplantation) has become standard therapy and is widely accepted as a modality for prolonging life and improving its quality in carefully selected patients. Currently, approximately 4000 heart transplants are performed worldwide each year with survival rates of 81%, 74%, and 68% at 1, 3, and 5 years, respectively (Taylor et al., 2006).

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

THE PROBLEM OF ALLOGRAFT REJECTION

Heart transplantation is the definitive therapy for end-stage HF but is complicated by recipient immune responses to the transplanted heart (allograft), often resulting in cardiac allograft rejection. Different forms of rejection are recognized, including hyperacute rejection (Kemnitz et al., 1991), acute cellular rejection (ACR) (Billingham et al., 1990; Rodriguez, 2003; Stewart et al., 2005; Winters and McManus, 1996; Winters et al., 1998), acute antibody-mediated rejection (AMR), mixed rejection (Book et al., 2003; Michaels et al., 2003a), and chronic rejection, also termed cardiac allograft vasculopathy (CAV). Hyperacute rejection is rare but may occur in the setting of circulating-preformed antibodies to the ABO blood group (in cases of ABO blood group incompatability) or to major histocompatability antigens on the cardiac allograft. This form of rejection manifests as severe graft failure within the first few minutes to hours after transplantation. ACR is the most common form of rejection and occurs in 30–50% of heart transplant recipients in the first year following transplantation (Hershberger et al., 2005). Most episodes occur within the first 3–6 months. This type of rejection is primarily mediated by T-lymphocytes and is characterized by lymphocytic infiltration within the myocardium, direct myocyte damage, and subsequent graft dysfunction. The severity of cellular rejection reflects the distribution and extent of inflammation and the presence or absence of myocyte damage. Acute AMR is mediated by B-lymphocytes and is characterized by antibody deposition on the cardiac allograft microvasculature, complement activation, and graft dysfunction. It is more likely to be associated with hemodynamic compromise compared with ACR, carries a worse prognosis, and is a strong risk factor for the early development of CAV (Michaels et al., 2003b). The prevalence of AMR has been reported to be between 15% and 20%, and it can occur independently of or in combination with ACR (Hammond et al., 1989; Michaels et al., 2003a). Chronic rejection occurs months to years after transplantation and is typically manifested as CAV and late graft failure. The mechanisms of chronic rejection are incompletely understood but may involve a proliferative response to both immunologically and nonimmunologically mediated endothelial injury with progressive intimal thickening within the coronary vessels.

IMMUNOSUPPRESSION STRATEGIES TO PREVENT REJECTION

Most clinically used immunosuppressive regimens consist of a combination of several agents used concurrently and utilize several general principles. The first general principle is that immune reactivity and tendency toward graft rejection are highest early (within the first 3–6 months) after graft implantation and decrease with time. Thus, most regimens employ the highest levels of immunosuppression immediately after transplantation and decrease those levels over the first year, eventually settling on the lowest maintenance levels of immune suppression that are compatible with preventing graft rejection and minimizing drug toxicities. The second general principle is to use low doses of several drugs without overlapping toxicities in preference to higher (and more toxic) doses of fewer drugs whenever feasible. The third principle is that too much or too intense immunosuppression is undesirable, because it leads to a myriad of undesirable effects, such as susceptibility to infection and malignancy. Finding the right balance between over- and under-immunosuppression in an individual patient is truly an art that utilizes science.

Most modern immunosuppressive protocols employ a three-drug regimen consisting of a calcineurin inhibitor (cyclosporine or tacrolimus), an antiproliferative agent (mycophenolate mofetil or azathioprine), and corticosteroids. The calcineurin inhibitors inhibit production of the cytokine interleukin-2, therefore preventing T-lymphocyte differentiation and proliferation. The antiproliferative agents exert their immunosuppressive effects by blocking purine synthesis and inhibiting proliferation of both T- and B-lymphocytes. Corticosteroids are non-specific anti-inflammatory agents that interrupt multiple steps in immune activation, including antigen presentation, cytokine production, and proliferation of lymphocytes. They are used in relatively high doses in the early post-transplant period and are tapered to low doses or discontinued after the first 6–12 months. A new class of agents called proliferation signal inhibitors (sirolimus and everolimus) work by inhibiting both T- and B-lymphocyte proliferation through G1 cell cycle blockade. These newer agents have become increasingly popular over the past 5 years due to their ability to prevent and retard progression of CAV and calcineurin-inhibitor-associated nephrotoxicity.

Significant advances have been made over the past two decades in moving from drugs that provide broad and non-specific immunosuppression to newer agents that provide more targeted immunosuppression through inhibition of lymphocyte activation and proliferation. Although use of these newer agents has decreased the incidence of both rejection and life-threatening infections, modern immunosuppressive regimens are inherently associated with drug- and class-specific toxicities, including metabolic derangements (hypertension, dyslipidemia, diabetes mellitus) and renal dysfunction. Additionally, heart transplant patients continue to have higher risk of developing opportunistic infections and malignancy compared to the general population due to suppression of their immune system. The risk of infection and neoplasm is typically related to the intensity and duration of immunosuppression.

Given the continued inherent toxicity of immunosuppression regimens, the variability of immunosuppression and susceptibility to drug toxicities among different individuals, and the constant aim to find a regimen that provides optimal immunosuppression with minimal side effects, there continues to be a great need for ways of monitoring immune suppression. Better tools for monitoring relative immunosuppression would allow for tailoring of immunosuppressive regimens to individual patients and aid in the development of new therapeutics (Table 27.1).

TABLE 27.1 Immunosuppression therapies

Method	Target	Selectivity
Total body irradiation	Bone marrow	+
Steroids	Lymphocytes/RES	+
Thoracic duct drainage	Lymphocytes	++
Antilymphocyte globuline	Lymphocytes	++
Azathioprine	Lymphocytes	++
Plasmapheresis	Antibodies	++
Cyclophosphamide	B-lymphocytes	++
Antithymocyte globuline	T-lymphocytes	++
Monoclonal CD3 antibodies	CD3+ T-lymphocytes	+++
Monoclonal CD4 antibodies	CD4+ T-lymphocytes	+++
Mycophelonate	De novo purine synthesis in lymphocytes	++++
Cyclosporine	IL2 inhibition in T-lymphocytes	++++
Tacrolimus	IL2 inhibition in T-lymphocytes	++++
Daclizumab	IL2 receptor antibodies	++++

TABLE 27.2 Grading system for endomyocardial biopsy

ISHLT Standardized cardiac biopsy grading
Acute cellular rejection

Old (1990) Grade	Revised (2004) Grade	Description
0	0R	No rejection
1A 1B 2	1R (mild)	Interstitial and/or perivascular mononuclear cell infiltrate with up to one focus of myocyte damage
3A	2R (moderate)	Two or more foci of mononuclear cell infiltrate with associated myocyte damage
3B 4	3R (severe)	Diffuse mononuclear and/or mixed inflammatory cell infiltrates with multiple foci of myocyte damage, with or without edema, hemorrhage, or vasculitis
Antibody-mediated rejection		
AMR 0		Negative for acute AMR
AMR 1		Positive immunofluorescence or immunoperoxidase staining for AMR (positive CD68, C4d).

R denotes "revised" grade. Adapted from Stewart et al. (2005).

CURRENT STRATEGIES FOR MONITORING TRANSPLANT REJECTION

Most patients with acute rejection are asymptomatic and have no clinical findings of cardiac allograft dysfunction. Additionally, the signs and symptoms of rejection, when present, are non-specific and may only manifest in the late stages of rejection. Thus, close surveillance of heart transplant recipients for acute rejection is critical. Patients are typically monitored for rejection using a combination of clinical assessment, imaging and/or quantification of allograft function [echocardiography, multiple gated acquisition (MUGA) scan, measurement of intracardiac pressures and flows], in addition to sampling of the myocardium via the technique of endomyocardial biopsy (EMB). Protocols for the timing of rejection surveillance are variable among transplant programs but are generally chosen to match the observed frequency of rejection episodes, which is clearly highest in the early postoperative period. Most programs perform rejection surveillance on a weekly basis for the first 4–6 postoperative weeks and then with diminishing frequency in a stable patient but at a minimum of every 3 months for the first postoperative year and at 3–6 months intervals after the first year.

The gold standard for diagnosing ACR has remained the EMB. In this procedure, several pieces of heart tissue from the inner portion (endomyocardium) of the right ventricle are removed with a rigid catheter advanced into the heart via a vein in the patient's neck or groin area. The specimens are then evaluated under a microscope for evidence of inflammatory infiltrates and myocyte injury. A uniform and standardized grading scheme for grading of ACR was developed by the International Society of Heart and Lung Transplantation (ISHLT) in 1990 and recently revised in 2004 (Stewart et al., 2005) (Table 27.2). Rejection therapy in the form of augmented immunosuppression is typically given for rejection grades of 3A/2R or higher and for rejection episodes that are associated with allograft dysfunction. The EMB, compared to other modalities of rejection surveillance, has the advantage of identifying rejection prior to the development of cardiac allograft dysfunction. However, the procedure is invasive, expensive, subject to sampling error, inter-observer variability, and causes morbidity (0.5–1.5%) (Nielsen

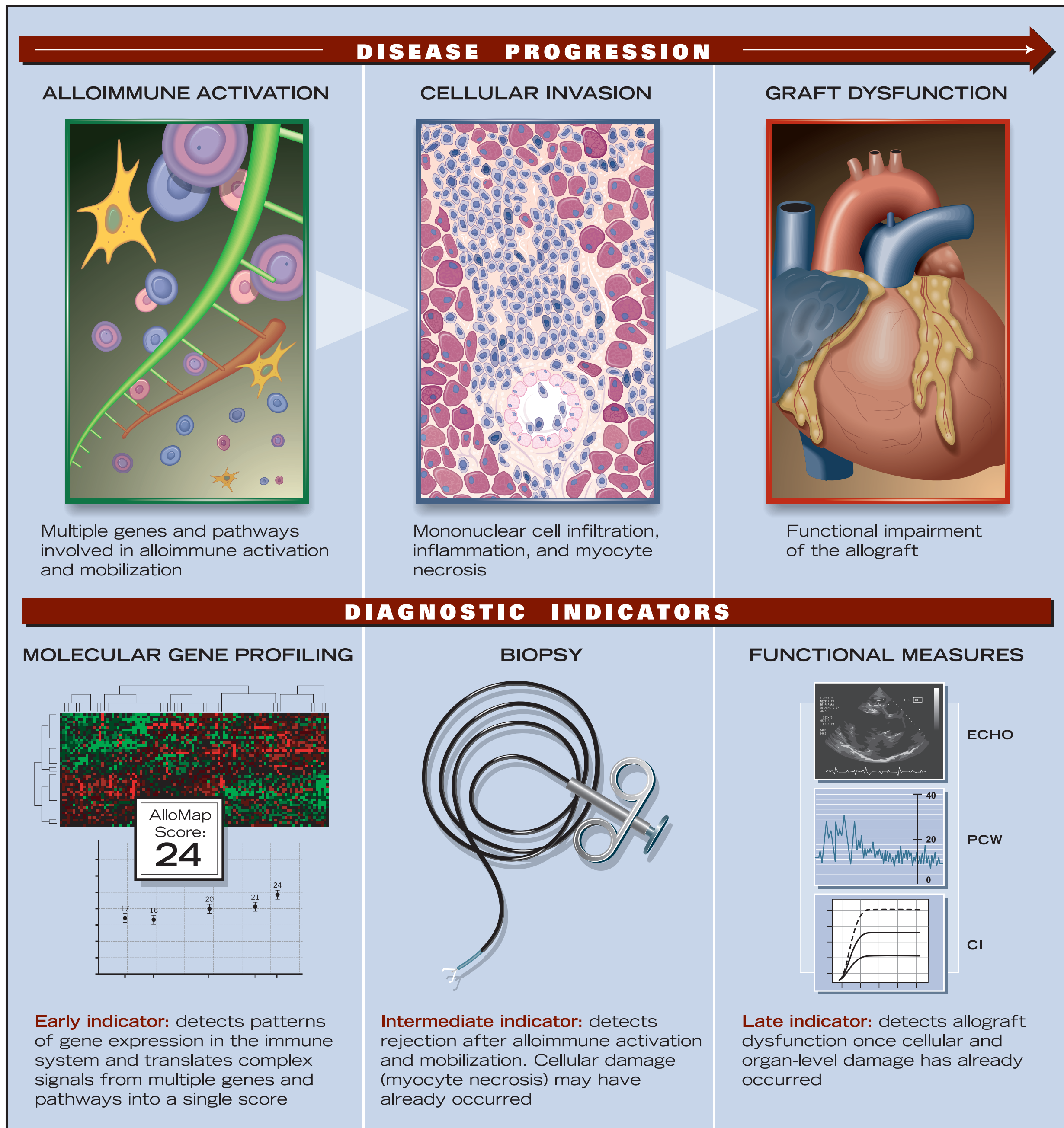


Figure 27.1 Progression and diagnosis of acute cellular rejection within the cardiac allograft.

et al., 1993; Winters and McManus, 1996; Winters et al., 1998). Biopsy-related complications include damage to the tricuspid valve, myocardial perforation, and excessive bleeding. Although non-invasive alternatives to EMB are clearly needed, methods such as echocardiography, ultrasonic myocardial backscatter, radionuclide imaging, magnetic resonance imaging, intra-myocardial electrograms, measurement of serum cardiac markers such as cardiac troponin I and B-type natriuretic peptide, and multiparametric immune monitoring have been difficult to validate and implement (Deng et al., 2005).

Most of the currently employed strategies identify cellular rejection long after immune activation has occurred. In many cases, myocyte injury and allograft dysfunction have already occurred (Figure 27.1). Despite activation of a broad range of immune pathways occurring in rejection, it has not been possible to identify specific molecular clues that inform on the degree of rejection in the transplanted organ. Clearly, greater information regarding the genes and proteins that are specifically activated in relation to transplant rejection is needed to provide a measure of rejection activity. Based on advances in the human

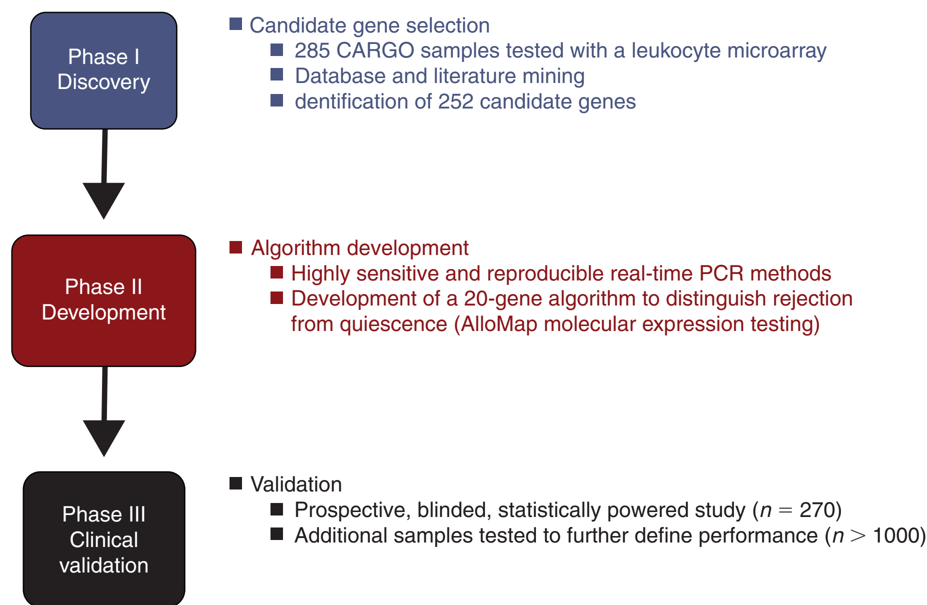


Figure 27.2 Approach to the development of a gene expression profiling test for cardiac transplant rejection.

genome project and based on the development of microarray technology and related computational methodology, a group of clinical and molecular scientists hypothesized that peripheral blood leukocyte expression profiling might reflect differential activity in the two states of acute rejection and quiescence within the allotransplanted heart.

THE CARGO CLINICAL STUDY

As circulating peripheral blood mononuclear cells (PBMC) reflect the status of a host immune system and responses to the allograft, measurement of PBMC gene expression should provide useful diagnostic or prognostic information for cardiac transplant recipients. Based on this hypothesis, studies have now been performed (Deng et al., 2006a; Horwitz et al., 2004; Schoels et al., 2004) to identify genes and pathways predictive of rejection in peripheral blood and to develop a molecular test based on this information. To identify the correlation between gene expression in circulating mononuclear cells and the relative degree of cardiac allograft rejection, a multicenter prospective study called the Cardiac Allograft Rejection Gene Expression Observational (CARGO) was initiated. Based on the genes and pathways identified as correlates of rejection, the CARGO study resulted in the development and validation of a peripheral blood gene expression test for rejection. The test was developed using DNA microarray technology and real-time PCR. The CARGO study had as a primary objective the development of a molecular diagnostic test from PBMC samples to discriminate between a quiescent state (original ISHLT Grade 0 rejection; revised ISHLT Grade 0R) and moderate/severe rejection (original ISHLT Grade >3A; revised ISHLT Grade >2R) in cardiac transplant recipients. Patients at eight heart transplant centers were followed prospectively with blood sampling performed at the time of post-transplant visits. Biopsies were graded by both local pathologists and three independent core pathologists blinded to clinical data. The scheme for gene discovery, test development, and validation from CARGO is shown in Figure 27.2.

DEVELOPMENT OF A GENE EXPRESSION SIGNATURE FOR CARDIAC TRANSPLANT REJECTION

Microarrays were used to identify genes in PBMC samples ($n = 285$) from cardiac transplant recipients that are associated with moderate/severe rejection or the absence of rejection. A set of 252 genes were selected for real-time PCR assay development from the microarray data and from genes in the published literature with known or hypothesized roles in cardiac allograft rejection (knowledge-based genes). While an empirical genomic approach should provide a comprehensive look at the genome, technical limitations of microarray technologies suggested that additional genes from existing knowledge should also be evaluated even if they were not selected from the microarray experiments. The limitations of microarrays include low sensitivity for genes expressed at low levels or in small subsets of circulating cells and nonspecificity of array probes for individual transcripts or splice variants. The 62 genes from the real-time PCR panel found to be reproducibly and differentially expressed between rejection and quiescence states were evaluated in a training set of 145 PBMC samples, and a multigene algorithm was developed using linear discriminant analysis for the ability to classify samples as having no rejection (ISHLT 0R) or moderate/severe rejection (ISHLT >2R). The optimal number of genes required for best classification was determined to be 11, with 9 genes being

included as controls. A gene expression profiling (GEP) test for detecting cardiac allograft rejection was then developed using PCR-based quantification of expression of these genes and their relative contribution to a probability score, based on the classification algorithm (Table 27.2). Finally, a validation cohort of independent samples from CARGO (63 samples from 63 patients) was used to validate the performance of the multigene test. These samples were independent from those used for the gene selection and algorithm development studies described above. This approach is critical for genomic studies that are fraught with false positive results due to the very large excess of genes to samples (Deng et al., 2006a). In this study, the test was shown to accurately detect the absence of moderate/severe rejection and thus identify a state of "quiescence" in the allograft. Using a single, pre-defined threshold of 20 (scores below 20 indicating a low probability of rejection), the test distinguished moderate/severe biopsy-defined rejection from quiescence ($p = 0.0018$) and had an agreement of 84% (95% CI 66–94%) with ISHLT Grade ≥3A/2R rejection compared with EMB. Beyond 1-year post-transplant, patient-GEP scores below the threshold also had a negative predictive value (NPV) for Grade ≥3A/2R rejection of >99%.

Based on these data, the GEP has been commercialized by XDx, Inc., Brisbane, CA. Their AlloMap™ test, which became available in January 2005, employs quantitative PCR of the 20 genes in triplicate and provides a score ranging from 0 to 40, with lower scores being associated with a very low likelihood of moderate/severe cardiac allograft rejection. A score of <30–34 has been used by many heart transplant centers to identify patients with a low probability of acute rejection. Details on development and validation of the GEP test have been published elsewhere (Deng et al., 2006b).

PATHWAYS MONITORED BY THE GEP (AlloMap™) TEST

A total of 62 genes were identified by real-time PCR in the validation phase of the CARGO study as being able to distinguish quiescent cardiac transplant patients from those who had moderate/severe biopsy-defined rejection. Study of the GeneOntology (http://www.geneontology.org) annotations associated with these genes, and in particular the 11 GEP test genes (Deng et al., 2006b), identified several different pathways (Table 27.3) associated with the rejection process. The pathways include T-cell priming, state of immunosuppression, changes in platelet phenotype, and systemic responses to inflammation in the allograft. T-cell receptor chains were upregulated. Another cluster involved hematopoiesis, reflecting an increase in erythroid progenitors, potentially stimulated by circulating byproducts of immune activity in the allograft (Hammer et al., 1998). Alternatively, a certain level of hematopoietic activity may be required to have or sustain an acute rejection response. Down-regulation of several genes was also observed. A number of genes of unknown function showed significant up- or down-regulation with rejection, requiring further work to elucidate their roles. In the context of the phenotypic changes that a naive T-cell acquires when it is primed in the lymph node (Mempel et al., 2004), PDCD1, a negative costimulatory molecule that minimizes the autoreactivity of an aggressive effector T-cell, is one of those candidates in the GEP classifier. Primed T-cells need to be able to traffic into sites of inflammation and acquire new trafficking molecules including ITGA4 (another GEP classifier gene).

TABLE 27.3 Genes and pathways represented in the final GEP (Allomap™) diagnostic test

Genes in AlloMap test	Corresponding pathways
SEMA7A	Macrophage activation/PMNs
IL1R2, FLT3, ITGAM	Steroid responsiveness
PF4, G6B	Platelet production
MIR, WDR40A	RBC production (hematopoiesis)
PDCD1, ITGA4	T-cell activation and regulation
RHOU	Cell morphology

VARIABILITY OF THE BIOPSY GOLD STANDARD AND RELATIONSHIP TO THE GEP (AlloMap™) SCORE

The CARGO study design originally assumed a "gold standard" clinical endpoint of biopsy-based detection of ACR. However, results from CARGO demonstrated that this "gold standard" was limited by considerable inter-observer variability among local pathologists, and use of GEP testing may help reduce the variability of diagnosing ACR inherent in the biopsy (Marboe et al., 2005). Therefore, in the development and evaluation of the GEP test, rejection was defined by a local pathologist's grading and by interpretation by a panel of three experienced, independent, blinded pathologists (central pathologists) who re-read each case. GEP scores were progressively higher, on average, as the rate of concordance among all pathologists' readings increased for identifying Grade >3A/2R rejection. When 1 of 4 pathologists diagnosed Grade ≥3A/2R, the average GEP score was 28.5. When 4 of 4 agreed on Grade ≥3A/2R, the average GEP score was 33 (Marboe et al., 2006).

DISCORDANCE BETWEEN BIOPSY GRADE AND MOLECULAR SCORE

Certainly the situation can arise where the GEP score is discordant from the histologic grading. A "positive" biopsy (ISHLT

Grade ≥3A/2R) and low GEP score (<34) are uncommon but may hypothetically be seen when local pathologists misdiagnose rejection, either overgrading or misinterpreting (e.g., Quilty lesions) the histology (Marboe et al., 2005). Additionally, a subset of focal rejection may be benign; in one study, more than 90% of Grade 3A/2R biopsies with two foci of focal moderate rejection diagnosed after 1-year post-transplant, resolved without therapy (Winters et al., 1995). Finally, molecular testing and biopsy measure different processes which may be discordant (e.g., lagging clearance of infiltrate in "resolving rejection"). In the CARGO study, rates of positive biopsy and low molecular score were low. For example, with a threshold score of 34 beyond 1-year post-transplant, the NPV of the test for Grade ≥2R/3A rejection was 99.2%. Thus, a Grade ≥2R/3A would be expected to occur with a low GEP score in 8 of 1000 tests in a population similar to CARGO. Conversely, a "negative" biopsy and high molecular score may be observed. Several hypotheses may explain this phenomenon, including early or focal rejection that may not be detected on the biopsy due to sampling error, alloimmune activation in the absence of cellular rejection on the biopsy, immune activation relating to conditions other than ACR (e.g., AMR, CAV/chronic rejection, or infection), or a quiescent state in a chronic and clinically stable heart transplant recipient. In the CARGO study, the rate of samples tested that had a negative biopsy with a high GEP score increased with time post-transplantation.

EFFECT OF TIME POST-TRANSPLANTATION ON PERFORMANCE OF THE GEP TEST

Samples from CARGO were used to derive the performance characteristics for the GEP test across a range of scoring thresholds. Since rejection rates and average GEP scores are known to vary with time post-transplant, performance characteristics were reported for defined time intervals (<6, 6–12, and ≥12 months). NPV and positive predictive value (PPV) were calculated for each threshold with respect to Grade ≥3A/2R rejection as defined by both local and central pathologists. The percent above threshold is the estimated overall rate of positive tests for an outpatient population. When interpreting the PPV for the GEP test for Grade ≥3A/2R rejection, it must be recognized that this parameter is highly dependent on the time-dependent prevalence of rejection in this population. In the first 6 months post-transplant, the risk of rejection is significantly higher than in later periods, and the PPV of the GEP test can be as high as 20–40% with high scores. Beyond 1-year post-transplant, the rate of Grade ≥3A/2R rejection drops to low levels (<3% of biopsies in CARGO), and the PPV of the test declines as well. This phenomenon is important to consider when interpreting high scores in patients ≥1 year post-transplant, who are at low risk for rejection. The expectation in the majority of these patients is that they do not have concurrent Grade ≥3A/2R rejection on biopsy. The GEP score associated with quiescence also rises with time post-transplant, probably related to down titration of corticosteroids and overall immunosuppression. Since the risk of rejection decreases with time and the NPV of the GEP score remains high over the range of scores between 30 and 35, the higher GEP score thresholds may be used to identify patients at very low risk for rejection during later periods post-transplant. Thus, implementation guidelines based on early clinical experience with the test recommend that GEP thresholds should vary by time post-transplant (0–6, 6–12, and ≥12 months post-transplant) (Starling et al., 2006).

RELATIONSHIP OF GEP SCORE TO CORTICOSTEROID DOSE

Quiescent GEP scores rise with time post-transplant throughout the first year, with the steepest part of this rise in the first 6 months (Figure 27.3). Corticosteroid dosing is the clinical variable most strongly associated with this pattern (Starling et al., 2005). Expression of a number of genes in the GEP algorithm correlates with steroid dose (*ITGAM*, *IL1R2*, *FLT3*). The pattern of change in expression of these genes with decreasing steroid dose is the same as the pattern observed with rejection. Steroid dosing is rapidly reduced in the first 6 months and progressively lowered during the remainder of the first year and beyond. Available data from the CARGO registry suggest that prednisone doses of <20 mg/day do not significantly influence the GEP score.

PREDICTION OF FUTURE ACR BY MOLECULAR SCORE

Peripheral blood gene expression patterns in 104 clinically stable heart transplant patients without initial histological evidence of acute rejection (ISHLT Grade 0 or 1A/1R) were evaluated in a nested case-control study within CARGO, to assess the ability of the GEP test to predict rejection in the subsequent 12 weeks. The cases included 39 patients who developed subsequently moderate to severe rejection (ISHLT Grade >3A/2R) within 12 weeks, and the controls included the remaining 65 patients who did not experience rejection during this time period. In addition, changes in individual gene expression patterns and their relationship with rejection therapy were studied. The gene expression score was significantly higher in patients with future moderate/severe rejection ($p = 0.01$), and the difference was more statistically significant within 180 days post-transplant ($p = 0.0004$). In this period, no patients with gene expression scores ≤20 developed moderate/severe rejection within 12 weeks, while 58% of patients with gene expression scores ≥30 did go on to develop rejection. Individual genes that most significantly predicted future rejection included *IL1R2* and *FLT3*, both corticosteroid responsive genes that decreased in expression

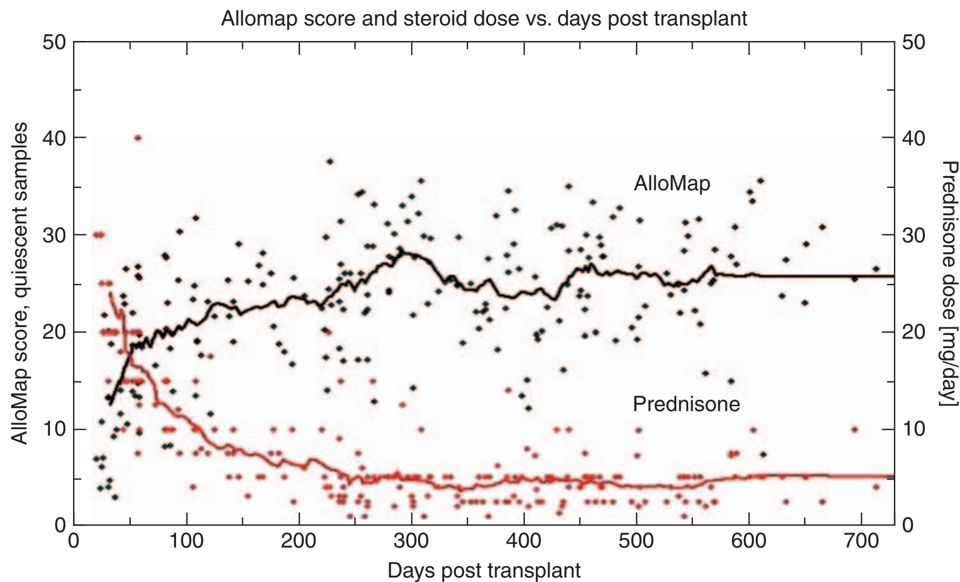


Figure 27.3 The GEP (AlloMap™) score and steroid dose versus days post-transplant.

before rejection, and *PCD1*, a marker of T-cell activation, that increased prior to rejection. Anti-rejection therapy resulted in a significant decrease in gene expression score (p=0.01). These data suggest that peripheral blood gene expression patterns can anticipate future moderate/severe cardiac allograft rejection in advance of EMB and may allow pre-emptive augmentation of baseline immunosuppression (Mehra et al., 2007).

CLINICAL USE OF THE AlloMap™ TEST

AlloMap™ is commercially available through the CLIA-certified XDx reference laboratory for use in heart transplant recipients who are 2 months or beyond post-transplantation and has been used clinically by US transplant centers since January 2005. Experience with rejection surveillance protocols incorporating GEP testing has provided additional insight into the performance characteristics of this test when used in a real-time clinical setting and across a wide spectrum of time periods post-transplantation. To date, transplant centers have used the AlloMap™ test in conjunction with the biopsy or in lieu of the biopsy in patients who are ≥3 months post-transplant. Indications for the GEP test vary according to the time period in which it is used. Centers have used the test within the first year to identify individuals at future risk of rejection and to assist in weaning of corticosteroids and immunosuppression. Beyond the first year post-transplantation, the test is typically used to identify patients at low risk of rejection that may safely be managed without routine biopsies. Finally, the AlloMap™ test has been used to noninvasively exclude rejection in patients with ambiguous signs or symptoms, in patients with inadequate biopsy specimens, and in patients with difficult vascular access.

The pooled data from early clinical implementation of GEP testing at several large US transplant centers (follow-up date March 31, 2006) have been analyzed and reported (Starling et al., 2004). Two hundred and forty-three clinical AlloMap™ measurements were included, 32 (13.2%) during the 6–12-month period, 192 (79.0%) during the 1–5-year period, and 19 (7.8%) during the >5-year period. The most important observation was the CARGO study confirmation of a high NPV (100%) with respect to ISHLT ≥2R/3A biopsy grades. It should be recognized that the mean duration post-transplant in this cohort was longer than the 14.5 months in the CARGO study. The data suggest that the test characteristics, when applied to patients who are ≥6 months post-transplant, are similar to those derived from the CARGO study. However, in this patient population, a higher test threshold (<34), which still maintains excellent NPV of >99%, may be appropriate as the AlloMap™ scores tend to rise with time post-transplant in clinically stable patients with histologically confirmed "quiescence" (Starling et al., 2006). A second important observation derived from clinical experience is that a subset of patients have consistently high longitudinal AlloMap™ scores but yet do not have associated Grade ≥3A/2R rejection on their biopsies. Therefore, some transplant centers have discontinued biopsies or transitioned to non-invasive rejection surveillance protocols in patients with consecutively high scores and quiescent biopsies. When used as part of a non-invasive strategy of rejection surveillance, the GEP test is typically combined with clinical and echocardiographic assessment of graft function to identify patients that may safely be managed without a heart biopsy. A clinical algorithm utilizing GEP testing to noninvasively manage heart transplant patients

who are beyond 1-year post-transplantation has been published (Pham et al., 2007).

While the performance of the AlloMap™ test has been validated in a large number of transplant patients, the clinical outcomes associated with using a gene expression-based strategy to monitor for rejection are currently unknown. A multicenter randomized clinical trial is underway to evaluate a GEP-based strategy, compared to a biopsy-based strategy, for assessing rejection in heart transplant recipients who are >6 months–5 years post-transplant. The "Invasive Monitoring Attenuation through Gene Expression (IMAGE)" will evaluate the impact of these two strategies with respect to clinically meaningful outcomes, such as graft dysfunction, rejection with hemodynamic compromise, and death. Additionally, the study will evaluate the incidence of biopsy-related complications, quality of life, and resource utilization among the two groups of patients.

2009 UPDATE

The AlloMap® Molecular Expression test received clearance by the US Food and Drug Administration (FDA) in August 2008 to aid in the identification of heart transplant recipients with stable allograft function who have a low probability of moderate/severe acute cellular rejection at the time of testing in conjunction with standard clinical assessment (US Food and Drug Administration, 2008). AlloMap® is the third *in vitro* diagnostic multivariate index assay (IVDMIA) cleared by the FDA. An IVDMIA is a test system that employs data from one or more *in vitro* assays, in conjunction with an algorithm, to generate a single, patient-specific result. The other FDA-approved tests under this category include a microarray-based multigene expression test that provides breast cancer prognostic information, as well as another microarray-based test that can help identify the type of cancer in a tumor sample by comparing the gene expression patterns within the sample with known gene expression patterns from malignant tumor types stored in a database.

Clinical studies evaluating the role of the GEP in heart transplantation are ongoing, and their findings should provide additional insight into use of this technology for detecting cardiac allograft rejection and other transplant-related conditions. In addition to detecting acute cellular rejection, small single-center studies have evaluated the role of GEP in identifying antibody-mediated rejection (Cadeiras et al., 2008), cardiac allograft vasculopathy (Yamani et al., 2007b), and post-transplant ischemic injury (Yamani et al., 2007a).

Additionally, the results from several large-scale observational and randomized prospective studies are expected to be announced in 2009–2010. The CARGO II study, involving over 700 heart transplant recipients from 13 European and 4 North American centers, was recently completed. The results of this study, expected in 2009, should provide additional information on the performance characteristics of the AlloMap® test in an international patient population. The IMAGE study, which has now enrolled over 500 heart transplant recipients from 12 large US heart transplant centers, is also expected to be completed in the fourth quarter of 2009 and will directly evaluate the impact of GEP-based versus biopsy-based rejection surveillance strategies on clinical endpoints. Additionally, the combined genetic and clinical information gathered from these studies may help investigators identify unique genetic profiles and mechanisms responsible for antibody-mediated rejection and cardiac allograft vasculopathy, thus potentially expanding the future role of GEP testing in the management of heart transplant recipients.

2009 UPDATE REFERENCES

Cadeiras, M., Bayern, M., Burke, E. et al. (2008). Gene expression profiles of patients with antibody-mediated rejection after cardiac transplantation. *J Heart Lung Transplant* 27(8), 932–934.

US Food and Drug Administration. (2008). FDA clears test to help doctors manage heart transplant patients. http://www.fda.gov/bbs/topics/NEWS/2008/NEW01877.html.

Yamani, M.H., Taylor, D.O., Haire, C., Smedira, N. and Starling, R.C. (2007a). Post-transplant ischemic injury is associated with up-regulated AlloMap gene expression. *Clin Transplant* 21(4), 523–525.

Yamani, M.H., Taylor, D.O., Rodriguez, E.R. et al. (2007b). Transplant vasculopathy is associated with increased AlloMap gene expression score. *J Heart Lung Transplant* 26(4), 403–406.

REFERENCES

Billingham, M.E., Cary, N.R., Hammond, M.E., Kemnitz, J., Marboe, C., McCallister, H.A., Snovar, D.C., Winters, G.L. and Zerbe, A. (1990). A working formulation for the standardization of nomenclature in the diagnosis of heart and lung rejection: Heart Rejection Study Group. The International Society for Heart Transplantation. *J Heart Transplant* 9, 587–593.

Book, W.M., Kelley, L. and Gravanis, M.B. (2003). Fulminant mixed humoral and cellular rejection in a cardiac transplant recipient: A

review of the histologic findings and literature. *J Heart Lung Transplant* 22, 604–607.

Deng, M.C., John, R., Baron, H., Itescu, S. and Suciu-Foca, N. (2005). Principles of transplantation immunology. In *Bone Disease of Organ Transplantation* (E. Shane and J. Compston, eds.), Elsevier, New York, pp. 3–30.

Deng, M.C., Eisen, H.J. and Mehra, M.R. (2006a). Methodological challenges of genomic research – the CARGO study. *Am J Transplant* 6, 1086–1087.

Deng, M.C., Eisen, H.J., Mehra, M.R., Billingham, M., Marboe, C.C., Berry, G., Kobashigawa, J., Johnson, F.L., Starling, R.C., Murali, S. et al. (2006b). Noninvasive discrimination of rejection in cardiac allograft recipients using gene expression profiling. *Am J Transplant* 6, 150–160.

Hammer, C., Reichenspurner, R., Meiser, B. and Reichart, B. (1998). Cytoimmunology in monitoring: The Munich experience. *Transplant Proc* 30, 873–874.

Hammond, E.H., Yowell, R.L., Nunoda, S., Menlove, R.L., Renlund, D.G., Bristow, M.R., Gay, W.A., Jr, Jones, K.W. and O'Connell, J.B. (1989). Vascular (humoral) rejection in heart transplantation: Pathologic observations and clinical implications. *J Heart Transplant* 8, 430–443.

Hershberger, R.E., Starling, R.C., Eisen, H.J., Bergh, C.H., Kormos, R.L., Love, R.B., Van Bakel, A., Gordon, R.D., Popat, R., Cockey, L. et al. (2005). Daclizumab to prevent rejection after cardiac transplantation. *N Engl J Med* 352, 2705–2713.

Horwitz, P.A., Tsai, E.J., Putt, M.E., Gilmore, J.M., Lepore, J.J., Parmacek, M.S., Kao, A.C., Desai, S.S., Goldberg, L.R., Brozena, S.C. et al. (2004). Detection of cardiac allograft rejection and response to immunosuppressive therapy with peripheral blood gene expression. *Circulation* 110, 3815–3821.

Kemnitz, J., Cremer, J., Restrepo-Specht, I., Haverich, A., Ziemer, G., Heublein, B., Borst, H.G., Uysal, A. and Georgii, A. (1991). Hyperacute rejection in heart allografts. Case studies.. *Pathol Res Pract* 187, 23–29.

Marboe, C.C., Billingham, M., Eisen, H., Deng, M.C., Baron, H., Mehra, M., Hunt, S., Wohlgemuth, J., Mahmood, I., Prentice, J. et al. (2005). Nodular endocardial infiltrates (Quilty lesions) cause significant variability in diagnosis of ISHLT Grade 2 and 3A rejection in cardiac allograft recipients. *J Heart Lung Transplant* 24, S219–S226.

Marboe, C., Deng, M.C., Berry, G., Billingham, M., Eisen, H.J., Pauly, D., Baron, H., Klingler, T.M., Lai, P.G., Mahmood, I. et al. (2006). Increased molecular testing scores associated with agreement among cardiac pathologists for the diagnosis of ISHLT 3A and higher rejection. *J Heart Lung Transplant* 25, 105–106.

Mehra, M.R., Kobashigawa, J.A., Deng, M.C., Fang, K.C., Klingler, T.M., Lal, P.G., Rosenberg, S., Uber, P.A., Starling, R.C., Murali, S., Pauly, D.F., Dedrick, R., Walker, M.G., Zeevi, A. and Eisen, H.J. (2007). CARGO Investigators. Transcriptional signals of T-cell and corticosteroid-sensitive genes are associated with future acute cellular rejection in cardiac allografts. *J Heart Lung Transplant* 12, 1255–1263.

Mempel, T.R., Henrickson, S.E. and Von Andrian, U.H. (2004). T-cell priming by dendritic cells in lymph nodes occurs in three distinct phases. *Nature* 427, 154–159.

Michaels, P.J., Fishbein, M.C. and Colvin, R.B. (2003). *Antibody-mediated rejection of human organ transplants*. Springer, New York.

Michaels, P.J., Espejo, M.L., Kobashigawa, J., Alejos, J.C., Burch, C., Takemoto, S., Reed, E.F. and Fishbein, M.C. (2003). Humoral rejection in cardiac transplantation: Risk factors, hemodynamic consequences and relationship to transplant coronary artery disease. *J Heart Lung Transplant* 22, 58–69.

Nielsen, H., Sorensen, F.B., Nielsen, B., Bagger, J.P., Thayssen, P. and Baandrup, U. (1993). Reproducibility of the acute rejection diagnosis in human cardiac allografts. The Stanford Classification and the International Grading System. *J Heart Lung Transplant* 12, 239–243.

Pham, M.X., Deng, M.C., Kfoury, A.G., Teuteberg, J.J., Starling, R.C. and Valantine, H. (2007). Molecular testing for long-term rejection surveillance in heart transplant recipients: Design of the Invasive Monitoring Attenuation through Gene Expression IMAGE) trial. *J Heart Lung Transplant* 26, 808–814.

Rodriguez, E.R. (2003). The pathology of heart transplant biopsy specimens: Revisiting the 1990 ISHLT working formulation. *J Heart Lung Transplant* 22, 3–15.

Rose, E.A., Gelijns, A.C., Moskowitz, A.J., Heitjan, D.F., Stevenson, L.W., Dembitsky, W., Long, J.W., Ascheim, D.D., Tierney, A.R., Levitan, R.G. et al. (2001). Long-term mechanical left ventricular assistance for end-stage heart failure. *N Engl J Med* 345, 1435–1443.

Schoels, M., Dengler, T.J., Richter, R., Meuer, S.C. and Giese, T. (2004). Detection of cardiac allograft rejection by real-time PCR analysis of circulating mononuclear cells. *Clin Transplant* 18, 513–517.

Starling, R.C., Deng, M.C., Kobashiganva, J.A., Walther, D., Wohlgemuth, J., Rosenberg, S. and Mehra, M. (2005). The influence of corticosteroids on the alloimmune molecular signature for cardiac allograft rejection. *J Heart Lung Transplant* 24, S65–S66.

Starling, R.C., Pham, M., Valantine, H., Miller, L., Eisen, H., Rodriguez, E.R., Taylor, D.O., Yamani, M.H., Kobashigawa, J., McCurry, K. et al. (2006). Molecular testing in the management of cardiac transplant recipients: Initial clinical experience. *J Heart Lung Transplant* 25, 1389–1395.

Stewart, S., Winters, G.L., Fishbein, M.C., Tazelaar, H.D., Kobashigawa, J., Abrams, J., Andersen, C.B., Angelini, A., Berry, G.J., Burke, M.M. et al. (2005). Revision of the 1990 working formulation for the standardization of nomenclature in the diagnosis of heart rejection. *J Heart Lung Transplant* 24, 1710–1720.

Taylor, D.O., Edwards, L.B., Boucek, M.M., Trulock, E.P., Waltz, D.A., Keck, B.M. and Hertz, M.I. (2006). Registry of the International Society for Heart and Lung Transplantation: Twenty-third official adult heart transplantation report – 2006. *J Heart Lung Transplant* 25, 869–879.

Winters, G.L. and McManus, B.M. (1996). Consistencies and controversies in the application of the International Society for Heart and Lung Transplantation working formulation for heart transplant biopsy specimens. Rapamycin Cardiac Rejection Treatment Trial Pathologists. *J Heart Lung Transplant* 15, 728–735.

Winters, G.L., Loh, E. and Schoen, F.J. (1995). Natural history of focal moderate cardiac allograft rejection. Is treatment warranted? *Circulation* 91, 1975–1980.

Winters, G.L., Marboe, C.C. and Billingham, M.E. (1998). The International Society for Heart and Lung Transplantation grading system for heart transplant biopsy specimens: Clarification and commentary. *J Heart Lung Transplant* 17, 754–760.

CHAPTER

Genetics and Genomics of Hypertrophic Cardiomyopathy

J. Martijn Bos, Steve R. Ommen and Michael J. Ackerman

INTRODUCTION

Hypertrophic cardiomyopathy (HCM) is a primary disorder of the myocardium associated with increase in cardiac mass and typically, asymmetric, predominantly left ventricular hypertrophy. Affecting 1 in 500 people, HCM is a disease of profound phenotypic and genotypic heterogeneity. Clinically, HCM can be characterized by a completely asymptomatic course to severe cardiac symptoms or even sudden cardiac death (SCD). In 1989, the molecular underpinnings of HCM were exposed with the discovery of the first locus of familial HCM, which was followed in 1990 with the identification of a mutation in the *MYH7*-encoded beta myosin heavy chain. Since then, hundreds of mutations scattered amongst at least 10 myofilament genes confer the pathogenetic substrate for this "disease of the sarcomere/myofilament." More recently, the genetic spectrum of HCM has expanded to encompass mutations in Z-disc–associated genes (Z-disc HCM). Furthermore, seemingly unexplained cardiac hypertrophy mimicking HCM can be the (primary) feature of some syndromes or metabolic diseases.

However, the underlying genetic mechanism for approximately 30–50% of HCM remains to be elucidated and although many genotype–phenotype relationships have been observed, only few carry enough clinical significance to aid physicians predicting genotype status or course and prognosis of the disease. Current treatment for HCM is mostly aimed at symptom relief, although studies are underway to design pharmacologic treatments for prevention of hypertrophy or individualized treatment based on one's genetic fingerprint.

DEFINITIONS, CLINICAL PRESENTATION, AND DIAGNOSIS

Background

Hypertrophic cardiomyopathy is defined as unexplained left ventricular hypertrophy (LVH) in the absence of precipitating factors such as hypertension or aortic stenosis. HCM is a disease of enormous phenotypic and genotypic heterogeneity. Affecting 1 in 500 people, it is the most prevalent genetic cardiovascular disease, and more importantly the most common cause of SCD in young athletes (Maron, 2002). HCM can manifest with negligible to extreme hypertrophy, minimal to extensive fibrosis and myocyte disarray, absent to severe left ventricular outflow tract obstruction (LVOTO), and distinct patterns of hypertrophy.

Nomenclature

Hypertrophic cardiomyopathy was described fully for the first time by Teare in 1958 as "asymmetrical hypertrophy of the heart in young adults" (Teare, 1958). Over the past half century, HCM has since been known by a confusing array of names, reflecting its clinical heterogeneity and uncommon occurrence in daily

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

practice. In 1968, WHO defined cardiomyopathies as "diseases of different and often unknown etiology in which the dominant feature is cardiomegaly and heart failure" (Abelmann, 1984). In 1980, cardiomyopathies were newly defined as "heart muscle diseases of unknown cause," thereby differentiating it from specific identified heart muscle diseases of known cause such as myocarditis. Throughout the years, names such as idiopathic hypertrophic subaortic stenosis (IHSS), (Braunwald et al., 1964), muscular subaortic stenosis (Pollick et al., 1982), and hypertrophic obstructive cardiomyopathy (HOCM) (Schoendube et al., 1995) have been used widely and interchangeably to define the same disease. In 1995, the WHO/International Society and Federation of Cardiology Task Force on cardiomyopathies classified the different cardiomyopathies by dominant pathophysiology or, if possible, by etiological/pathogenetic factors (Richardson et al., 1996). The four most important cardiomyopathies – dilated cardiomyopathy (DCM), restrictive cardiomyopathy (RCM), arrhythmogenic right ventricular cardiomyopathy (ARVC), and hypertrophic cardiomyopathy (HCM) – were recognized, next to a number of specific and mostly acquired cardiomyopathies, like ischemic- and inflammatory cardiomyopathies (Richardson et al., 1996). Accordingly, HCM is described as left and/or right ventricular hypertrophy, usually asymmetric and involving the interventricular septum with predominant autosomal dominant inheritance that can be caused by mutations in sarcomeric contractile proteins (Richardson et al., 1996). On a microscopic level, HCM is characterized by the classical triad of cardiomyocyte hypertrophy, fibrosis, and myofibrillar disarray.

Clinical Presentation and Diagnosis

The clinical presentation of HCM is underscored by extreme variability from an asymptomatic course to that of severe heart failure, arrhythmias, and SCD. Many patients remain asymptomatic or only mildly symptomatic throughout the course of life. HCM commonly manifests between the second and the fourth decades of life but can present at the extremes of age. Infants and young children may present with severe hypertrophy leading to heart failure, and these patients have poor prognosis. More often, SCD can be the tragic sentinel event for HCM in children, adolescents, and young adults. The most common symptoms at presentation of disease are exertional dyspnea, chest pain, and syncope or presyncope. Approximately 5% of patients with HCM progress to "end-stage" disease characterized by LV dilatation and heart failure. In such cases cardiac transplantation may be considered. Other serious life-threatening complications include embolic stroke and cardiac arrhythmias.

Echocardiography

Conventional two-dimensional echocardiography is the diagnostic modality of choice for the clinical diagnosis of HCM. The disease is characterized by otherwise unexplained and usually asymmetric, diffuse or segmental hypertrophy associated with a non-dilated left ventricle (LV) independent of presence or absence of LV outflow obstruction. A left ventricular wall thickness of ≤12 mm is typically regarded as normal, with measurements of 13–15 mm labeled as "borderline hypertrophy." A maximal LV end-diastolic wall thickness exceeding 15 mm represents the absolute dimension generally accepted for the clinical diagnosis of HCM in adults (in children, 2 or more standard deviations from the mean relative to body surface area) (Maron et al., 2003). Echocardiography can also provide details of location and degree of hypertrophy. In general, while there are innumerable morphologic appearances of the heart, the following four different morphological subtypes of HCM can be recognized in most cases (Table 28.1, Figure 28.1): sigmoid septum, reverse septal curvature, apical-, and neutral contour variant.

TABLE 28.1 Septal morphologies in HCM

Septal morphology	Description
Sigmoid septum	– Septum concave to cavity with pronounced septal bulge – Ovoid LV cavity
Reverse septal curvature	– Predominant mid-septal convexity toward LV cavity – Crescent-shaped cavity
Apical variant	– Predominant apical distribution of hypertrophy
Neutral variant	– Overall straight or variable convexity, predominantly neither convex nor concave

Dynamic LVOTO is a common feature of HCM but is not required for the diagnosis of HCM. The existence of LVOTO is diagnosed by demonstration of a resting or provocable Doppler gradient of >30 mmHg. LVOTO is produced by the interaction of the hypertrophied septum and systolic anterior motion (SAM) of the mitral valve. The latter results from abnormal blood flow vectors across the valve, and abnormal anterior positioning of the valve and its support structures. Variable severity, posteriorly directed mitral regurgitation is a common finding.

Most patients at presentation manifest some degree of impaired diastolic function ranging from abnormal relaxation to severe myocardial stiffness, elevated left ventricular end-diastolic pressure (LVEDP), elevated atrial pressure, and pulmonary congestion leading to exercise intolerance and fatigue. Systolic cardiac function, as measured by ejection fraction, is usually preserved. It is notable that more elegant measures of systolic performance, such as tissue velocity and strain imaging, suggest a decrease in systolic function in patients with HCM, and may even be present in gene-mutation carriers before hypertrophy can be detected. "End-stage disease" is characterized by LV dilatation, poor systolic function, and heart failure.

Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) constitutes an important additional diagnostic tool especially in patients with suboptimal echocardiography or unusual segmental involvement of myocardium. Delayed gadolinium enhancement in MRI is an excellent

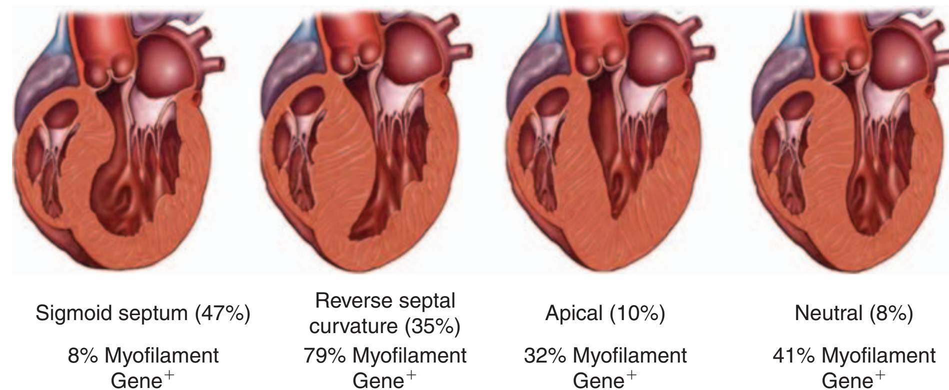


Figure 28.1 Septal morphologies in HCM. Shown are the most common septal morphologies in HCM, their distribution in a large cohort of patients with HCM as well as the yield of genetic testing for each morphological subgroup.

tool to identify areas of scarring, replacement fibrosis, or irreversible myocardial injury. There are some recent reports of higher risk of SCD associated with these finding. In a cohort of 42 patients, myocardial hyper-enhancement was found in 79%. Extent of hyper-enhancement was greater in patients with progressive disease (28.5% versus 8.7%, $p < 0.001$) and in patients with two or more risk factors for SCD (15.7% versus 8.6%, $p = 0.02$) (Moon et al., 2003).

MOLECULAR GENETICS OF HCM

Sarcomeric/Myofilament HCM

Since the sentinel discovery of the first locus for familial HCM on chromosome 14 (1989) and first mutations involving the *MYH7*-encoded beta myosin heavy chain (1990) as the pathogenic basis for HCM (Geisterfer-Lowrance et al., 1990; Jarcho et al., 1989), over 400 mutations scattered among at least 24 genes encoding various sarcomeric proteins (myofilaments and Z-disc–associated proteins) and calcium-handling proteins have been identified (Table 28.2, Figure 28.2). The most common genetically mediated form of HCM is myofilament HCM, with hundreds of disease-associated mutations in eight genes-encoding proteins critical to the sarcomere's thick myofilament (beta myosin heavy chain [*MYH7*] [Geisterfer-Lowrance et al., 1990]; regulatory myosin light chain [*MYL2*] and essential myosin light chain [*MYL3*] [Poetter et al., 1996]), intermediate myofilament (myosin-binding protein C [*MYBPC3*] [Watkins et al., 1995]), and thin myofilament (cardiac troponin T [*TNNT2*], alpha-tropomyosin [*TPM1*] [Thierfelder et al., 1994], cardiac troponin I [*TNNI3*] [Kimura et al., 1997]), troponin C [*TNNC1*] [Landstorm et al., 2008] and actin ([*ACTC*] [Mogensen et al., 1999; Olson et al., 2002a]). Targeted screening of giant sarcomeric *TTN*-encoded titin, which extends throughout half of the sarcomere, has thus far revealed only one mutation (Satoh et al. 1999). More recently, mutations have been described in the myofilament protein alpha myosin heavy chain encoded by *MYH6* (Niimura et al., 2002). The prevalence of mutations in the eight most common myofilament-associated genes, currently comprising the commercially available HCM genetic tests in different international cohorts, ranges from 30% to 61%, leaving still a large number of patients with genetically unexplained disease (Van Driest et al., 2005a).

Z-disc and Calcium-Handling HCM

Over the last few years, the spectrum of HCM-associated genes has expanded outside the myofilament to encompass additional subgroups that could be classified as "Z-disc HCM" and "calcium-handling HCM." Due to its close proximity to the contractile apparatus of the myofilaments and its specific structure–function relationship with regard to cyto-architecture, as well as its role in the stretch-sensor mechanism of the sarcomere, recent attention has been focused on the proteins that comprise the cardiac Z-disc. Initial mutations were described in *CSRP3*-encoded muscle LIM protein (Geier et al., 2003) and *TCAP*-encoded telethonin (Hayashi et al., 2004), an observation replicated in a large cohort of unrelated patients with HCM (Bos et al., 2006). Recently, mutations in patients with HCM have been reported in *LDB3*-encoded LIM domain binding 3, *ACTN2*-encoded alpha-actinin-2, and *VCL*-encoded vinculin/metavinculin (Theis et al., 2006). Interestingly, although the first HCM-associated mutation in vinculin was found in the cardiac-specific insert of the gene, yielding the protein called metavinculin (Vasile et al., 2006a), the follow-up study also identified a mutation in the ubiquitously expressed protein vinculin (Vasile et al., 2006b). In 2007, a mutation in the Z-disc–associated *MYOZ2*-encoded myozenin 2 was reported as a novel gene for HCM (Osio et al., 2007).

As the critical ion in the excitation–contraction coupling of the cardiomyocyte, calcium and proteins involved in calcium-induced calcium release (CICR) have always been of high interest in the pathogenesis of HCM. Although with very low frequency, mutations have been described in the promoter – and coding region of *PLN*-encoded phospholamban – an important

TABLE 28.2 Summary of HCM susceptibility genes

	Gene	Locus	Protein	Estimated frequency (%)
Myofilament HCM				
Giant filament	*TTN*	2q24.3	Titin	<1
Thick filament	*MYH7*[a]	14q11.2–q12	β-myosin heavy chain	15–25
	MYH6	14q11.2–q12	α-myosin heavy chain	<1
	MYL2[a]	12q23–q24.3	Ventricular regulatory myosin light chain	<2
	MYL3[a]	3p21.2–p21.3	Ventricular essential myosin light chain	<1
Intermediate filament	*MYBPC3*[a]	11p11.2	Cardiac myosin-binding protein C	15–25
Thin filament	*TNNT2*[a]	1q32	Cardiac troponin T	<5
	TNNI3[a]	19p13.4	Cardiac troponin I	<5
	TPM1[a]	15q22.1	α-tropomyosin	<5
	ACTC[a]	15q14	α-cardiac actin	<1
	TNNC1[a]	3p21.3–p14.3	Cardiac troponin C	<1
Z-disc HCM				
	LBD3	10q22.2–q23.3	LIM binding domain 3 (alias: ZASP)	1–5
	CSRP3	11p15.1	Muscle LIM protein	<1
	TCAP	17q12–q21.1	Telethonin	<1
	VCL	10q22.1–q23	Vinculin/metavinculin	<1
	ACTN2	1q42–q43	α-actinin 2	<1
	MYOZ2	4q26–q27	Myozenin 2	<1
Calcium-handling HCM				
	RYR2	1q42.1–q43	Cardiac ryanodine receptor	<1
	JPH2	20q12	Junctophilin-2	<1
	PLN	6q22.1	Phospholamban	<1

[a]Available as part of commercially available genetic test for HCM.

inhibitor of cardiac muscle sarcoplasmic reticulum Ca(2+)-ATPase (SERCA) (Haghighi et al., 2006; Minamisawa et al., 2003), the *RyR2*-encoded cardiac ryanodine receptor (Fujino et al., 2006), and the *JPH2*-encoded type 2 junctophilin (Landstrom et al., 2007).

Metabolic and Syndromal HCM-mimicry

The last genetic subgroup of unexplained cardiac hypertrophy is the one comprising metabolic and syndromal HCM-mimickers – in which cardiomyopathy is a sometimes sole presenting feature (Table 28.3). In 2001, the first mutations in *PRKAG2*-encoded AMP-activated protein kinase gamma 2, a protein involved in the energy homeostasis of the heart, were described in two families with severe HCM and aberrant AV-conduction in some individuals (Blair et al., 2001). Further studies showed that patients with mutations in this gene lack the HCM characteristics of myocyte – and myofibrillar disarray – but show newly formed vacuoles filled with glycogen-associated granules. This glycogen storage disease therefore seemed to mimic HCM, distinguishing itself by electrophysiological abnormalities, particularly ventricular pre-excitation (Arad et al., 2002, 2005). In 2005, Arad et al. described mutations in lysosome-associated membrane protein2 encoded by *LAMP2* (Danon's syndrome) and protein kinase gamma 2 encoded by *PRKAG2* in glycogen storage disease-associated genes mimicking the clinical phenotype of HCM (Arad et al., 2005). A recent community-based study showed that in 50 healthy individuals with idiopathic LVH, participants with sarcomere gene protein and storage mutations essentially were indistinguishable clinically from those without mutations (Morita et al., 2006).

In 2005, a mutation in *FXN*-encoded frataxin associated with Friedrich ataxia was described in a patient with HCM. Although this patient also harbored a myofilament mutation in *MYBPC3*-encoded myosin binding protein C, functional characterization showed significant influence of the *FXN*-mutant on the phenotype, suggesting that the observed alterations in energetics may act in synergy with the present myofilament mutation (Van Driest et al., 2005b). Akin to PRKAG2 and LAMP2, Fabry's disease can express predominant cardiac features of seemingly unexplained LVH. Over the years, mutations in *GLA*-encoded alpha-galactosidose A have been found in patients with this multisystem disorder (Nakao et al., 1995; Sachdev et al., 2002; Sakuraba et al., 1990).

Genotype–Phenotype Relationships in HCM

For over the past decade, multiple studies have tried to identify phenotypic characteristics most indicative of myofilament HCM to facilitate genetic counseling and strategically direct clinical genetic testing (Ackerman et al., 2002; Richard et al.,

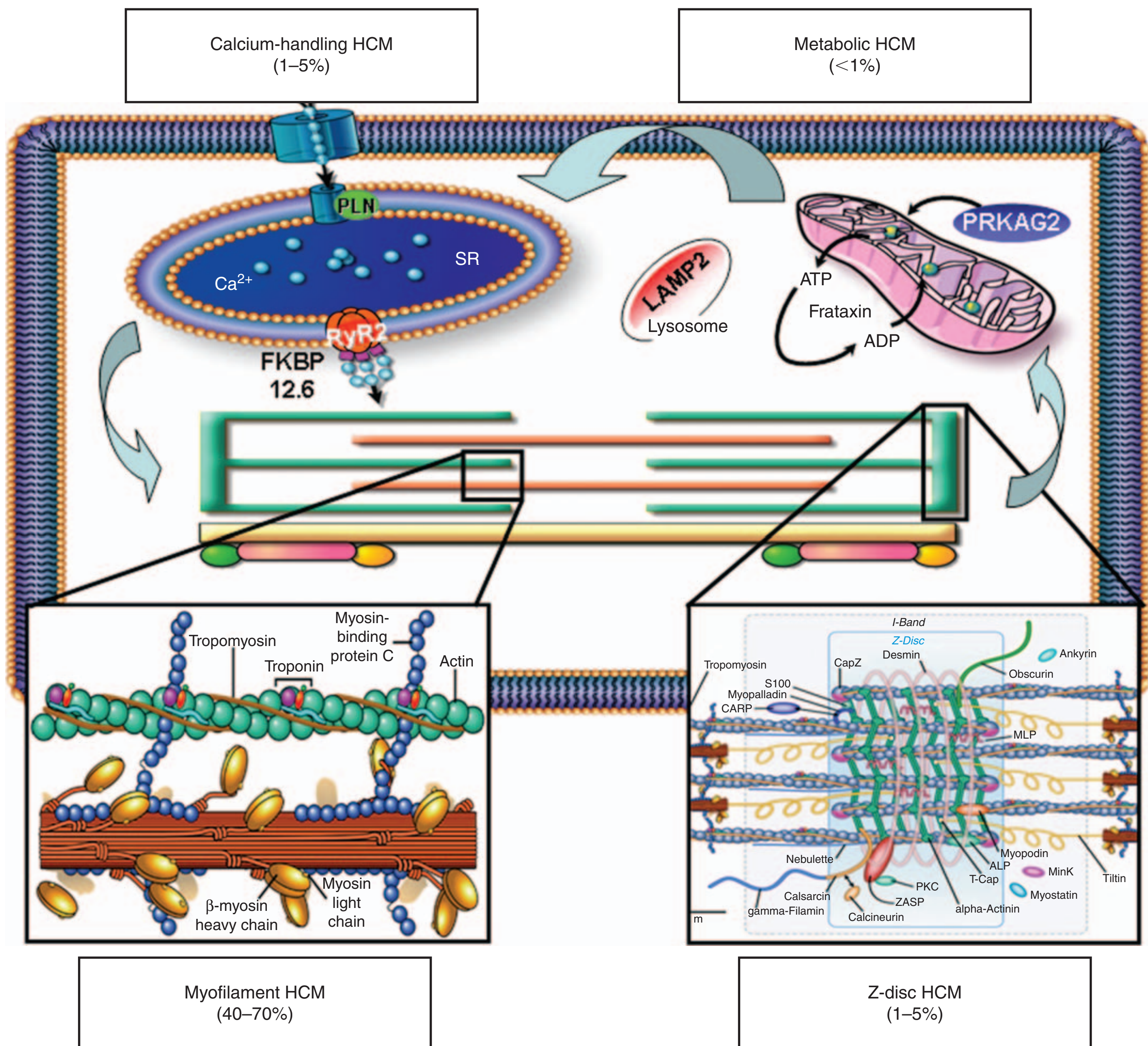


Figure 28.2 Subgroups of genetic HCM. Shown are the most important functional subgroups of genetically mediated HCM. Blue arrows indicate the functional relationship between the different elements.

2003; Van Driest et al., 2002, 2003, 2004a, b; Woo et al., 2003). Up until 2001, it was thought that specific mutations in these myofilament genes were inherently "benign" or "malignant" (Anan et al., 1994; Coviello et al., 1997; Elliott et al., 2000; Moolman et al., 1997; Niimura et al., 2002; Seidman et al., 2001; Varnava et al., 1999; Watkins et al., 1992). However, these studies were based on highly penetrant, single families with HCM, and later genotype–phenotype studies involving a large cohort of unrelated patients have indicated that great caution must be exercised with assigning particular prognostic significance to any particular mutation (Ackerman et al., 2002; Van Driest et al., 2002, 2004b). Furthermore, these studies demonstrated that the two most common forms of genetically mediated HCM – *MYH7*-HCM and *MYBPC3*-HCM – were phenotypically indistinguishable (Van Driest et al., 2004b). While several phenotype–genotype relationships have emerged to enrich the yield of genetic testing, these patient profiles have not been particularly clinically informative on an individual level.

TABLE 28.3 Metabolic and Syndromal-HCM-mimickers

Gene	Locus	Protein	Syndrome
PRKAG2[a]	7q35-q36.36	AMP-activated protein kinase	WPW/HCM
LAMP2[a]	Xq24	Lysosome-associated membrane protein 2	Danon's syndrome
GAA	17q25.2-q25.3	Alpha-1,4-glucosidase deficiency	Pompe's disease
GLA[a]	Xq22	Alpha-galactosidase A	Fabry's disease
FXN	9q13	Frataxin	Friedrich's ataxia
PTPN11[b]	12q24.1	Protein tyrosine phosphatase, non-receptor type 11	Noonan's syndrome

Note: In patients with cardiac hypertrophy secondary to Noonan's syndrome, *PTPN11* mutations are quite uncommon (<10%) compared to >50% yield expected for Noonan's syndrome without LVH.
[a]Available as part of commercially available genetic testing.
[b]*PTPN11* is also associated with LEOPARD syndrome and is allelic to Noonan's syndrome.

Recently, an important discovery, linking the echocardiographically determined septal morphology to the underlying genetic substrate, was made (Binder et al., 2006). The first link to be drawn between septal morphologies was a result of a HCM study by Lever and colleagues in the 1980s, where septal contour (Table 28.1) was found to be age-dependent with a predominance of sigmoidal-HCM noted in the elderly (Lever et al., 1989). In the early 1990s, an early genotype–phenotype observation involving a small number of patients and family members showed that patients with mutations in the beta myosin heavy chain (*MYH7*-HCM) generally had reversed curvature septal contours (reverse curve-HCM) (Solomon et al., 1993).

Analysis of the echocardiograms of 382 previously genotyped and published patients (Van Driest et al., 2003, 2004b, c), revealed that sigmoidal-HCM (47% of cohort) and reverse curve-HCM (35% of cohort) were the two most prevalent anatomical subtypes of HCM and that the septal contour was the strongest predictor for the presence of a myofilament mutation, regardless of age (Binder et al., 2006). In fact, multivariate analysis in this cohort demonstrated that septal morphology was the only, independent predictor of myofilament HCM with an odds ratio of 21 ($p < 0.001$) when reverse curve morphology was present (Figure 28.1; Binder et al., 2006).

The yield from the commercially equivalent HCM genetic research test for myofilament HCM was 79% in reverse curve-HCM but only 8% in patients with sigmoidal-HCM. Of the smaller subgroup of patients with apical HCM, 32% had a mutation in one of the elements of the cardiac myofilament (Binder et al., 2006). These observations may facilitate echo-guided genetic testing by enabling informed genetic counseling about the *a priori* probability of a positive genetic test based on the patient's expressed anatomical phenotype of HCM (Figure 28.1).

With the majority of known myofilament proteins studied, except for a complete analysis of the giant protein *TTN*-encoded titin, recent research has focused on proteins beyond the cardiac myofilaments, especially proteins involved in the cytoarchitecture and cardiac stretch-sensor mechanism of the cardiomyocyte localized to the cardiac Z-disc (Figure 28.2). The Z-disc is an intricate assembly of proteins at the Z-line of the cardiomyocyte sarcomere. Proteins of the Z-disc are important in the structural and mechanical stability of the sarcomere as they appear to serve as a docking station for transcription factors, Ca^{2+} signaling proteins, kinases, and phosphatases (Frank et al., 2006; Pyle et al., 2004). In addition, this assembly of proteins seems to serve as a way station for proteins that regulate transcription by aiding in their controlled translocation between the nucleus and the Z-disc (Frank et al., 2006; Pyle et al., 2004).

A main implication for the Z-disc is its involvement in the cardiomyocyte stretch sensing and response systems (Knoll et al., 2002). Mutations in these proteins have been implicated as HCM (Bos et al., 2006; Geier et al., 2003; Hayashi et al., 2004; Vasile et al., 2006a, b) and DCM susceptibility genes (Bos et al., 2006; Geier et al., 2003; Hayashi et al., 2004; Mohapatra et al., 2003; Olson et al., 2002b; Vasile et al., 2006a). Additionally, it has become appreciated that these divergent cardiomyopathic phenotypes of HCM and DCM are partially allelic disorders with *ACTC*, *MYH7*, *TNNT2*, *TPM1*, *MYBPC3*, *TTN*, *MLP*, *TCAP*, and *VCL* established as both HCM and DCM susceptibility genes (Daehmlow et al., 2002; Geier et al., 2003; Gerull et al., 2002; Hayashi et al., 2004; Kamisago et al., 2000; Mohapatra et al., 2003; Olson et al., 2000, 2001; Vasile et al., 2006a).

Recently, mutations in *ACTN2*-encoded alpha-actinin-2 (ACTN2) and *LDB3*-encoded LIM domain binding 3 (LDB3) as novel HCM susceptibility genes (Theis et al., 2006) were described. Linking reverse curve-HCM to the presence of myofilament mutation, and recognizing that the Z-disc may transduce multiple-signaling pathways during stress, translating into hypertrophic responses, cell growth and remodeling (Frey et al., 2004), it was observed that Z-disc HCM, in contrast to myofilament HCM, is preferentially sigmoidal. In fact, 11 out of 13 patients with Z-disc HCM had a sigmoidal septal contour and no reverse septal curvatures were seen (Theis et al., 2006). It is speculated that Z-disc HCM leads to a hypertrophic response that is expressed in the areas of highest stress (i.e., LVOT) and therefore predisposes to a sigmoidal septal contour.

New Insights and Approaches to Genomics of HCM

Not only in molecular genetics but also in other fields of "-omics," novel pathways underlying the pathophysiology of this heterogeneous disease have been identified using several new techniques to study large-scale transcriptional changes (Churchill, 2002; Holland, 2002; Velculescu et al., 1997). A transcriptomic approach

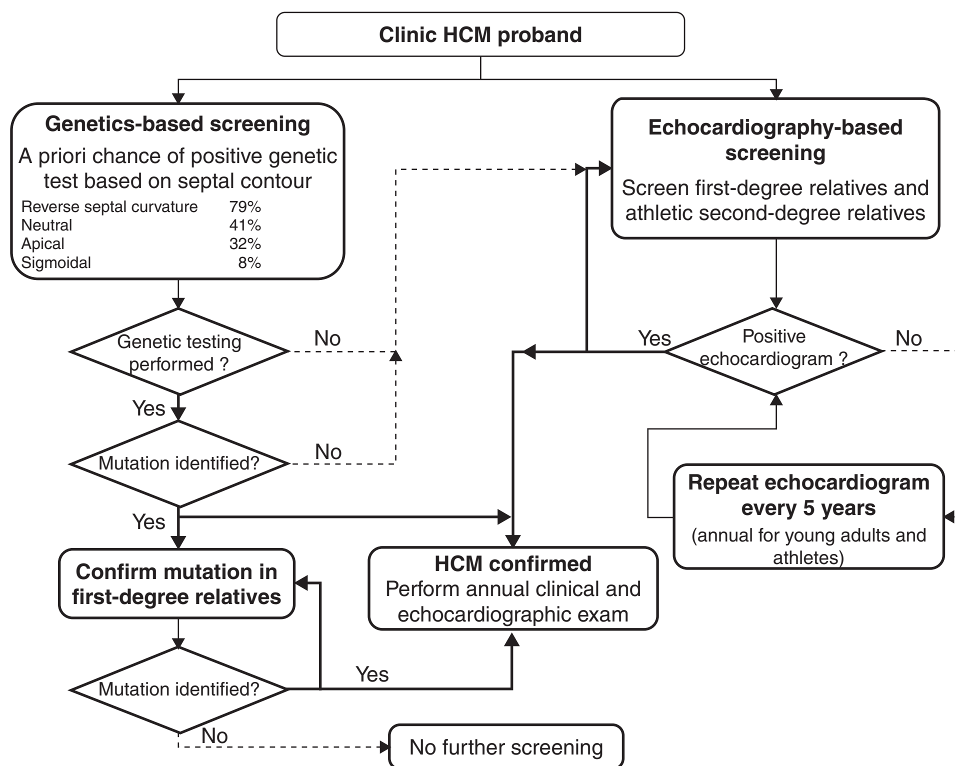


Figure 28.3 Genetic- and echocardiographic-based screening in HCM. Flow-chart showing decision to follow in genetic- and echocardiography-based screening in HCM. Noted are the *a priori* chances for a positive genetic test result based on the echocardiographic-scored septal contour, as well as the steps to follow if a patient chooses not to pursue genetic testing.

using microarray, a technique that enables one to give a snapshot view of gene expression, combined with complex analytic tools, can identify genes that seem to be co-regulated and thereby form a transcriptional network of genes and pathways. Microarray chips can hold over 10,000 genes and can be utilized to compare expression levels in certain disease states with healthy controls. In 2002, Hwang et al. studied RNA from heart failure patients with either HCM or DCM and found 192 genes to be upregulated in both, as well as several genes differentially expressed between the two diseases providing information on different pathways and genes involved in the pathogenesis (Hwang et al., 2002). More recently, Rajan et al. performed microarray analysis on ventricular tissue of two previously developed transgenic HCM mice carrying mutations in alpha-tropomyosin (TPM1). Studying 22,600 genes, they discovered 754 differentially expressed genes between transgenic and non-transgenic mice, of which 266 were differentially regulated between the two different mutant hearts showing most significant changes in genes belonging to the "secreted/extracellular matrix" (upregulation) and "metabolic enzymes" (downregulation) (Rajan et al., 2006).

Microarray techniques were also used by Sayed et al. when they studied the effect of aortic constriction on expression of microRNAs, fundamental regulators consisting of non-coding RNA molecules that silence genes through post-transcriptional regulation first described by Lee et al. in 1993 (Lee et al., 1993). In this recent study, they found that microRNAs, especially microRNA-1 (miR-1), play an important role in the development of cardiac hypertrophy, where downregulation of miR-1 leads to relief of its growth-related target genes, protein synthesis, and increased cell size showing a completely new approach to the -omics of HCM (Sayed et al., 2007).

In 2007, a novel-sensitive messenger RNA (mRNA) profiling technology, PMAGE (polony multiplex analysis of gene expression), which detects mRNAs as rare as one transcript per three cells, was developed (Kim et al., 2007). Using this new technique, early transcriptional changes preceding pathological manifestations were identified in mice with HCM-causing mutations, including low-abundance mRNA encoding signaling molecules and transcription factors that participate in the disease pathogenesis (Kim et al., 2007).

SCREENING AND TREATMENT FOR HCM

Family Screening in HCM

Genetic and clinical screening of family members with HCM plays an important role in the early diagnosis of HCM. Figure 28.3 shows one possible algorithm to follow after diagnosis of a proband with HCM. In general, all first-degree relatives and probably "athletic" second-degree relatives of an index case of HCM

should be screened by an ECG and an echocardiogram. Annual screenings are recommended for young persons (12–25 years) and athletes and thereafter every 5 years. If the HCM-causing mutation is known, first-degree relatives should have confirmatory genetic testing of the mutation in addition to the screening ECG and echocardiogram. Depending on the established familial versus sporadic pattern, confirmatory genetic testing should proceed in concentric circles of relatedness. For example, if the mutation is established in the patient's father, the patient's paternal grandparents should be tested, and if necessary, the paternal aunts and uncles and so forth. If the putative HCM-associated mutation is not found in a phenotype-negative family member, screening can be stopped. However, a decision to cease surveillance for HCM in a relative hinges critically on the certainty of the identified gene/mutation and its causative link as well as the complete absence of any traditional evidence used to diagnose HCM clinically (i.e., asymptomatic and normal echo).

Follow-up and Sports Participation

Intense physical exertion can potentially trigger SCD in individuals with HCM. According to the 2005 Bethesda Conference recommendations, athletes with HCM should be excluded from participation in contact sports as well as most organized competitive sports, with the possible exception of low-intensity sports classified as class IA sports (i.e., golf, bowling, cricket, billiards, and riflery) (Maron et al., 2005). Per the guidelines, the presence of an implantable cardioverter defibrillator (ICD) does not alter these recommendations. This restrictive approach is loosened for the patient with genotype-positive but phenotype-negative HCM. All patients with HCM should undergo, on an annual basis, careful personal and family history record, two-dimensional echocardiography, 12-lead ECG, 24–48-h ambulatory Holter electrocardiogram, and exercise stress testing (for evaluation of exercise tolerance, blood pressure, and ventricular tachyarrhythmias).

Pharmacological Therapy for Obstructive HCM

The primary goal of pharmacologic therapy in obstructive HCM is to decrease symptoms. Understanding the pathophysiology of the LVOTO is crucial to devising appropriate medication recommendations. The LVOTO in HCM is dynamic and highly load-dependent. The obstruction is augmented with decreases in preload (volume status) and afterload (vasodilation) and with increases in contractility. Simple physical exertion such as walking can cause all of these to occur simultaneously. As patients with HCM typically only have symptoms with effort, the goal of medical therapy is to decrease the effort-related augmentation of LVOTO.

Beta Blockers

Due to negative inotropic and negative chronotropic effects, beta blockers are the traditional mainstay of HCM therapy. Beta blockers are used in symptomatic patients with or without obstruction to control heart failure and anginal chest pain. The dose–response relationship of these medications varies significantly from patient to patient. Commonly used beta blockers include propranolol, atenolol, metoprolol, and nadolol (Elliott et al., 2004; Spirito et al., 2006).

Calcium-Blocker Therapy

The calcium channel antagonist verapamil is another drug used in HCM for its negative inotropic effect. It should be avoided in infants and used with caution in patients with heart failure and/or very significant obstruction (Maron et al., 2003). While both verapamil and diltiazem have negative inotropic and negative chronotropic effects, in some patients, vasodilatory action can predominate and paradoxically increase the severity of symptoms. Dihydropyridine-type calcium antagonists are pure vasodilators that should be avoided in patients with HCM.

Disopyramide

Disopyramide is a negative inotrope and type 1-A antiarrhythmic agent, which may help some patients with obstruction. It decreases cardiac output in nonobstructive HCM and is used primarily in patients not responding to beta blocker and/or calcium channel blocker therapy. There are no data that beta blockers, calcium channel blockers, or disopyramide alter the risk of sudden death (Maron et al., 2003).

Drugs to Be Used with Caution in HCM

Angiotensin-converting enzymes inhibitors (ACE inhibitors), angiotensin II blockers, nifedipine, and other pure afterload reducing agents should be used with caution, as afterload reduction may worsen LVOTO (Roberts et al., 2005; Spirito et al., 2006). Beta adrenergic agents like dopamine, dobutamine, or epinephrine and agents with increased inotropic activity may worsen LVOTO (Maron et al., 2003). Likewise, rapid or aggressive diuresis can decrease preload and worsen LVOTO.

Pharmacogenomics

Currently, there is no therapy available specifically designed for specific gene mutations underlying the disease nor a therapy that has been shown to reverse the hypertrophic process. Although polymorphisms in the renin-angiotensin-aldosterone system (RAAS) modify the phenotype of HCM, particularly *MYBPC3*-HCM (Perkins et al., 2005), a direct correlation with genotype-specific drug treatment has not been shown. However, with growing evidence that ACE inhibitors especially combined with low doses of aldosterone receptor blockers can prevent the progression of hypertrophy and fibrosis (Fraccarollo et al., 2003, 2005; Kalkman et al., 1999; Kambara et al., 2003; Monteiro de Resende et al., 2006), one can envision that, with increasing knowledge of the genomic background of HCM, specific therapies will emerge in the near future. In other cases for example, the proper and prompt recognition of an HCM phenocopy such as cardiac Fabry's disease can facilitate gene-specific pharmacotherapy including in this particular example, enzyme-replacement therapy. Albeit rare, such clinical sleuthing can enable early treatment and prevent the progression of the disease.

The Role and Impact of Non-Pharmacological Therapy

Septal Myectomy Surgery

Ventricular septal myectomy remains the gold standard for treating drug refractory, symptomatic obstructive HCM; a procedure during which a piece of hypertrophied septum is removed in order to relieve the obstruction (Maron et al., 2004; Nishimura et al., 2004; Spirito et al., 2006). Surgery is usually indicated in patients with peak instantaneous LVOT Doppler gradient of 50 mmHg or higher under rest or provocation and/or severely symptomatic patients (NYHA class 3 or 4, (Maron et al., 2003, 2004; Spirito et al., 2006). This profile represents approximately 5% of patients with HCM (Elliott et al., 2004). More extensive, extended septal myectomy involving the anterolateral papillary muscle and mitral valvuloplasty may be needed in patients with abnormal papillary muscle apparatus and mitral valve abnormalities (Ommen et al., 2005). The surgical mortality is <1% in most major centers (Maron et al., 2004; Poliac et al., 2006). Long-term survival after surgical myectomy is equal to that observed in the general population (Ommen et al., 2005). Surgery provides long-term improvement in LVOT gradient, mitral valve regurgitation, and symptomatic improvement (Maron et al., 2004; Ommen et al., 2005; Poliac et al., 2006).

Septal Ablation

Alcohol septal ablation technique using ethanol (95% alcohol 1–3 ml) is injected in specific septal branches of the left anterior descending artery producing a controlled septal infarction often providing dramatic symptomatic improvement in some patients (Faber et al., 1998; Gietzen et al., 1999; Kimmelstiel et al., 2004; Knight et al., 1997; Lakkis et al., 1998). The criteria for patient selection for alcohol septal ablation are similar to myectomy with the following *caveat* – the impact of alcohol septal ablation on SCD risk is unknown. Scarring associated with alcohol septal ablation may create a permanent arrhythmogenic substrate (Maron et al., 2003). Complications include complete atrioventricular block requiring permanent pacemakers (5–10% of patients), large myocardial infarction, acute mitral valve regurgitation, ventricular fibrillation (VF), and death (2–4%, Nishimura et al., 2004; Roberts et al., 2005; Spirito et al., 2006). Alcohol septal ablation is not suitable for patients with LVOTO secondary to abnormal mitral valve apparatus and unusual location of hypertrophy away from the area supplied by septal perforator. Given the unknown future risks of alcohol septal ablation, it is not recommended in children or young adults (Maron et al., 2003).

Implantable Cardioverter Defibrillator

The implantable cardiac defibrillator (ICD) plays an important role in primary and secondary prevention of sudden death of patient with HCM. The three main functions of the ICD are detection of arrhythmia; delivery of appropriate electrical therapy; and storage of diagnostic information, including electrocardiograms and details of treated episodes. In a multicenter study of ICDs in patients with HCM, the device intervened appropriately, terminating VT/VF, at a rate of

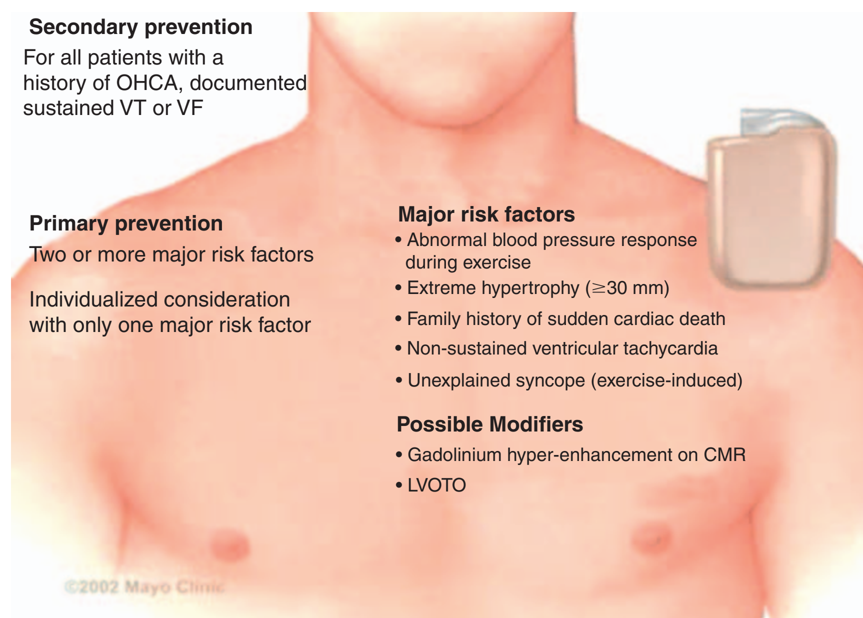


Figure 28.4 The major indications for implantation of an ICD in HCM: secondary prevention or primary prevention, and major risk factors. CMR, cardiac magnetic resonance; LVOTO, left ventricular outflow tract obstruction; OHCA, out of hospital cardiac arrest; VF, ventricular fibrillation; VT, ventricular tachycardia.

5% per year for those patients implanted as primary prevention and 11% per year for secondary prevention, over an average follow-up of 3 years (Maron et al., 2000). The indications for implantation of an ICD are listed in Figure 28.4 and can be summarized as secondary prevention for all patients with a history of out of hospital cardiac arrest (OHCA), documented sustained VT or VF, or as primary prevention in patients with two or more major risk factors.

Dual-Chamber Pacing

There has been a great deal of debate surrounding the use of pacing as a means of relieving ventricular obstruction. Some studies have shown a beneficial effect, while others demonstrated significant placebo effect (Maron et al., 2003). The average decrease in LVOTO gradient with pacing ranged from a modest 25% to 40% and varied substantially (Maron et al., 2003). There is evidence to suggest that appropriately used dual-chamber pacing may decrease LVOT gradient and provide symptomatic relief (Maron et al., 1999). Thus, there may be a limited role of dual-chamber pacing in a select group of patients, for example patients with advanced age (>65 years) and higher surgical risk. There is no evidence to suggest any change in SCD risk or disease progression (Boriani et al., 2004).

Maze Procedures

Surgical Maze procedure combined with myectomy may be a feasible therapeutic option in HCM with LVOTO and atrial fibrillation (AF). There are small case series reporting low operative mortality and morbidity and a high likelihood of patients remaining in sinus rhythm post procedure (Chen et al., 2004). Larger studies with longer follow-up are needed to better define the risks and benefits of surgical Maze procedure in HCM.

2009 UPDATE

With research ongoing and recognizing that approximately 20% of reverse curve-HCM and over 80% of sigmoidal-HCM remains genetically unexplained, novel HCM-susceptibility genes have been discovered recently. In 2007, Arimura et al. found two mutations in one patient with HCM in *OBSCN*-encoded obscurin that altered interaction with the giant myofilament protein titin (Arimura et al., 2007). In a larger study, 3 of 384 patients with HCM were mutation-positive in the *ANKRD1*-encoded cardiac ankyrin repeat protein (Arimura et al., 2009).

The release of the complete human genome sequence and the enormity of variation in individuals show a growing role for modifier genes and the search for effect by genome-wide studies. In 2007, Daw et al. identified multiple loci with suggestive linkage for modifying cardiac hypertrophy. Effect sizes on left ventricular mass on this cohort of 100 patients ranged from an ~8 g shift attributed to one locus for the common allele to 90 g shift for another locus' uncommon allele (Daw et al., 2007). In 2008, sex hormone polymorphisms were shown to modify the HCM phenotype (Lind et al., 2008). Fewer CAG repeats in *AR*-encoded androgen receptor were associated with greater myocardial wall thickness in males ($p=0.008$), and male carriers of the A-allele in the promoter of *ESR1*-encoded estrogen receptor 1 (SNP rs6915267) exhibited a 11% decrease in LV wall thickness ($p=0.047$) compared to GG-homozygote males (Lind et al., 2008). HCM modifier polymorphisms like these could contribute to the clinical differences observed between men and women with HCM (Bos et al., 2008; Olivotto et al., 2005).

A recent longitudinal study in a large cohort of Italian patients with HCM has shown an increased risk of cardiovascular death, nonfatal stroke, or progression to NYHA class III/IV among patients with a positive HCM genetic test involving any of the myofilament genes compared to those patients with a negative genetic test (25% versus 7%, respectively; $p=0.002$). Multivariate analysis showed that myofilament positive HCM (i.e., a positive genetic test) was the strongest predictor of an adverse outcome (hazard ratio 4.27 (CI 1.43 – 12.48, $p=0.008$, Olivotto et al., 2008). Furthermore, patients with a positive genetic test had greater probability of developing severe LV systolic dysfunction ($p=0.021$) and restrictive LV filling ($p=0.018$, Olivotto et al., 2008).

In the emerging field of microRNA's (miR's) and their role in cardiac development and (hypertrophic) heart disease, several new studies linking miR's and hypertrophy have been published. Utilizing two mouse models of pathological hypertrophy – transverse aortic constriction (TAC) and calcineurin transgenic mice – 6 miR's were up-regulated, which *in vitro,* were sufficient to induce hypertrophic growth of cardiomyocytes (van Rooij et al., 2006). A transgenic mouse model over-expressing miR-195 showed that a single miRNA could induce pathological hypertrophy and heart failure (van Rooij et al., 2006). Multiple studies have since been published with miRNA expression profiles in different settings, *in vivo* and *in vitro*, of cardiac hypertrophy demonstrating the potential role of miR's in (hypertrophic) heart disease and possibly HCM (Ikeda et al., 2007; Sayed et al., 2007; Tatsuguchi et al., 2007; Thum et al., 2007; van Rooij et al., 2006, 2007).

2009 UPDATE REFERENCES

Arimura, T., Bos, J.M., Sato, A., Kubo, T., Okamoto, H., Nishi, H., Harada, H., Koga, Y., Moulik, M., Doi, Y. et al. (2009). Ankrd1-encoded cardiac ankyrin repeat protein mutations in hypertrophic cardiomyopathy. *J Am Coll Cardiol*.

Arimura, T., Matsumoto, Y., Okazaki, O., Hayashi, T., Takahashi, M., Inagaki, N., Hinohara, K., Ashizawa, N., Yano, K. and Kimura, A. (2007). Structural analysis of obscurin gene in hypertrophic cardiomyopathy. *Biochem Biophys Res Commun* 362(2), 281–287.

Bos, J.M., Theis, J.L., Tajik, A.J., Gersh, B.J., Ommen, S.R. and Ackerman, M.J. (2008). Relationship between sex, shape, and substrate in hypertrophic cardiomyopathy. *Am Heart J* 155(6), 1128–1134.

Daw, E.W., Chen, S.N., Czernuszewicz, G., Lombardi, R., Lu, Y., Ma, J., Roberts, R., Shete, S. and Marian, A.J. (2007). Genome-wide mapping of modifier chromosomal loci for human hypertrophic cardiomyopathy. *Hum Mol Genet* 16(20), 2463–2471.

Ikeda, S., Kong, S.W., Lu, J., Bisping, E., Zhang, H., Allen, P.D., Golub, T.R., Pieske, B. and Pu, W.T. (2007). Altered microRNA expression in human heart disease. *Physiol Genomics* 31(3), 367–373.

Lind, J.M., Chiu, C., Ingles, J., Yeates, L., Humphries, S.E., Heather, A.K. and Semsarian, C. (2008). Sex hormone receptor gene variation associated with phenotype in male hypertrophic cardiomyopathy patients. *J Mol Cell Cardiol* 45(2), 217–222.

Olivotto, I., Girolami, F., Ackerman, M.J., Nistri, S., Bos, J.M., Zachara, E., Ommen, S.R., Theis, J.L., Vaubel, R.A., Re, F. et al. (2008). Myofilament protein gene mutation screening and outcome of patients with hypertrophic cardiomyopathy. *Mayo Clin Proc* 83(6), 630–638.

Olivotto, I., Maron, M.S., Adabag, A.S., Casey, S.A., Vargiu, D., Link, M.S., Udelson, J.E., Cecchi, F. and Maron, B.J. (2005). Gender-related differences in the clinical presentation and outcome of hypertrophic cardiomyopathy. *J Am Coll Cardiol* 46(3), 480–487.

Sayed, D., Hong, C., Chen, I.Y., Lypowy, J. and Abdellatif, M. (2007). MicroRNAs play an essential role in the development of cardiac hypertrophy. *Circ Res* 100(3), 416–424.

Tatsuguchi, M., Seok, H.Y., Callis, T.E., Thomson, J.M., Chen, J.F., Newman, M., Rojas, M., Hammond, S.M. and Wang, D.Z. (2007). Expression of micrornas is dynamically regulated during cardiomyocyte hypertrophy. *J Mol Cell Cardiol* 42(6), 1137–1141.

Thum, T., Galuppo, P., Wolf, C., Fiedler, J., Kneitz, S., van Laake, L.W., Doevendans, P.A., Mummery, C.L., Borlak, J., Haverich, A. et al. (2007). Micrornas in the human heart: A clue to fetal gene reprogramming in heart failure. *Circulation* 116(3), 258–267.

van Rooij, E., Sutherland, L.B., Liu, N., Williams, A.H., McAnally, J., Gerard, R.D., Richardson, J.A. and Olson, E.N. (2006). A signature pattern of stress-responsive micrornas that can evoke cardiac hypertrophy and heart failure. *Proc Natl Acad Sci USA* 103(48), 18255–18260.

van Rooij, E., Sutherland, L.B., Qi, X., Richardson, J.A., Hill, J. and Olson, E.N. (2007). Control of stress-dependent cardiac growth and gene expression by a microrna. *Science* 316(5824), 575–579.

REFERENCES

Abelmann, W.H. (1984). Classification and natural history of primary myocardial disease. *Prog Cardiovasc Dis* 27(2), 73–94.

Ackerman, M.J., Van Driest, S.V., Ommen, S.R., Will, M.L., Nishimura, R.A., Tajik, A.J. and Gersh, B.J. (2002). Prevalence and age-dependence of malignant mutations in the beta-myosin heavy chain and troponin T gene in hypertrophic cardiomyopathy: A comprehensive outpatient perspective. *J Am Coll Cardiol* 39(12), 2042–2048.

Anan, R., Greve, G., Thierfelder, L., Watkins, H., McKenna, W.J., Solomon, S., Vecchio, C., Shono, H., Nakao, S., Tanaka, H. et al. (1994). Prognostic implications of novel beta cardiac myosin heavy chain gene mutations that cause familial hypertrophic cardiomyopathy. *J Clin Invest* 93(1), 280–285.

Arad, M., Benson, D.W., Perez-Atayde, A.R., McKenna, W.J., Sparks, E.A., Kanter, R.J., McGarry, K., Seidman, J.G. and Seidman, C.E. (2002). Constitutively active AMP kinase mutations cause glycogen storage disease mimicking hypertrophic cardiomyopathy. *J Clin Invest* 109(3), 357–362.

Arad, M., Maron, B.J., Gorham, J.M., Johnson, W.H., Jr, Saul, J.P., Perez-Atayde, A.R., Spirito, P., Wright, G.B., Kanter, R.J., Seidman, C.E. et al. (2005). Glycogen storage diseases presenting as hypertrophic cardiomyopathy. *N Engl J Med* 352(4), 362–372.

Binder, J., Ommen, S.R., Gersh, B.J., Van Driest, S.L., Tajik, A.J., Nishimura, R.A. and Ackerman, M.J. (2006). Echocardiography-guided genetic testing in hypertrophic cardiomyopathy: septal morphological features predict the presence of myofilament mutations. *Mayo Clin Proc* 81(4), 459–467.

Blair, E., Redwood, C., Ashrafian, H., Oliveira, M., Broxholme, J., Kerr, B., Salmon, A., Ostman-Smith, I. and Watkins, H. (2001). Mutations in the gamma(2) subunit of AMP-activated protein kinase cause familial hypertrophic cardiomyopathy: Evidence for the central role of energy compromise in disease pathogenesis. *Hum Mol Genet* 10(11), 1215–1220.

Boriani, G., Maron, B.J., Shen, W.K. and Spirito, P. (2004). Prevention of sudden death in hypertrophic cardiomyopathy: But which defibrillator for which patient? *Circulation* 110(15), e438–e442.

Bos, J.M., Poley, R.N., Ny, M., Tester, D.J., Xu, X., Vatta, M., Towbin, J.A., Gersh, B.J., Ommen, S.R. and Ackerman, M.J. (2006). Genotype-phenotype relationships involving hypertrophic cardiomyopathy-associated mutations in titin, muscle Lim protein, and telethonin. *Mol Genet Metab* 88(1), 78–85.

Braunwald, E., Lambrew, C.T., Rockoff, S.D., Ross, J., Jr and Morrow, A.G. (1964). Idiopathic hypertrophic subaortic stenosis. I. A description of the disease based upon an analysis of 64 patients. *Circulation* 30(Suppl 4), 3–119.

Chen, M.S., McCarthy, P.M., Lever, H.M., Smedira, N.G. and Lytle, B.L. (2004). Effectiveness of atrial fibrillation surgery in patients with hypertrophic cardiomyopathy. *Am J Cardiol* 93(3), 373–375.

Churchill, G.A. (2002). Fundamentals of experimental design for cDNA microarrays. *Nat Genet* 32(Suppl), 490–495.

Coviello, D.A., Maron, B.J., Spirito, P., Watkins, H., Vosberg, H.P., Thierfelder, L., Schoen, F.J., Seidman, J.G. and Seidman, C.E. (1997). Clinical features of hypertrophic cardiomyopathy caused by mutation of a hot spot in the alpha-tropomyosin gene. *J Am Coll Cardiol* 29(3), 635–640.

Daehmlow, S., Erdmann, J., Knueppel, T., Gille, C., Froemmel, C., Hummel, M., Hetzer, R. and Regitz-Zagrosek, V. (2002). Novel mutations in sarcomeric protein genes in dilated cardiomyopathy. *Biochem Biophys Res Commun* 298(1), 116–120.

Elliott, P. and McKenna, W.J. (2004). Hypertrophic cardiomyopathy. *Lancet* 363(9424), 1881–1891.

Elliott, P.M., Poloniecki, J., Dickie, S., Sharma, S., Monserrat, L., Varnava, A., Mahon, N.G. and McKenna, W.J. (2000). Sudden death in hypertrophic cardiomyopathy: Identification of high risk patients. *J Am Coll Cardiol* 36(7), 2212–2218.

Faber, L., Seggewiss, H. and Gleichmann, U. (1998). Percutaneous transluminal septal myocardial ablation in hypertrophic obstructive cardiomyopathy: Results with respect to intraprocedural myocardial contrast echocardiography. *Circulation* 98(22), 2415–2421.

Fraccarollo, D., Galuppo, P., Hildemann, S., Christ, M., Ertl, G. and Bauersachs, J. (2003). Additive improvement of left ventricular remodeling and neurohormonal activation by aldosterone receptor blockade with eplerenone and ACE inhibition in rats with myocardial infarction. *J Am Coll Cardiol* 42(9), 1666–1673.

Fraccarollo, D., Galuppo, P., Schmidt, I., Ertl, G. and Bauersachs, J. (2005). Additive amelioration of left ventricular remodeling and molecular alterations by combined aldosterone and angiotensin receptor blockade after myocardial infarction. *Cardiovasc Res* 67(1), 97–105.

Frank, D., Kuhn, C., Katus, H.A. and Frey, N. (2006). The sarcomeric Z-disc: A nodal point in signalling and disease. *J Mol Med*, 1–23.

Frey, N., Katus, H.A., Olson, E.N. and Hill, J.A. (2004). Hypertrophy of the heart: A new therapeutic target?. *Circulation* 109(13), 1580–1589.

Fujino, N., Ino, H., Hayashi, K., Uchiyama, K., Nagata, M., Konno, T., Katoh, H., Sakamoto, Y. and Tsubokawa, T. (2006). A novel missense mutation in cardiac ryanodine receptor gene as a possible cause of hypertrophic cardiomyopathy: Evidence from familial analysis. *Circulation* 114(18). II-165–915.

Geier, C., Perrot, A., Ozcelik, C., Binner, P., Counsell, D., Hoffmann, K., Pilz, B., Martiniak, Y., Gehmlich, K., van der Ven, P.F. et al. (2003). Mutations in the human muscle Lim protein gene in families with hypertrophic cardiomyopathy. *Circulation* 107(10), 1390–1395.

Geisterfer-Lowrance, A.A., Kass, S., Tanigawa, G., Vosberg, H., McKenna, W., Seidman, C.E. and Seidman, J.G. (1990). A molecular basis for familial hypertrophic cardiomyopathy: A beta cardiac myosin heavy chain gene missense mutation. *Cell* 62, 999–1006.

Gerull, B., Gramlich, M., Atherton, J., McNabb, M., Trombitas, K., Sasse-Klaassen, S., Seidman, J.G., Seidman, C., Granzier, H., Labeit, S. et al. (2002). Mutations of TTN, encoding the giant muscle filament titin, cause familial dilated cardiomyopathy. *Nat Genet* 30(2), 201–204.

Gietzen, F.H., Leuner, C.J., Raute-Kreinsen, U., Dellmann, A., Hegselmann, J., Strunk-Mueller, C. and Kuhn, H.J. (1999). Acute and long-term results after transcoronary ablation of septal hypertrophy (TASH). Catheter interventional treatment for hypertrophic obstructive cardiomyopathy. *Eur Heart J* 20(18), 1342–1354.

Haghighi, K., Kolokathis, F., Gramolini, A.O., Waggoner, J.R., Pater, L., Lynch, R.A., Fan, G.C., Tsiapras, D., Parekh, R.R., Dorn, G.W., 2nd et al. (2006). A mutation in the human phospholamban gene, deleting arginine 14, results in lethal, hereditary cardiomyopathy. *Proc Natl Acad Sci USA* 103(5), 1388–1393.

Hayashi, T., Arimura, T., Itoh-Satoh, M., Ueda, K., Hohda, S., Inagaki, N., Takahashi, M., Hori, H., Yasunami, M., Nishi, H. et al. (2004). TCAP gene mutations in hypertrophic cardiomyopathy and dilated cardiomyopathy. *J Am Coll Cardiol* 44(11), 2192–2201.

Holland, M.J. (2002). Transcript abundance in yeast varies over six orders of magnitude. *J Biol Chem* 277(17), 14363–14366.

Hwang, J.J., Allen, P.D., Tseng, G.C., Lam, C.W., Fananapazir, L., Dzau, V.J. and Liew, C.C. (2002). Microarray gene expression profiles in dilated and hypertrophic cardiomyopathic end-stage heart failure. *Physiol Genomics* 10(1), 31–44.

Jarcho, J.A., McKenna, W., Pare, J.A., Solomon, S.D., Holcombe, R.F., Dickie, S., Levi, T., Donis-Keller, H., Seidman, J.G. and Seidman, C.E. (1989). Mapping a gene for familial hypertrophic cardiomyopathy to chromosome 14q1. *N Engl J Med* 321(20), 1372–1378.

Kalkman, E.A., van Haren, P., Saxena, P.R. and Schoemaker, R.G. (1999). Early captopril prevents myocardial infarction-induced hypertrophy but not angiogenesis. *Eur J Pharmacol* 369(3), 339–348.

Kambara, A., Holycross, B.J., Wung, P., Schanbacher, B., Ghosh, S., McCune, S.A., Bauer, J.A. and Kwiatkowski, P. (2003). Combined effects of low-dose oral spironolactone and captopril therapy in a rat model of spontaneous hypertension and heart failure. *J Cardiovasc Pharmacol* 41(6), 830–837.

Kamisago, M., Sharma, S.D., DePalma, S.R., Solomon, S., Sharma, P., McDonough, B., Smoot, L., Mullen, M.P., Woolf, P.K., Wigle, E.D. et al. (2000). Mutations in sarcomere protein genes as a cause of dilated cardiomyopathy. *N Engl J Med* 343(23), 1688–1696.

Kim, J.B., Porreca, G.J., Song, L., Greenway, S.C., Gorham, J.M., Church, G.M., Seidman, C.E. and Seidman, J.G. (2007). Polony multiplex analysis of gene expression (PMAGE) in mouse hypertrophic cardiomyopathy. *Science* 316(5830), 1481–1484.

Kimmelstiel, C.D. and Maron, B.J. (2004). Role of Percutaneous Septal Ablation in Hypertrophic Obstructive Cardiomyopathy. *Circulation* 109(4), 452–456.

Kimura, A., Harada, H., Park, J.E., Nishi, H., Satoh, M., Takahashi, M., Hiroi, S., Sasaoka, T., Ohbuchi, N., Nakamura, T. et al. (1997). Mutations in the cardiac troponin I gene associated with hypertrophic cardiomyopathy. *Nat Genet* 16(4), 379–382.

Knight, C., Kurbaan, A.S., Seggewiss, H., Henein, M., Gunning, M., Harrington, D., Fassbender, D., Gleichmann, U. and Sigwart, U. (1997). Nonsurgical septal reduction for hypertrophic obstructive cardiomyopathy: Outcome in the first series of patients. *Circulation* 95(8), 2075–2081.

Knoll, R., Hoshijima, M., Hoffman, H.M., Person, V., Lorenzen-Schmidt, I., Bang, M.L., Hayashi, T., Shiga, N., Yasukawa, H., Schaper, W. et al. (2002). The cardiac mechanical stretch sensor machinery involves a Z disc complex that is defective in a subset of human dilated cardiomyopathy. *Cell* 111(7), 943–955.

Lakkis, N.M., Nagueh, S.F., Kleiman, N.S., Killip, D., He, Z.X., Verani, M.S., Roberts, R. and Spencer, W.H., 3rd (1998). Echocardiography-guided ethanol septal reduction for hypertrophic obstructive cardiomyopathy. *Circulation* 98(17), 1750–1755.

Landstrom, A.P., Parvatiyar, M.S., Pinto, J.R., Marquardt, M.L., Bos, J.M., Ommen, S.R., Potter, J.D., Ackerman, M.J. (2008). Defining *TNNC1*-encoded cardiac Troponin C as novel target gene for hypertrophic cardiomyopathy. *J Mol Cell Cardiol* In Press.

Landstrom, A., Batalden, K.B., Weisleder, N., Bos, J.M., Tester, D.J., Ommen, S.R., Wehrens, X.H.T., Claycomb, W.C., Ko, J., Hwang, M. et al. (2007). Mutations in *JPH2*-encoded junctophilin-2 associated with hypertrophic cardiomyopathy in humans. *J Mol Cell Cardiol* 42, 1026–1035.

Lee, R.C., Feinbaum, R.L. and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75(5), 843–854.

Lever, H.M., Karam, R.F., Currie, P.J. and Healy, B.P. (1989). Hypertrophic cardiomyopathy in the elderly. Distinctions from the young based on cardiac shape. *Circulation* 79(3), 580–589.

Maron, B.J. (2002). Hypertrophic cardiomyopathy: A systematic review. *JAMA* 287(10), 1308–1320.

Maron, B.J., Nishimura, R.A., McKenna, W.J., Rakowski, H., Josephson, M.E. and Kieval, R.S. (1999). Assessment of permanent dual-chamber pacing as a treatment for drug-refractory symptomatic patients with obstructive hypertrophic cardiomyopathy. A randomized, double-Blind, crossover study (M-PATHY). *Circulation* 99(22), 2927–2933.

Maron, B.J., Shen, W.K., Link, M.S., Epstein, A.E., Almquist, A.K., Daubert, J.P., Bardy, G.H., Favale, S., Rea, R.F., Boriani, G. et al. (2000). Efficacy of implantable cardioverter-defibrillators for the prevention of sudden death in patients with hypertrophic cardiomyopathy [See Comments]. *N Engl J Med* 342(6), 365–373.

Maron, B.J., McKenna, W.J., Danielson, G.K., Kappenberger, L.J., Kuhn, H.J., Seidman, C.E., Shah, P.M., Spencer, W.H., 3rd, Spirito, P., Ten Cate, F.J. et al. (2003). American College of Cardiology/European Society of Cardiology clinical expert consensus document on hypertrophic cardiomyopathy. A Report of the American College of Cardiology Foundation Task Force on Clinical Expert Consensus Documents and the European Society of Cardiology Committee for Practice Guidelines. *J Am Coll Cardiol* 42(9), 1687–1713.

Maron, B.J., Dearani, J.A., Ommen, S.R., Maron, M.S., Schaff, H.V., Gersh, B.J. and Nishimura, R.A. (2004). The case for surgery in obstructive hypertrophic cardiomyopathy. *J Am Coll Cardiol* 44(10), 2044–2053.

Maron, B.J., Ackerman, M.J., Nishimura, R.A., Pyeritz, R.E., Towbin, J.A. and Udelson, J.E. (2005). Task Force 4: HCM and other cardiomyopathies, mitral valve prolapse, myocarditis, and Marfan syndrome. *J Am Coll Cardiol* 45(8), 1340–1345.

Minamisawa, S., Sato, Y., Tatsuguchi, Y., Fujino, T., Imamura, S., Uetsuka, Y., Nakazawa, M. and Matsuoka, R. (2003). Mutation of the phospholamban promoter associated with hypertrophic cardiomyopathy. *Biochem Biophys Res Commun* 304(1), 1–4.

Mogensen, J., Klausen, I.C., Pedersen, A.K., Egeblad, H., Bross, P., Kruse, T.A., Gregersen, N., Hansen, P.S., Daandrup, U. and Borglum, A.D. (1999). Alpha-cardiac cctin is a novel disease gene in familial hypertrophic cardiomyopathy. *J Clin Invest* 103(10), R39–R43.

Mohapatra, B., Jimenez, S., Lin, J.H., Bowles, K.R., Coveler, K.J., Marx, J.G., Chrisco, M.A., Murphy, R.T., Lurie, P.R., Schwartz, R.J. et al. (2003). Mutations in the muscle Lim protein and alpha-actinin-2 genes in dilated cardiomyopathy and endocardial fibroelastosis. *Mol Genet Metab* 80(1-2), 207–215.

Monteiro de Resende, M., Kriegel, A.J. and Greene, A.S. (2006). Combined effects of low-dose spironolactone and captopril therapy in a rat model of genetic hypertrophic cardiomyopathy. *J Cardiovasc Pharmacol* 48(6), 265–273.

Moolman, J.C., Corfield, V.A., Posen, B., Ngumbela, K., Seidman, C., Brink, P.A. and Watkins, H. (1997). Sudden death due to troponin T mutations. *J Am Coll Cardiol* 29(3), 549–555.

Moon, J.C., McKenna, W.J., McCrohon, J.A., Elliott, P.M., Smith, G.C. and Pennell, D.J. (2003). Toward clinical risk assessment in hypertrophic cardiomyopathy with gadolinium cardiovascular magnetic resonance. *J Am Coll Cardiol* 41(9), 1561–1567.

Morita, H., Larson, M.G., Barr, S.C., Vasan, R.S., O'Donnell, C.J., Hirschhorn, J.N., Levy, D., Corey, D., Seidman, C.E., Seidman, J.G. et al. (2006). Single-gene mutations and increased left ventricular wall thickness in the community: The framingham heart study. *Circulation* 113(23), 2697–2705.

Nakao, S., Takenaka, T., Maeda, M., Kodama, C., Tanaka, A., Tahara, M., Yoshida, A., Kuriyama, M., Hayashibe, H., Sakuraba, H. et al. (1995). An atypical variant of Fabry's disease in men with left ventricular hypertrophy. *N Engl J Med* 333(5), 288–293.

Niimura, H., Patton, K.K., McKenna, W.J., Soults, J., Maron, B.J., Seidman, J.G. and Seidman, C.E. (2002). Sarcomere protein gene mutations in hypertrophic cardiomyopathy of the elderly. *Circulation* 105(4), 446–451.

Nishimura, R.A. and Holmes, D.R., Jr (2004). Clinical practice. Hypertrophic obstructive cardiomyopathy. *N Engl J Med* 350(13), 1320–1327.

Olson, T.M., Doan, T.P., Kishimoto, N.Y., Whitby, F.G., Ackerman, M.J. and Fananapazir, L. (2000). Inherited and *de novo* mutations in the cardiac actin gene cause hypertrophic cardiomyopathy. *J Mol Cell Cardiol* 32, 1687–1694.

Olson, T.M., Illenberger, S., Kishimoto, N.Y., Huttelmaier, S., Keating, M.T. and Jockusch, B.M. (2002a). Metavinculin mutations alter actin interaction in dilated cardiomyopathy. *Circulation* 105(4), 431–437.

Olson, T.M., Karst, M.L., Whitby, F.G. and Driscoll, D.J. (2002b). Myosin light chain mutation causes autosomal recessive cardiomyopathy with mid-cavitary hypertrophy and restrictive physiology. *Circulation* 105(20), 2337–2340.

Olson, T.M., Kishimoto, N.Y., Whitby, F.G. and Michels, V.V. (2001). Mutations that alter the surface charge of alpha-tropomyosin are associated with dilated cardiomyopathy. *J Mol Cell Cardiol* 33(4), 723–732.

Ommen, S.R., Maron, B.J., Olivotto, I., Maron, M.S., Cecchi, F., Betocchi, S., Gersh, B.J., Ackerman, M.J., McCully, R.B., Dearani, J.A. et al. (2005). Long-term effects of surgical septal myectomy on survival in patients with obstructive hypertrophic cardiomyopathy. *J Am Coll Cardiol* 46(3), 470–476.

Osio, A., Tan, L., Chen, S.N., Lombardi, R., Nagueh, S.F., Shete, S., Roberts, R., Willerson, J.T. and Marian, A.J. (2007). Myozenin 2 is a novel gene for human hypertrophic cardiomyopathy. *Circ Res*.

Perkins, M.J., Van Driest, S.L., Ellsworth, E.G., Will, M.L., Gersh, B.J., Ommen, S.R. and Ackerman, M.J. (2005). Gene-specific modifying effects of pro-LVH polymorphisms involving the renin-angiotensin-aldosterone system among 389 unrelated patients with hypertrophic cardiomyopathy. *Eur Heart J* 26(22), 2457–2462.

Poetter, K., Jiang, H., Hassanzadeh, S., Master, S.R., Chang, A., Dalakas, M.C., Rayment, I., Sellers, J.R., Fananapazir, L. and Epstein, N.D. (1996). Mutations in either the essential or regulatory light chains of myosin are associated with a rare myopathy in human heart and skeletal muscle. *Nat Genet* 13(1), 63–69.

Poliac, L.C., Barron, M.E. and Maron, B.J. (2006). Hypertrophic cardiomyopathy. *Anesthesiology* 104(1), 183–192.

Pollick, C., Morgan, C.D., Gilbert, B.W., Rakowski, H. and Wigle, E.D. (1982). Muscular subaortic stenosis: The temporal relationship between systolic anterior motion of the anterior mitral leaflet and the pressure gradient. *Circulation* 66(5), 1087–1094.

Pyle, W.G. and Solaro, R.J. (2004). At the crossroads of myocardial signaling: The role of Z-discs in intracellular signaling and cardiac function. *Circ Res* 94(3), 296–305.

Rajan, S., Williams, S.S., Jagatheesan, G., Ahmed, R.P., Fuller-Bicer, G., Schwartz, A., Aronow, B.J. and Wieczorek, D.F. (2006). Microarray analysis of gene expression during early stages of mild and severe cardiac hypertrophy. *Physiol Genomics* 27(3), 309–317.

Richard, P., Charron, P., Carrier, L., Ledeuil, C., Cheav, T., Pichereau, C., Benaiche, A., Isnard, R., Dubourg, O., Burban, M. et al. (2003). Hypertrophic cardiomyopathy: Distribution of disease genes, spectrum of mutations, and implications for a molecular diagnosis strategy. *Circulation* 107(17), 2227–2232.

Richardson, P., McKenna, W., Bristow, M., Maisch, B., Mautner, B., O'Connell, J., Olsen, E., Thiene, G., Goodwin, J., Gyarfas, I. et al. (1996). Report of the 1995 World Health Organization/International Society and Federation of Cardiology Task Force on the Definition and Classification of Cardiomyopathies. *Circulation* 93(5), 841–842.

Roberts, R. and Sigwart, U. (2005). Current concepts of the pathogenesis and treatment of hypertrophic cardiomyopathy. *Circulation* 112(2), 293–296.

Sachdev, B., Takenaka, T., Teraguchi, H., Tei, C., Lee, P., McKenna, W.J. and Elliott, P.M. (2002). Prevalence of Anderson-Fabry disease in male patients with late onset hypertrophic cardiomyopathy. *Circulation* 105(12), 1407–1411.

Sakuraba, H., Oshima, A., Fukuhara, Y., Shimmoto, M., Nagao, Y., Bishop, D.F., Desnick, R.J. and Suzuki, Y. (1990). Identification of point mutations in the alpha-galactosidase a gene in classical and atypical hemizygotes with Fabry disease. *Am J Hum Genet* 47(5), 784–789.

Satoh, M., Takahashi, M., Sakamoto, T., Hiroe, M., Marumo, F. and Kimura, A. (1999). Structural analysis of the titin gene in hypertrophic cardiomyopathy: Identification of a novel disease gene. *Biochem Biophys Res Commun* 262, 411–417.

Sayed, D., Hong, C., Chen, I.Y., Lypowy, J. and Abdellatif, M. (2007). Micrornas play an essential role in the development of cardiac hypertrophy. *Circ Res* 100(3), 416–424.

Schoendube, F.A., Klues, H.G., Reith, S., Flachskampf, F.A., Hanrath, P. and Messmer, B.J. (1995). Long-term clinical and echocardiographic follow-up after surgical correction of hypertrophic obstructive cardiomyopathy with extended myectomy and reconstruction of the subvalvular mitral apparatus. *Circulation* 92(9 Suppl), II122–II127.

Seidman, J.G. and Seidman, C. (2001). The genetic basis for cardiomyopathy: From mutation identification to mechanistic paradigms. *Cell* 104(4), 557–567.

Solomon, S.D., Wolff, S., Watkins, H., Ridker, P.M., Come, P., McKenna, W.J., Seidman, C.E. and Lee, R.T. (1993). Left ventricular hypertrophy and morphology in familial hypertrophic cardiomyopathy associated with mutations of the beta-myosin heavy chain gene. *J Am Coll Cardiol* 22(2), 498–505.

Spirito, P. and Autore, C. (2006). Management of hypertrophic cardiomyopathy. *BMJ* 332(7552), 1251–1255.

Teare, D. (1958). Asymmetrical hypertrophy of the heart in young adults. *Br Heart J* 20(1), 1–8.

Theis, J.L., Martijn Bos, J., Bartleson, V.B., Will, M.L., Binder, J., Vatta, M., Towbin, J.A., Gersh, B.J., Ommen, S.R. and Ackerman, M.J. (2006). Echocardiographic-determined septal morphology in Z-disc hypertrophic cardiomyopathy. *Biochem Biophys Res Commun* 351(4), 896–902.

Thierfelder, L., Watkins, H., MacRae, C., Lamas, R., McKenna, W., Vosberg, H.P., Seidman, J.G. and Seidman, C.E. (1994). Alpha-tropomyosin and cardiac troponin T mutations cause familial hypertrophic cardiomyopathy: A disease of the sarcomere. *Cell* 77(5), 701–712.

Van Driest, S.V., Ackerman, M.J., Ommen, S.R., Shakur, R., Will, M.L., Nishimura, R.A., Tajik, A.J. and Gersh, B.J. (2002). Prevalence and severity of "Benign" mutations in the beta myosin heavy chain, cardiac troponin-T, and alpha tropomyosin genes in hypertrophic cardiomyopathy. *Circulation* 106, 3085–3090.

Van Driest, S.L., Ellsworth, E.G., Ommen, S.R., Tajik, A.J., Gersh, B.J. and Ackerman, M.J. (2003). Prevalence and spectrum of thin filament mutations in an outpatient referral population with hypertrophic cardiomyopathy. *Circulation* 108, 445–451.

Van Driest, S.L., Jaeger, M.A., Ommen, S.R., Will, M.L., Gersh, B.J., Tajik, A.J. and Ackerman, M.J. (2004). Comprehensive analysis of the beta-myosin heavy chain gene in 389 unrelated patients with hypertrophic cardiomyopathy. *J Am Coll Cardiol* 44(3), 602–610.

Van Driest, S.L., Maron, B.J. and Ackerman, M.J. (2004). From malignant mutations to malignant domains: The continuing search for prognostic significance in the mutant genes causing hypertrophic cardiomyopathy. [Comment]. *Heart (British Cardiac Society)* 90(1), 7–8.

Van Driest, S.L., Vasile, V.C., Ommen, S.R., Will, M.L., Gersh, B.J., Nishimura, R.A., Tajik, A.J. and Ackerman, M.J. (2004). Myosin binding protein C mutations and compound herterozygosity in hypertrophic cardiomyopathy. *J Am Coll Cardiol* 44(9), 1903–1910.

Van Driest, S.L., Ommen, S.R., Tajik, A.J., Gersh, B.J. and Ackerman, M.J. (2005). Sarcomeric genotyping in hypertrophic cardiomyopathy. *Mayo Clin Proc* 80(4), 463–469.

Van Driest, S.L., Gakh, O., Ommen, S.R., Isaya, G. and Ackerman, M.J. (2005). Molecular and functional characterization of a human frataxin mutation found in hypertrophic cardiomyopathy. *Mol Genet Metab* 2, 1096–7192.

Varnava, A., Baboonian, C., Davison, F., de Cruz, L., Elliott, P.M., Davies, M.J. and McKenna, W.J. (1999). A new mutation of the cardiac troponin T gene causing familial hypertrophic cardiomyopathy without left ventricular hypertrophy. *Heart* 82(5), 621–624.

Vasile, V.C., Will, M.L., Ommen, S.R., Edwards, W.D., Olson, T.M. and Ackerman, M.J. (2006a). Identification of a metavinculin missense mutation, R975w, associated with both hypertrophic and dilated cardiomyopathy. *Mol Genet Metab* 87(2), 169–174.

Vasile, V.C., Ommen, S.R., Edwards, W.D. and Ackerman, M.J. (2006b). A missense mutation in a ubiquitously expressed protein, vinculin, confers susceptibility to hypertrophic cardiomyopathy. *Biochem Biophys Res Commun* 345(3), 998–1003.

Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett, D.E., Jr, Hieter, P., Vogelstein, B. and Kinzler, K.W. (1997). Characterization of the yeast transcriptome. *Cell* 88(2), 243–251.

Watkins, H., Rosenzweig, A., Hwang, D.S., Levi, T., McKenna, W., Seidman, C.E. and Seidman, J.G. (1992). Characteristics and prognostic implications of myosin missense mutations in familial hypertrophic cardiomyopathy. *N Engl J Med* 326(17), 1108–1114.

Watkins, H., Conner, D., Thierfelder, L., Jarcho, J.A., MacRae, C., McKenna, W.J., Maron, B.J., Seidman, J.G. and Seidman, C.E. (1995). Mutations in the cardiac myosin binding protein-C gene on chromosome 11 cause familial hypertrophic cardiomyopathy. *Nat Genet* 11, 434–437.

Woo, A., Rakowski, H., Liew, J.C., Zhao, M.S., Liew, C.C., Parker, T.G., Zeller, M., Wigle, E.D. and Sole, M.J. (2003). Mutations of the beta myosin heavy chain gene in hypertrophic cardiomyopathy: Critical functional sites determine prognosis. *Heart (British Cardiac Society)* 89(10), 1179–1185.

CHAPTER

Genetics and Genomics of Arrhythmias

Jeffrey A. Towbin and Matteo Vatta

INTRODUCTION

Arrhythmias typically are thought to occur due to primary or secondary abnormalities in cardiac electrophysiology. These abnormalities can include primary alterations in myocardial conduction and repolarization or those occurring as a result of structural heart disease. A significant portion of the individuals found to have abnormalities of rhythm and conduction are now known to have a genetic basis, with familial inheritance notable. Familial inheritance of arrhythmias and conduction disorders indicates that genetic factors play an integral role in development of these abnormalities. Understanding the underlying genetic defects responsible for these disorders has indeed provided insights into the mechanisms leading to the clinical picture and promises to impact the therapeutic strategies used in the care of these patients. Over the course of the last 15 years, our understanding of the genetic abnormalities in a variety of cardiomyopathies, long QT syndrome (LQTS), Brugada syndrome, catecholaminergic polymorphic ventricular tachycardia (CPVT), and newer disorders has led to general concepts of cardiovascular disease. In this chapter, the clinical features and management of arrhythmia disorders and conduction disease will be discussed and the current understanding of the genetic and genomic abnormalities associated with these disorders will be reviewed.

SPECIFIC CARDIAC ARRHYTHMIAS

The entire cardiac electrical system can be affected by genetic abnormalities, leading to atrial and ventricular tachyarrhythmias, sinus node dysfunction (SND), or atrioventricular block (AV block). Primary and secondary arrhythmias (i.e., those associated with structural heart disease) are discussed below.

PRIMARY ABNORMALITIES IN CARDIAC RHYTHM: VENTRICULAR TACHYARRHYTHMIAS

Long QT Syndromes

The long QT syndromes (LQTS) are primary disorders of cardiac repolarization (Schwartz et al., 2000a) in which prolongation of the QT interval corrected for heart rate (QTc) is seen on the surface electrocardiogram (ECG) along with abnormalities of T-wave morphology and sinus bradycardia (Figure 29.1). Syncope, seizures, and sudden death are the clinical features that are commonly seen, occurring due to ventricular tachycardia (VT), especially polymorphic VT or *torsades de pointes* (Figure 29.2), which can degenerate into ventricular fibrillation (VF). *Torsades de pointes* is common in all forms of LQTS, and is defined as "turning of the points" describing the varying axis of the QRS complex during VT (Schwartz et al., 2000a). This arrhythmia is a subset of polymorphic VT, to be distinguished from monomorphic VT which has a different morphology and mechanism of development (El-Sherif et al., 1997; Viskin et al., 1996). Monomorphic VT usually results from reentrant mechanisms and typically can be induced by programmed electrical stimulation. On the other hand, polymorphic VT cannot be induced using programmed electrical stimulation, suggesting the mechanism is unlikely to be

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

primary reentry. *Torsades de pointes*, in fact, is thought to be initiated by abnormal automaticity and then maintained by reentrant mechanisms (Antzelevitch, 1999). Development of early afterdepolarizations (EADs) appear to be an important mechanism whereby drug-induced action potential prolongation initiates *torsades* (January and Riddle, 1989; Verduyn et al., 1997). Under normal conditions, the ventricle is activated from subendocardium to epicardium by impulses arising in the subendocardial Purkinje network. Mapping data in animal models support the idea that the initial beat in *torsades* arises in the subendocardium, consistent with a triggered beat arising from an EAD in the Purkinje system (El-Sherif et al., 1996). The conditions that evoke EADs markedly prolong repolarization in the mid-myocardium (M cell region) as well, resulting in a situation where propagation of EAD-related triggered beats from the subendocardium may be blocked in regions where M cell action potentials have become especially long, setting up intramural reentrant excitation with a circuit that varies from beat to beat (Antzelevitch et al., 1995). This perhaps accounts for the distinctive morphology of *torsades de pointes* (Asano et al., 1997; El-Sherif et al., 1997).

Recently, the concept has emerged that defects in currents important for repolarization prolong the action potential but are not directly arrhythmogenic. Rather, action potential duration creates a milieu in which genetically normal, drug-unmodified ion channels or other electrogenic phenomena further prolong repolarization and precipitate arrhythmias.

The LQTS have been classified into acquired and genetically inherited forms. Acquired long QT syndrome (aLQTS) is the most common form of LQTS, with drug-induced LQTS particularly common. Drugs implicated in aLQTS include antiarrhythmic agents such as quinidine or sotalol, tricyclic antidepressants, antibiotics (especially macrolide antibiotics such as erythromycin), antihistamines such as terfenidate, and inhalational anesthetics. Acquired LQTS has also been seen in association with metabolic derangements including hypokalemia, hypomagnesemia, and hypercalcemia. Additionally, aLQTS has been identified in patients with other cardiac diseases such as cardiomyopathies and myocardial ischemia, as well as under circumstances of intracranial disease (i.e., intracranial surgery, subarachnoid hemorrhage, and increased intracranial pressure).

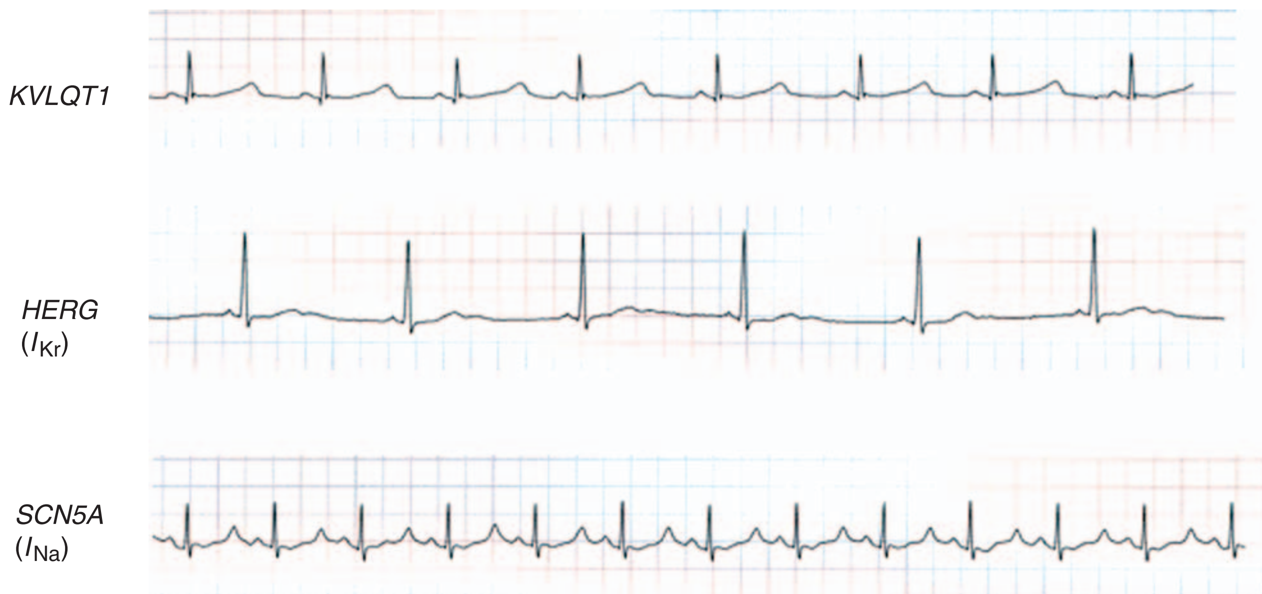


Figure 29.1 Electrocardiograms demonstrating LQTS associated with the three major genes causing LQTS (with permission).

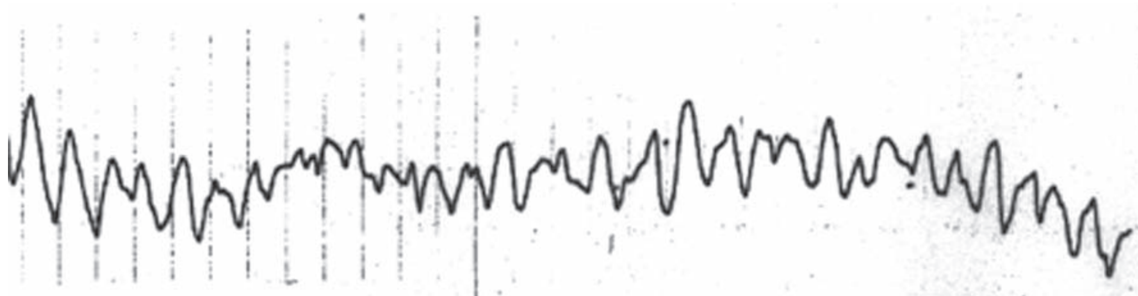


Figure 29.2 *Torsade de pointes* polymorphic ventricular tachycardia.

Four forms of inherited LQTS are known: Romano-Ward LQTS (Romano et al., 1963; Ward, 1964) and Jervell and Lange-Nielsen syndrome (Jervell and Lange-Nielsen, 1957) are the classically described forms of LQTS. More recently, two other disorders, Andersen syndrome (Andersen et al., 1971) and Timothy syndrome (Marks et al., 1995; Splawski et al., 2004) have been described, studied and their underlying causes determined. Another disorder, sudden infant death syndrome (SIDS) has also been found to be "LQTS-like" in some babies (Schwartz et al., 1998).

Table 29.1 Diagnostic criteria in LQTS

Clinical finding	Points
Electrocardiographic findings[a]	
QTc[b]	
>480 $ms^{1/2}$	3
460–470 $ms^{1/2}$	2
450 (male) $ms^{1/2}$	1
Torsades de pointes[c]	2
T-wave alterans	1
Notched T wave in three leads	1
Low heart rate for age[d]	0.5
Clinical history	
Syncope[c]	
With stress	2
Without stress	1
Congenital deafness	0.5
Family history[e]	
Family members with definite LQTS[f]	1
Unexplained sudden cardiac death below age 30 among immediate family members	0.5

Scoring: <1 point = low probability of LQTS; 2–3 points = intermediate probability of LQTS; >4 points = high probability of LQTS.

[a] In the absence of medications or disorders known to affect these ECG features.

[b] QTc calculated by Bazett's formula, where $QTc = QT/\sqrt{RR}$.

[c] Mutually exclusive.

[d] Resting heart rate below the second percentile for age.

[e] The same family member cannot be counted twice.

[f] Definite LQTS is defined by an LQTS score >4.

Romano-Ward syndrome is characterized by autosomal dominant inheritance with reduced penetrance and is the most common inherited form of LQTS, estimated to have an incidence of 1 in 10,000 live births worldwide (Priori et al., 1999a, b; Vatta et al., 2000). Jervell and Lange-Nielsen syndrome is a seemingly rare condition, with an estimated incidence of 1 in 1.6 million live births (Priori et al., 1999a, b; Vatta et al., 2000). This disorder, which is defined as LQTS associated with sensorineural deafness, has been described as having autosomal recessive inheritance since its initial description in 1957 (Jervell and Lange-Nielsen, 1957). In the past decade, molecular genetic studies, have clarified its inheritance, suggesting it to be autosomal dominant LQTS associated with autosomal recessive deafness. Finally, SIDS has been shown to be another presentation of inherited LQTS (Schwartz et al., 1998). Supportive data was provided by the identification of mutations in the cardiac sodium channel gene, *SCN5A*, by Ackerman et al. (2001), followed by identification of *HERG/KCNH2* mutations in other infants with SIDS (Christiansen et al., 2005).

Clinical Features of LQTS

LQTS is typically identified in individuals presenting with syncope, seizures or sudden cardiac death which results from episodic ventricular tachydysrhythmias, particularly *torsades de pointes* and VF (Priori et al., 1999a, b; Schwartz et al., 2000a; Vatta et al., 2000). Cardiac arrhythmias have been reported in up to 24% of cases, but the risk of sudden death has been estimated to be less than 1% per year. These estimates, however, are almost certainly inaccurate, since a significant percentage of the 300,000–400,000 sudden deaths occurring in the United States yearly (Myerburg, 1997) are likely to be the result of this disorder (which goes unrecognized) and many living patients are asymptomatic and can have normal QT intervals on screening ECGs due to reduced penetrance. Today, many asymptomatic family members are being identified via screening ECGs and molecular genetic family screening, and, therefore, better estimates are likely in the future.

Since LQTS may be difficult to diagnose, a set of criteria have been developed by Schwartz et al. (1985, 1993, 2000a). These criteria use a point system in the diagnostic scheme, relying heavily on classic features (Table 29.1). This approach is thought to improve diagnostic accuracy by including major criteria (prolonged QTc >440 ms; stress-induced syncope; family history of LQTS) and minor criteria (congenital deafness; T-wave alternans; relative bradycardia; abnormal ventricular repolarization) (Figure 29.3).

The clinical features of LQTS, which occur due to arrhythmias, are typically associated with triggering events. The most

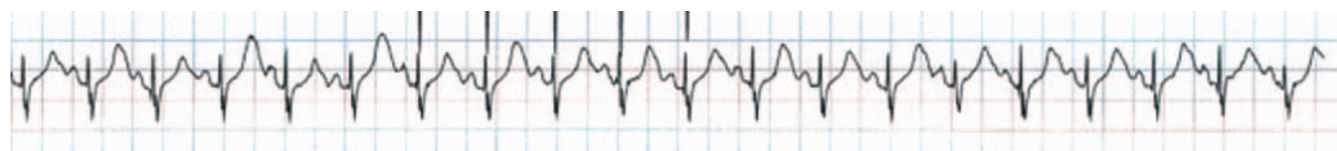


Figure 29.3 T-wave alternans in a child with LQTS.

well described triggers include exercise, anxiety or excitement, auditory events (i.e., telephone or alarm clock ringing), swimming or diving into a pool, and being postpartum. Some patients have events or die during sleep, which is thought to be associated with bradycardia.

In most cases of LQTS, no other abnormalities occur. However, sensorineural deafness is associated with the Jervell and Lange-Nielsen form of LQTS, distinguishing it from the Romano-Ward syndrome. LQTS has been reported to occur in approximately 1% of children with congenital deafness and this has led to the recommendation for screening ECGs in this group of children. A study from Thailand demonstrated a possible prevalence of 0.7% of Jervell and Lange-Nielsen in children with sensorineural deafness at a school for the deaf (Sopontammarak et al., 2003). These children also appear to have more severe ECG abnormalities than Romano-Ward syndrome patients and have a worse prognosis. Other associated abnormalities have rarely been reported in more complex patients with LQTS. Marks et al., (1995) described LQTS patients with syndactyly; in this patient subset, LQTS appears to be quite severe, with a high percentage dying in infancy. In most of these patients, the initial presentation included 2:1 AV block due to marked QT prolongation. These patients were initially found in sporadic cases, with both genders being represented, but more recently families have been identified and a new syndrome termed Timothy syndrome (Splawski et al., 2004), has been labeled (see section on "Timothy Syndrome"). Other patients with LQTS have been described with associated diabetes mellitus or asthma (Bellavere et al., 1988; Rosero et al., 1999), as well as those with potassium-sensitive periodic paralysis, dysmorphic features and skeletal abnormalities termed Andersen syndrome (Andersen et al., 1971).

Genetics of LQTS

As previously noted, four forms of inherited LQTS have been described, including autosomal dominant (Romano-Ward syndrome), autosomal recessive (Jervell and Lange-Nielsen syndrome), and other complex forms (Andersen syndrome and Timothy syndrome), in addition to sporadic cases. Over the past 15 years, the genetic aspects of all four forms of LQTS have been unraveled. In the case of the most common inherited form, Romano-Ward syndrome, the key genes have been identified for all of the mapped subtypes. In 1991, Keating and colleagues identified genetic linkage to the short arm of chromosome 11 (11p15.5) in several families with Romano-Ward syndrome (Keating et al., 1991a, b). Shortly thereafter, we demonstrated genetic heterogeneity and this was confirmed by several laboratories subsequently (Benhorin et al., 1993; Curran et al., 1993; Towbin et al., 1992, 1994). In a collaborative effort (Jiang et al., 1994), linkage was shown for several families to two new loci, the long arm of chromosome 7 (7q35-36) and the short arm of chromosome 3 (3p21). The three loci were later termed *LQT1* (11p15.5), *LQT2* (7q35-56), and *LQT3* (3p21). A fourth locus (*LQT4*) on chromosome 4q (4q25-27) was later described as well (Schott et al., 1995). Two other genes, both located on chromosome 21q22 (*LQT5*, *LQT6*), were later identified (Abbott et al., 1999; Barhanin et al., 1996; Sanguinetti et al., 1996a). More recently, other candidate genes and their loci were also identified (Table 29.2). Penetrance in Romano-Ward syndrome is reduced and in some families Romano-Ward syndrome appears to occur in a recessively inherited pattern (Priori et al., 1998, 1999c).

Gene Identification in Romano-Ward Syndrome: Ion Channel Gene Mutations

KVLQT1 or KCNQ1: The LQT1 Gene

The first of the genes mapped for LQTS, termed *LQT1*, required 5 years from the time that mapping to chromosome 11p15.5 was first reported to gene cloning (Table 29.2). This gene, originally named *KVLQT1*, but more recently called *KCNQ1*, is a potassium channel gene that consists of 16 exons, spans approximately 400 kb, and is widely expressed in human tissues including heart, inner ear, kidney, lung, placenta, and pancreas, but not in skeletal muscle, liver, or brain.

Analysis of the predicted amino acid sequence of KCNQ1 demonstrated that the gene encodes a potassium channel α-subunit (Barhanin et al., 1996; Sanguinetti et al., 1996a; Wang et al., 1996) with a conserved potassium-selective pore-signature sequence flanked by six membrane-spanning segments similar to shaker-type channels (Figure 29.4). A putative voltage sensor is found in the fourth membrane-spanning domains (S4) and the selective pore loop is between the fifth and sixth membrane-spanning domains (S5, S6). Biophysical characterization of the KCNQ1 protein confirmed that KCNQ1 is a voltage-gated potassium channel protein subunit that requires coassembly (Barhanin et al., 1996; Sanguinetti et al., 1996a) with a β-subunit called mink or KCNE1 to function properly (Figure 29.4). Expression of either KCNQ1 or KCNE1 alone results in inefficient (or no) current development. When both subunits are coexpressed in either mammalian cell lines or *Xenopus* oocytes, however, the slowly activating potassium current (I_{Ks}) is developed in cardiac myocytes. Combination of normal and mutant KVLQT1 subunits forms abnormal I_{Ks} channels, and these mutations are believed to act through a dominant-negative mechanism (the mutant form of KVLQT1 interferes with the function of the normal wild-type form through a "poison pill"-type mechanism) or a loss-of-function mechanism (only the mutant form loses activity) (Demolombe et al., 1998; Shalaby et al., 1997; Wollnick et al., 1997). In some cases, the protein is not normally trafficked to the membrane.

The vast majority of mutations in *KCNQ1* are heterozygous mutations in Romano-Ward syndrome patients (Chouabe et al., 1997, 2000; Duggal et al., 1998; Li et al., 1998; Splawski et al., 1997a; Towbin, 2006), and *KCNQ1* appears to be the most commonly mutated gene in LQTS.

HERG or KCNH2: The LQT2 Gene

After the initial localization of LQT2 to chromosome 7q35-36 by Jiang et al. (1994) (Table 29.2), candidate gene screening

Table 29.2 Long QT and related arrhythmia syndromes

Gene	Locus	Syndrome	Protein and subunit	Function and abnormality
KCNQ1	11p15.5	LQTS1	$K_V7.1\ \alpha$	$I_{Ks}\downarrow$KvLQT1
KCNH2	7q35	LQTS2	$K_V11.1\ \alpha$	$I_{Kr}\downarrow$ HERG
SCN5A	3p21	LQTS3	$Na_V1.5\ \alpha$	$I_{Na}\uparrow$
ANKβ	4q25	LQTS4	Ankyrin-B	$I_{Na,K}\downarrow$ $I_{NCX}\downarrow$
KCNE1	21q22.1	LQTS5	minK β	$I_{Ks}\downarrow$
KCNE2	21q22.1	LQTS6	MiRP1 β	$I_{Kr}\downarrow$
KCNJ2	17q23	LQTS7	Kir2.1 α	$I_{K1}\downarrow$
CACNA1C	12p13.3	LQTS8	$Ca_V1.2\ \alpha_{1c}$	$I_{Ca,L}\uparrow$
CAV3	3p25	LQTS9	Caveolin3	$I_{Na}\uparrow$
SCN4B	11q23	LQTS10	$Na_V1.5\ \beta4$	$I_{Na}\uparrow$
KCNQ1	11p15.5	JLNS1	$K_V7.1\ \alpha$	$I_{Ks}\downarrow$ KvLQT1
KCNE1	21q22.1	JLNS2	minK β	$I_{Ks}\downarrow$
SCN5A	3p21	LQTS3, SIDS1	$Na_V1.5\ \alpha$	$I_{Na}\uparrow$
KCNQ1	11p15.5	LQTS1, SIDS2	$K_V7.1\ \alpha$	$I_{Ks}\downarrow\uparrow$ KvLQT1
KCNH2	7q35	LQTS2, SIDS3	$K_V11.1\ \alpha$	$I_{Kr}\downarrow$ HERG
CAV3	3p25	LQTS9, SIDS4	Caveolin3	$I_{Na}\uparrow$
KCNE2	21q22.1	LQTS2, SIDS5	MiRP1 β	$I_{Kr}\downarrow$

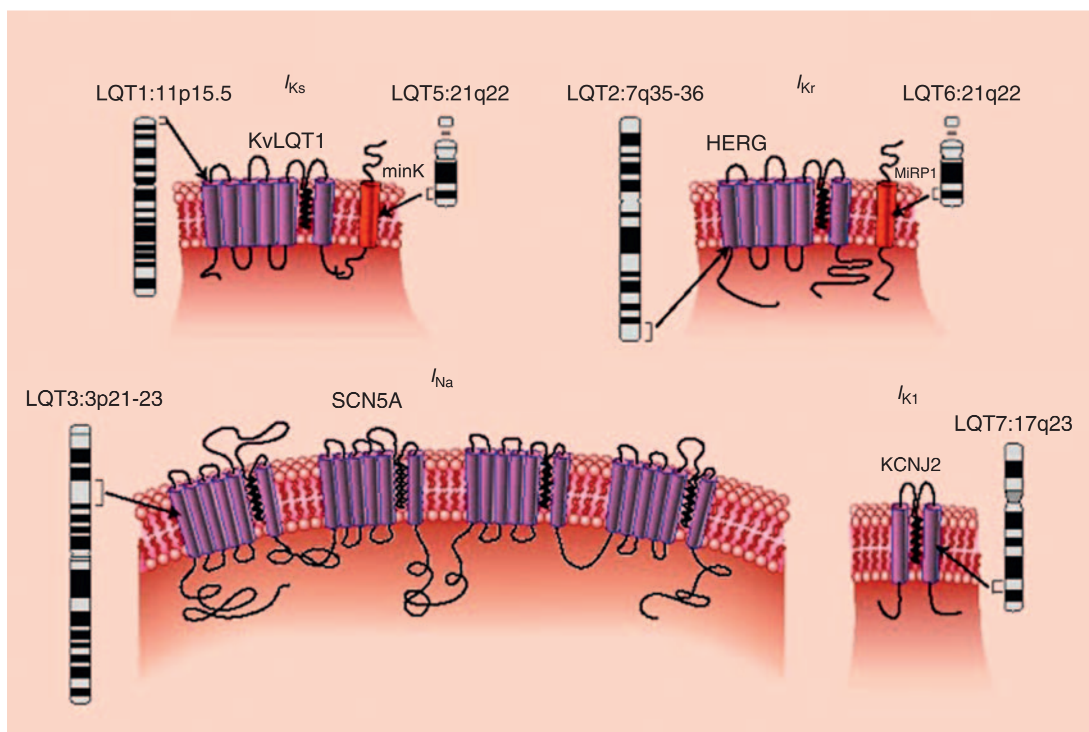


Figure 29.4 Chromosomal loci and genes responsible for LQTS.

identified mutations in *HERG* (human ether-a-go-go-related gene), a cardiac potassium channel gene originally cloned from a brain cDNA library (Warmke and Ganetzky, 1994) and that is expressed in neural crest-derived neurons (Arcangeli et al., 1997), microglia (Pennefather et al., 1998), a wide variety of tumor cell lines (Bianchi et al., 1998), and the heart (Curran et al., 1995).

The *KCNH2* gene consists of 16 exons and spans 55 kb of genomic sequence (Curran et al., 1995). The predicted topology of HERG is shown in Figure 29.4 and is similar to KCNQ1 but, unlike KCNQ1, KCNH2 has extensive intracellular amino- and carboxyl-termini, with a region in the carboxyl-terminal domain having sequence similarity to nucleotide binding domains (NBDs).

Electrophysiologic and biophysical characterization of KCNH2 expressed in *Xenopus* oocytes established that this protein encodes the rapidly activating delayed-rectifier potassium current I_{Kr} (Sanguinetti et al., 1995; Trudeau et al., 1995) and electrophysiologic studies of LQTS-associated mutations demonstrated a loss-of-function or a dominant-negative mechanism (Sanguinetti et al., 1996b) of action. In addition, protein trafficking abnormalities have been shown to occur (Furutani et al., 1999; Zhou et al., 1998). More recently, Zhang et al. (2004) identified an intronic variant in *KCNH2* (T1945 + 6C) in which splicing assay analysis showing downstream intron retention and complementary DNA with the retained intron 7 failed to produce functional channels, consistent with potential disease-causing dysfunction. This finding potentially expands the disease-causing mechanisms in LQTS. This channel has been shown to coassemble with β-subunits for normal function, similar to that seen in I_{Ks}. Abbott et al. (1999) identified MiRP1 (KCNE2) as a β-subunit for KCNH2 (see section on "*MiRP1*: The *LQT6* Gene"). The identification of these initial genes which encoded ion channels, suggested that the "final common pathway" of LQTS is ion channel disruption (Towbin, 2000).

SCN5A: The LQT3 Gene

Utilization of the candidate gene approach established that the gene responsible for chromosome 3-linked LQTS (*LQT3*) (George et al., 1995) is the cardiac sodium channel gene *SCN5A* (Wang et al., 1995a, b) (Table 29.2). SCN5A is highly expressed in human myocardium and brain, but not in skeletal muscle, liver, or uterus (Hartmann et al., 1999; Wang et al., 1995a). It consists of 28 exons that span 80 kb and encodes a protein of 2016 amino acids with a putative structure that consists of four homologous domains (DI to DIV), each of which contains six membrane-spanning segments (S1 to S6) similar to the structure of the potassium channel α-subunits (Figure 29.4) (Gellens et al., 1992). Mutation analysis identified three mutations initially (Wang et al 1995a, b); and when expressed in *Xenopus* oocytes, all mutations generated a late phase of inactivation-resistant, mexiletine- and tetrodotoxin-sensitive whole-cell current via different mechanisms (Bennett et al., 1995; Dumaine et al., 1996). Two of the three mutations showed dispersed reopening after the initial transient, but the other mutation showed both dispersed reopening and long-lasting bursts. These results suggested that SCN5A mutations act through a gain-of-function mechanism (the mutant channel functions normally, but with altered properties such as delayed inactivation) and that the mechanism of chromosome 3-linked LQTS is persistent nonactivated sodium current in the plateau phase of the action potential. An et al. (1998) also showed that not all mutations in SCN5A are associated with persistent current and demonstrated that SCN5A interacted with β-subunits. Other mutations were later identified (Benhorin et al., 1998).

MinK or KCNE1: The LQT5 Gene

The *minK* (IsK or *KCNE1*) gene was initially localized to chromosome 21 (21q22.1) and found to consist of three exons that span approximately 40 kb (Honore et al., 1991). It encodes a short protein consisting of 130 amino acids and has only one transmembrane-spanning segment with small extracellular and intercellular regions (Figure 29.4). When expressed in *Xenopus* oocytes, it produces potassium current that closely resembles the slowly activating delayed-rectifier potassium current I_{Ks} in cardiac cells (Barhanin et al., 1996; Sanguinetti et al., 1996a), but requires coexpression cardiac slowly activating delayed-rectifier I_{Ks} current to develop (Barhanin et al., 1996; Sanguinetti et al., 1996a). Bianchi et al. (1999) also showed that mutant KCNE1 results in abnormalities of I_{Ks} and I_{Kr} and in protein trafficking abnormalities. McDonald et al. (1997) showed that KCNE1 also interacts with KCNH2, regulating I_{Kr}. Splawski et al. (1997b) demonstrated that *KCNE1* mutations cause LQT5 when they identified mutations in two families with LQTS (Table 29.2). In both cases, mutations were identified that reduced I_{Ks} by shifting the voltage dependence of activation and accelerating channel deactivation. This was later confirmed by others (Chouabe et al., 2000; Duggal et al., 1998) and further supported by the fact that a mouse model with mutant *minK* (Vetter et al., 1996) developed a phenotype (that included deafness). The functional consequences of these mutations include delayed cardiac repolarization and hence, an increased risk of arrhythmias (Chouabe et al., 2000; Duggal et al., 1998; Franqueza et al., 1999; Vatta and Towbin, 2006).

MiRP1 or KCNE2: The LQT6 Gene

The *MiRP1* gene (the *minK-related peptide 1* or *KCNE2* gene) is a potassium channel gene encoding a small integral membrane subunit protein that assembles with KCNH2 (LQT2) to alter its function and enable full development of the I_{Kr} current (Figure 29.4). This is 123-amino acid channel protein has a single predicted transmembrane segment similar to that described for KCNE1. Chromosomal localization studies mapped the *KCNE2* gene to chromosome 21q22.1, within 79 kb of *KCNE1* (*minK*) and arrayed in opposite orientation. The open reading frames of these two genes share 34% identity and both are contained in a single exon, suggesting that they are related through gene duplication and divergent evolution (Abbott et al., 1999).

Three missense mutations associated with dysrhythmias were initially identified (Abbott et al., 1999) and biophysical analysis demonstrated that these mutants form channels

that open slowly and close rapidly, thus diminishing potassium currents. Therefore, like KCNE1, this channel protein acts as a β-subunit but, by itself, leads to risk of ventricular arrhythmia when mutated (Table 29.2).

Gene Identification in Romano-Ward Syndrome: Non-Ion Channel-Encoding Genes

Ankyrin-β or ANKβ: The LQT4 Gene

Initially mapped in 1995 to chromosome 4q25-25 by Schott and colleagues in a large French family with autosomal dominant LQTS associated with sinus bradycardia due to SND and atrial fibrillation (AF), the gene remained elusive until 2003. Mohler et al. (2003) identified *ankyrin-β* or *ankyrin-2* (*ANKβ*) as the disease-causing gene (Table 29.2). This gene encodes a protein with three major isoforms with molecular weights ranging from 440, 220, and 150 kDa generated by alternative splicing. The major isoform in the heart is the 220 kDa form. Ankyrins are adapter proteins that link integral membrane proteins to the spectrin-based cytoskeleton (Bennett and Baines, 2001) and contain three functional domains that consist of the membrane binding domain, the spectrin binding domain, and the regulatory domain. These proteins bind to several ion channel proteins, including the voltage-sensitive sodium channel (I_{Na}), anion exchanger (Cl^-/HCD_3 exchanger), sodium–potassium ATPase, sodium–calcium exchanger (NCX or $I_{Na—Ca}$), and calcium release channels including those mediated by the ryanodine receptor (RyR2) and inositol triphosphate (IP_3) receptor. Mutations in *ANKβ* appear to act by a loss-of-function mechanism. It is likely that mutations in this gene cause clinical phenotype by disrupting the cytoskeletal framework which results in compromise of channel function or of the trafficking of channels to their proper locale. A murine model of mutant *ANKβ* results in clinical similarities to that seen in humans and findings indicative of calcium handling abnormalities (Mohler et al., 2003) and sodium channel dysfunction (Chauhan et al., 2000) have been reported.

Caveolin-3 or CAV3: The LQT9 Gene

Caveolae are 50–100-nm omega-shaped microdomains of the plasmalemma which are particularly abundant in the cardiovascular system, including cardiomyocytes, endothelial cells, smooth muscle cells, macrophages, and fibroblasts. These proteins are involved in a variety of functions including vesicular trafficking and serve as platforms to organize and regulate a variety of signal transduction pathways. Caveolins, the principal proteins in caveolae, are found as three different isoforms (CAV1–CAV3) that are encoded by separate genes. Although CAV1 and CAV2 are expressed in most cell types, only CAV3 is specifically expressed in cardiomyocytes and skeletal muscle. Some cardiac ion channels have been specifically localized to caveolae, including the SCN5A-encoded voltage-gated sodium channel ($Na_v1.5$), the voltage-dependent potassium channel ($K_v1.5$), the sodium-calcium exchanger, and the L-type calcium channel. In the heart, a variety of other signaling molecules have been found

TABLE 29.3 Brugada and related arrhythmia syndromes

Gene	Locus	Syndrome	Protein and subunit	Function and abnormality
SCN5A	3p21	BrS1, CoD	$Na_V1.5$ α	I_{Na}↓
GPD1L	3p24	BrS2	G3PD1L	I_{Na}↓
SCN5A	3p21	SUND	$Na_V1.5$ α	I_{Na}↓
SCN5A	3p21	Progressive CoD	$Na_V1.5$ α	I_{Na}↓
SCN5A	3p21	BrS1, CoD, AA (SSS)	$Na_V1.5$ α	I_{Na}↓
SCN5A	3p21	BrS1, LQTS3	$Na_V1.5$ α	I_{Na}↓
SCN5A	3p21	BrS1, LQTS3, CoD	$Na_V1.5$ α	I_{Na}↓
SCN5A	3p21	iVF, CoD	$Na_V1.5$ α	I_{Na}↓
SCN5A	3p21	DCM, CoD, AA (AF)	$Na_V1.5$ α	I_{Na}↓
SCN5A	3p21	BrS1, SIDS1	$Na_V1.5$ α	I_{Na}↓
SCN5A	3p21	BrS1, CoD, SIDS1	$Na_V1.5$ α	I_{Na}↓

in caveolae, including the β2-adrenergic receptor. Vatta et al. (2006) identified multiple mutations in familial and sporadic cases of LQTS (Table 29.2) and electrophysiological analysis of these mutations demonstrated a two- to three-fold increase in the late sodium current compared with wild-type CAV3. In addition, CAV3 was shown to directly interact with SCN5A, the cardiac sodium channel. This gain-of-function of the cardiac sodium channel induced by mutations in CAV3, is similar to the SCN5A mechanism of disease.

Caveolin-3 is the second non-ion channel protein (after ankyrin-B) implicated in the pathogenesis of congenital LQTS. Consistent with the "final common pathway" hypothesis (Bowles et al., 2000; Towbin, 1998, 2000), genes encoding cardiac channel-interacting proteins, which secondarily disrupt ion channel function, may confer genetic susceptibility for LQTS. Caveolin-3 represents one of a potentially large group of such proteins. In addition, $Na_V1.5$ appears to be disrupted in a variety of arrhythmia-associated disorders, including LQTS and Brugada syndrome (Table 29.3) both possible substrates of SIDS (Ackerman et al., 2001; Towbin and Ackerman, 2001) and AV block (Benson et al., 2003).

SCN4B: The LQT10 Gene

Sodium channel β-subunits are critical regulatory proteins, and four of these β-subunits have been described, β1–β4. These regulatory proteins are encoded by the sodium channel β-subunit gene family known as $Na_{v\beta}$ and the four protein subunits include SCN1B–SCN4B. These β-subunits are all detectable in cardiac tissue and play a crucial role in cell adhesion, signal transduction, channel expression at the sarcolemma, and voltage dependence

of channel gating. These proteins are characterized by an extracellular N-terminal cleaved region, a transmembrane segment, and a cytoplasmic domain with a C-terminal tail. Studies have shown that $Na_{v\beta}$ subunits contain an extracellular Ig-like fold, often found in cell adhesion molecules that target ion channels to the sarcolemma and mediate interactions with signaling molecules. $Na_{v\beta}1$ and $Na_{v\beta}3$ are similar in sequence and associate noncovalently with β-subunits while $Na_{v\beta}2$ and $Na_{v\beta}4$ are related proteins that are disulfide-linked to β-subunits. Immunohistochemical studies in murine hearts indicate that the primary cardiac sodium channel in ventricular myocytes is composed of $Na_v1.5$ (SCN5A) plus β2 and/or β4 subunits. The β4-subunit cytoplasmic tail is thought to interact with the S6-binding site within the inner cavity of $Na_v1.5$. Recently, Madeiros-Domingo et al. (2007) have identified a mutation in the gene encoding the β4-subunit (SCN4B) in a multigenerational family with severe LQTS in several family members (Table 29.2), as well as with reduced penetrance in others. In addition, AV block, sudden cardiac death, and bradycardia were clinically apparent in this family. The authors identified a missense mutation in the family (L179F) which occurred in the transmembrane-spanning region of the protein. Biophysical analysis of this mutation expressed with the SCN5A β-subunit demonstrated an eightfold increase in late sodium current compared with SCN5A alone and threefold increase compared with SCN5A + WT-β4. This is consistent with the LQT3 biophysical phenotype and is similar to that seen in CAV3-related LQTS. Hence, this further confirms the "final common pathway hypothesis" that arrhythmias and conduction system disease occur as a consequence of ion channel dysfunction occurring either directly through ion channel mutations or secondarily due to dysfunction resulting from abnormal binding partner function or other indirect channel abnormality (such as, occurs with drugs) (Bowles et al., 2000; Towbin, 2000).

COMPLEX FORMS OF LQTS

Jervell and Lange-Nielsen Syndrome

Clinical Features

As noted previously, patients with Jervell and Lange-Nielsen syndrome (JLNS) have severe QT interval prolongation, episodic tachydysrhythmias including *torsades de pointes*, syncope and/or sudden death, and sensorineural deafness (Jervell and Lange-Nielsen, 1957; Schwartz et al., 2000). The deafness is autosomal recessive and severe while the LQTS is autosomal dominant (Vatta et al., 2000).

Genetics

Neyroud et al. (1997) reported the first molecular abnormality in patients with JLNS when they reported on two families in which three children were affected by JLNS, finding a novel homozygous deletion–insertion mutation of *KCNQ1* (Table 29.2). A deletion of 7 bp and an insertion of 8 bp at the same location led to premature termination at the C-terminal end of the KVLQT1 channel. At the same time, Splawski et al. (1997a) identified a homozygous insertion of a single nucleotide that caused a frameshift in the coding sequence after the second putative transmembrane domain (S2) of KCNQ1. Together, these data strongly suggested that at least one form of JLNS is caused by homozygous mutations in *KCNQ1*, which was confirmed by others (Chen et al., 1999; Chouabe et al., 1997; Schulze-Bahr et al., 1997; Tyson et al., 1997; Wollnick et al., 1997).

As a general rule, heterozygous mutations in *KCNQ1* cause Romano-Ward syndrome (LQTS only), whereas homozygous (or compound heterozygous) mutations in *KCNQ1* cause JLNS (LQTS and deafness). The hypothetical explanation suggests that although heterozygous KCNQ1 mutations act by a dominant-negative mechanism (Mohammad-Pannah et al., 1999), some functional KCNQ1 potassium channels still exist in the stria vascularis of the inner ear. Therefore, congenital deafness is averted in patients with heterozygous *KCNQ1* mutations. For patients with homozygous mutations, no functional KCNQ1 potassium channels can be formed. It was shown by *in situ* hybridization that KCNQ1 is expressed in the inner ear (Neyroud et al., 1997), suggesting that homozygous *KCNQ1* mutations can cause the dysfunction of potassium secretion in the inner ear and lead to deafness (Vetter et al., 1996). However, it should be noted that incomplete penetrance exists and not all heterozygous or homozygous mutations follow this rule (Priori et al., 1999c).

As with Romano-Ward syndrome, if *KCNQ1* mutations can cause the phenotype, it could be expected that *KCNE1* mutations could also be causative of the phenotype (JLNS). Schulze-Bahr et al. (1997), in fact, showed that mutations in *KCNE1* result in JLNS syndrome as well (Table 29.2), and this was confirmed subsequently (Duggal et al., 1998; Tyson et al., 1997). Hence, abnormal I_{Ks} current, whether it occurs due to homozygous or compound heterozygous mutations in *KCNQ1* or *KCNE1*, results in LQTS and deafness.

Andersen-Tawil Syndrome (LQT7)

Clinical Aspects

Andersen et al. (1971) identified a complex phenotype including ventricular arrhythmias, potassium-sensitive periodic paralysis, and dysmorphic features. The dysmorphisms included hypertelorism, broad nasal root, defects of the soft and hard palate, as well as short stature. More recently, skeletal abnormalities have broadened the phenotype (Andelfinger et al., 2002). These skeletal features include micrognathia, clinodactyly, syndactyly, and scoliosis. The associated cardiac abnormalities include QTc prolongation, VT, VF, and atrial arrhythmias. *Torsades de pointes* and bidirectional VT have been seen. In addition, repolarization abnormalities affecting late repolarization and resembling giant U waves are common. Sudden death has not been reported as a major risk in this disorder. Andelfinger et al. (2002) also reported sex-specific variable expression, as well as other clinical features including unilateral dysplastic kidney and congenital heart disease (bicusopid aortic valve, coarctation of the aorta, valvular pulmonic stenosis).

TABLE 29.4 Ca^{2+} dependent arrhythmia syndromes

Gene	Locus	Syndrome	Protein and subunit	Function and abnormality
CACNA1CA	12p13.3	TS1, ASD	$Ca_V1.2\ \alpha_{1C}$	$I_{Ca,L}\uparrow$
CACNA1C	12p13.3	TS2, ASD	$Ca_V1.2\ \alpha_{1C}$	$I_{Ca,L}\uparrow$
RyR2	1q42	CPVT1	RyR2 α	SR Ca^{2+}leak ↑
CASQ2	1p13.3	CPVT2	Calsequestrin	SR Ca^{2+} leak ↑
ANKβ	4q25	CPVT3	Ankyrin-B	SR Ca^{2+} leak ↑
KCNJ2	17q23	CPVT4, ATS1	Kir2.1 α	$I_{K1}\uparrow$
RyR2	1q42-q43	CPVT1, LQTS	RyR2 α	SR Ca^{2+} leak ↑
RyR2	1q42-q43	CPVT1, ARVC2	RyR2 α	SR Ca^{2+} leak ↑

Genetic Aspects

Kir 2.1 or KCNJ2: The LQT7 Gene

Andersen-Tawil syndrome was originally mapped to chromosome 17q23-q24.2 by Plaster et al. (2001) using genome-wide linkage analysis. Candidate gene mutation screening identified mutations in *KCNJ2* (Table 29.2), which encodes an inward rectifier potassium channel called Kir2.1 (Tristani-Firouzi et al., 2002). This channel is highly expressed in the heart and plays a role in phase 4 repolarization and in the resting membrane potential. Multiple gene mutations have been identified to date with relatively high penetrance noted. Functional studies have demonstrated reduction or suppression of I_{K1} by a haploinsufficiency or dominant-negative effect (Tristani-Firouzi et al., 2002; Lange et al., 2003). Lange et al. (2003) generated known *KCNJ2* mutants which did not yield any measurable potassium currents in CHO cells consistent with failure to form functional homomultimeric complexes and non-functional channels. In addition, Bendahhou et al. (2003) demonstrated that defective Kir 2.1 channels may not traffic to the membrane properly. This gene may play a role in developmental signaling pathways as well, which is believed to be the cause of the dysmorphisms (Andelfinger et al., 2002).

Analysis of a variety of *KCNJ2* mutations demonstrated that many of these mutations included residues implicated in binding membrane-associated phoephatidylinositol 4,5-biphosphate (PIP2) (Donaldson et al., 2003). It should be noted that nearly 40% of cases do not segregate with this gene, suggesting that genetic heterogeneity exists (Donaldson et al., 2004).

Timothy Syndrome

Clinical

This complex disorder is characterized by multisystem dysfunction including webbing of fingers and toes (syndactyly), immune deficiency, intermittent hypoglycemia, a cognitive immune deficiency, intermittent hypoglycemia, cognitive abnormalities, autism, dysmorphic features, various forms of congenital heart disease, and lethal arrhythmias associated with LQTS (Marks et al., 1995; Splawski et al., 2004).

Genetics

This autosomal dominant disorder was described initially by Marks et al. (1995), and more recently the genetic basis of this syndrome was described by Splawski et al. (2004). Mutations in Ca(V)1.2, the L-type calcium channel, which is expressed in all affected tissues, was shown to produce maintained inward Ca^{2+} overload in multiple cell types and, in the heart, causes prolonged Ca^{2+} current delaying cardiomyocyte repolarization and increased arrhythmia risk (Table 29.4).

Genotype–Phenotype Correlations in LQTS

Moss et al. (1995) showed that the ECG manifestations of LQTS were in great part determined by which channel is mutated. Different T-wave patterns were clearly evident when comparing tracings from patients with mutations in LQT1, LQT2, and LQT3. More recently, Zareba et al. (1998) showed that the mutated gene may result in a specific clinical phenotype with different triggers and may predict outcome. For instance, these authors suggested that mutations in LQT1 and LQT2 result in early symptoms (i.e., syncope) but the risk of sudden death was relatively low for any event. In contrast, mutations in LQT3 resulted in a paucity of symptoms but when symptoms occurred they were associated with a high likelihood of sudden death. Zareba et al. (2003a) showed that the location of mutations in *KCNQ1* plays no role in clinical course. However, Shimizu et al. (2004) showed mutation site-specific differences in arrhythmic risk and sensitivity to sympathetic stimulation in LQT1 patients with those having transmembrane mutations having a greater risk of early-onset cardiac events and greater sensitivity to sympathetic stimulation compared with patients having C-terminal mutations. Moss and colleagues showed that pore mutations in LQT2 have higher risk effects of age and gender based on genotype were also reported (Zareba et al., 2003b). In this case, Zareba and colleagues showed that during childhood, the risk of cardiac events was significantly higher in LQT1 and LQT3 males than females with no gender-related differences in the risk of cardiac events among LQT2 and LQT3 carriers. In adulthood, LQT2 and LQT1 females had a significantly higher

risk of cardiac events than respective males. The lethality of cardiac events was highest in LQT3 males and females, and higher in LQT1 and LQT2 males than females. Compound mutations, which reportedly occurs in ~8% of subjects, appears to cause longer QTc intervals, worse and more frequent cardiac events and, in general, a more severe phenotype than other mutations and worse disease than expected (Westenskow et al., 2004).

Neonates appear to have gene-specific clinical features as well. Lupoglazoff et al. (2004) studied 23 neonate probands and demonstrated that 2:1 AV block is common and usually associated with LQT2 mutations, with a poor prognosis during the first month of life. In contrast, LQT1 mutations in this patient subgroup appeared correlated with sinus bradycardia and good short-term prognosis with β-blocker therapy. Ackerman et al. (1999) and Moss et al. (1999) showed that swimming is a common trigger for symptoms in LQT1 patients while Wilde et al. (1999) have found auditory triggers to be common in LQT2. LQT3, on the other hand, appears to be associated with sleep-associated symptoms. More recently, Choi et al. (2004) evaluated 388 subjects with LQTS for LQT gene mutations and identified 43 individuals with swimming-related events (11%). Among this group, 65% had mutations in LQT1, with 21% having ryanodine receptor mutations (RyR2), and 5% with LQT2 mutations. Coupled with the findings by Moss and colleagues, it could be suggested that understanding the underlying cause of LQTS in any individual could be used to improve survival by prevention and gene-specific therapy.

Management of LQTS

At present, there are four classical modalities for treatment of LQTS that have withstood the test of time: (1) β-blockers (Moss, 1998, Moss et al., 2000; Priori et al., 2004; Villain et al., 2004;), (2) pacemakers (Eldar et al., 1987; Moss et al., 1991, Viskin, 2000), (3) left cervicothoracic sympathetic ganglionectomy (Moss and McDonald, 1970; Schwartz et al., 1991, 2004), and, most recently, (4) internal cardioverter defibrillator (ICD) (Ten Harkel et al., 2005; Zareba et al., 2003c). The mortality of untreated symptomatic patients with LQTS exceeds 20% in the year after their first syncopal episode and approaches 50% within 10 years of initial presentation (Schwartz, 1985). With institution of the classical therapy, this can be reduced to 2–4% in 5 years after initial presentation (Moss et al., 2000; Schwartz, 1985). Despite the lack of a placebo-controlled, randomized clinical trial, strong evidence supports the use of antiadrenergic interventions as the mainstay of therapy for most patients. The trigger for many life-threatening events appears to involve sudden increases in sympathetic activity (i.e., emotional or physical stress), and therefore, antiadrenergic therapy makes physiological sense. The β-adrenergic blocking agents prevent new syncopal episodes in approximately 65–75% of patients (Moss et al., 2000). Suppression of complex ventricular arrhythmias (i.e., couplets and VT) seems desirable. Villain et al. (2004) showed that a low incidence of cardiac events occur in children with LQTS treated with β-blockers. In the 122 children studied, only 4 deaths occurred. Of these, 111 children (92%) were treated with β-blockers alone, with no deaths and only 5 non-fatal cardiac events (4.5%) noted. None of these children had LQT1 mutations. Priori et al. (2004) studied 335 patients treated with β-blockers for an average of 5 years and found 10% of LQT1 patients, 23% of LQT2 patients, and 32% of LQT3 patients had events while being treated with the highest risk of events in LQT2 and LQT3 patients. Chatrath et al. (2004) found that 25% of LQTS probands had cardiac events, with LQT1 most prevalent. In addition, they found that the highest rate of events occurred during treatment with atenolol while propranolol appeared more protective. The addition of an IB agent (mexiletine) to β-blocker therapy may be helpful, particularly in patients with the LQT3 genotype (Schwartz et al., 1995). High-risk patients with drug resistant, symptomatic VT are referred for left cardiac sympathetic denervation which apparently provides additional protection (Moss and McDonald, 1970; Schwartz et al. 1991, 2004). In addition, the use of cardiac pacing as an adjunct to β-blockers appears to be most rational in patients with evidence of pause-dependent or bradycardia-dependent arrhythmias (Eldar et al., 1987; Moss et al., 1991; Viskin, 2000). Symptomatic bradycardia due to LQTS or induced by β-blocker therapy, should also be considered an indication for elective pacing. Beta-blocker therapy is monitored with treadmill exercise testing with the desired result a blunting of the heart rate response to exercise. However, Kaltman et al. (2003) suggested that little difference occurs between pretreated and treated patients regarding QTc at any phase, QTc dispersion or other QTc measures. Unfortunately, in some patients, this comes at the expense of excessive sinus bradycardia at rest and with minimal levels of exertion. Excessive fatigue, inattentiveness and irritability may result in discontinuance of therapy by the patient. Compliance, especially in the adolescent population, may be enhanced by returning the patient to a relatively normal lifestyle by the elimination of chronotropic incompetence with a pacemaker.

Other less time-tested therapies are also available. In some rare cases, *torsades de pointes* persists despite therapy with the classical modalities. The ICD has been used successfully in this setting (Groh et al., 1996; Platia et al., 1985). Up until recently, it had not been considered to be first-line therapy because shocks from the device can precipitate further emotional stress and set off a circuitous response of persistent malignant arrhythmias. However, the Multicenter Automatic Defibrillator Implantation Trial (MADIT), which demonstrated dramatic superiority of therapy with automatic implantable defibrillators over "best conventional therapy" in patients with coronary disease at high risk for ventricular arrhythmias (Moss et al., 1996), has made this therapeutic approach somewhat appealing. Automatic implantable defibrillators have been used more commonly in LQTS patients due to the results of the MADIT trial. However, whether this is the best approach is still not clear as long-term data is needed to determine the answer to this question. Zareba et al. (2003c) has shown 3-year follow-up analysis with 1.3% death in ICD patients versus 16% non-ICD patients. Reports of efficacy in neonates have been published as well (Ten Harkel et al., 2005).

Another new approach to treating patients with LQTS is the so-called "gene-specific" approach. With the identification of the precise molecular defect in some patients with LQTS, specific mechanism-based therapies have been devised and small therapeutic trials performed. Schwartz et al. (1995) were the first to use this approach when they used the sodium channel blocker mexiletine in patients with mutations in the sodium channel gene *SCN5A* (*LQT3*). In these patients the QTc was dramatically shortened in a statistically significant manner. Patients with potassium channel mutations (*HERG*, *LQT2*) treated with mexiletine had no change in the QTc. However, no data currently exists which demonstrates clinical efficacy of this approach in either decreasing the number of syncopal events or improving survival. Other sodium channel blockers have had similar effects (Rosero et al., 1999). Recently, Benhorin et al. (2000) showed that flecainide shortened the QTc in patients with a D1790G *SCN5A* mutation while lidocaine was ineffective suggesting that allele-specific therapies may be needed. Although the data is intriguing, use of sodium channel blockers alone for patients with *SCN5A* mutations should still be considered experimental. Other gene-specific trials have also been performed with similar results. Compton et al. (1996) used intravenous potassium to elevate the serum potassium to >4.8 in patients with *LQT2* and found significant shortening of the QTc. Again, no data on survival or symptom improvement exists for this therapy, but this has been confirmed (Choy et al., 1997). Etheridge et al. (2003) showed that QTc reduction may be large, QT dispersion and T-wave morphology improved, but outcome data was not available. Potassium channel openers may have the same effect (Shimizu et al., 1998). Other gene-specific therapeutic approaches are currently being developed.

Brugada Syndrome

Clinical Features

This disorder is characterized by ST-segment elevation in the right precordial leads (V1–V3) with or without right bundle branch block (RBBB) (Figure 29.5) and episodic VF (Brugada et al., 2005). The first identification of the ECG pattern of RBBB with ST-elevation in leads V1–V3 was reported by Osher and Wolff (1953). Shortly thereafter, Edeiken (1954) identified persistent ST-elevation without RBBB in 10 asymptomatic males and Levine et al. (1956) described ST-elevation in the right chest leads and conduction block in the right ventricle in patients with severe hyperkalemia. The first association of this ECG pattern with sudden death was described by Martini et al. (1989) and later by Aihara et al. (1990) and further confirmed in 1991 by Pedro and Josep Brugada (1991) who described four patients with sudden death and aborted sudden death with ECGs demonstrating RBBB and persistent ST-elevation in leads V1–V3. In 1992, these authors characterized what they believed to be a distinct clinical and ECG syndrome (Antzelevitch et al., 2003; Brugada et al., 1992, 1997). In some patients the surface ECG appears normal, even in familial cases. Provocation studies in the catheterization laboratory using ajmaline, flecainide, or procainamide

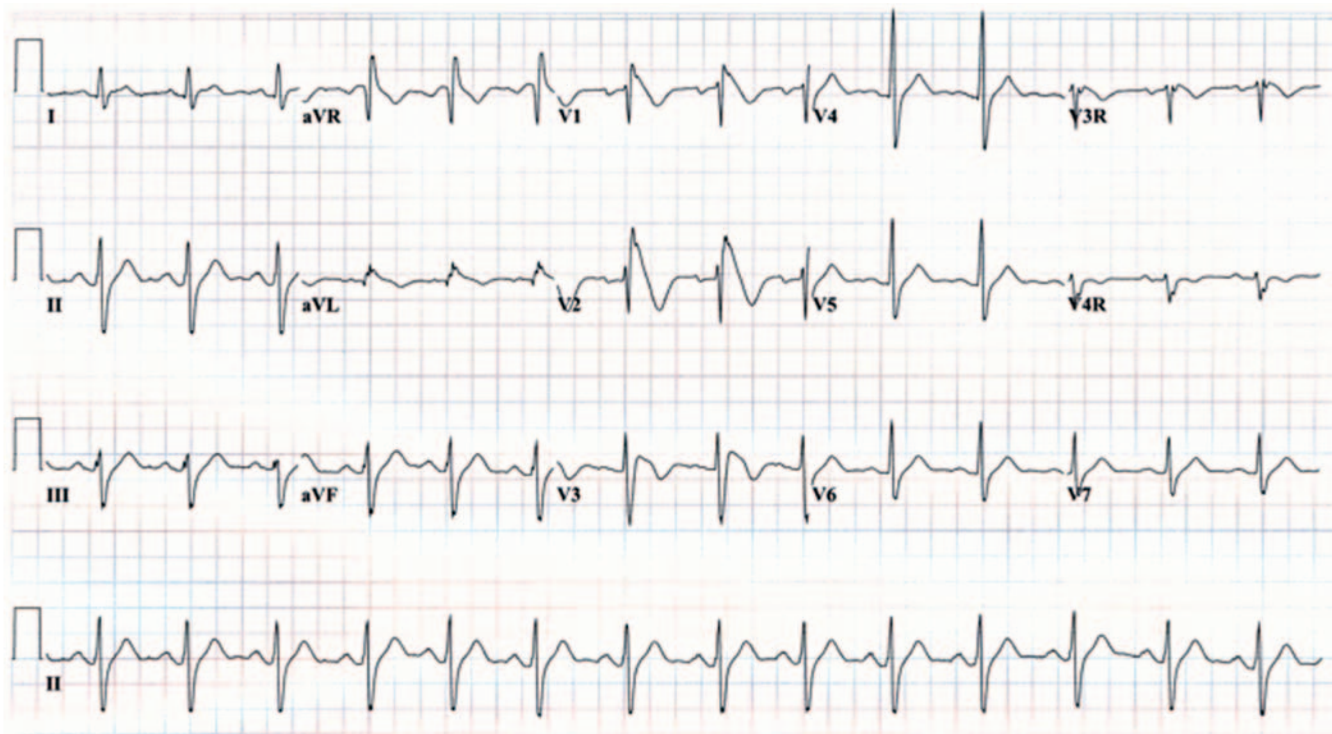


Figure 29.5 Brugada syndrome in a 4-year old. Note the right ventricular conduction delay and ST-segment elevation in leads V1, V2. Procainamide infusion worsened the ST-elevation.

will result in ST-segment elevation in the right precordial leads in affected patients (Brugada et al., 2000, 2003; Hong et al., 2004; Priori et al., 2002a; Rolf et al., 2003). Long term follow-up (Brugada et al., 2002) of patients and families with Brugada syndrome demonstrates sudden death is a persistent risk.

The finding of ST-elevation in the right chest leads has been observed in a variety of clinical and experimental settings and is not unique or diagnostic of Brugada syndrome by itself. Situations in which these ECG findings occur include electrolyte or metabolic disorders, pulmonary or inflammatory diseases, abnormalities of the central or peripheral nervous system. In the absence of these abnormalities, the term idiopathic ST-elevation is often used and may identify Brugada syndrome patients (Antzelevitch, 1999). Two consensus conferences have been held to define the clinical signs, diagnostic approaches, therapies, and diagnostic studies (Antzelevitch et al., 2005; Wilde et al., 2002a,b,c).

The ECG findings and associated sudden and unexpected death has been reported as a common problem in Southeast Asia where it most commonly affects men during sleep (Nademanee et al., 1997). This disorder, known as Sudden and Unexpected Death Syndrome (SUDS) or Sudden Unexpected Nocturnal Death Syndrome (SUNDS), has many names in Southeast Asia including bangungut (to rise and moan in sleep) in the Philippines; non-lai-tai (sleep-death) in Laos; lai-tai (died during sleep) in Thailand; and pokkuri (sudden and unexpectedly ceased phenomena) in Japan (Nademanee et al., 1997; Sangwatanaroj et al., 2002). General characteristics of SUDS include young, healthy males in whom death occurs suddenly with a groan, usually during sleep late at night. No precipitating factors are identified and autopsy findings are generally negative (Gotoh, 1976). Life-threatening ventricular tachydysrhythmias as a primary cause of SUDS have been demonstrated, with VF occurring in most cases (Hayashi et al., 1985).

SHORT QT INTERVAL SYNDROME

Clinical Features

In 2000, Gussak and colleagues identified a new familial clinical syndrome characterized by an abbreviated QTc interval (<300 ms), predisposition to life-threatening arrhythmias and a high rate of sudden death. AF may occur and on electrophysiologic evaluation, short refractory periods may be identified. Age of onset of symptoms may be young (<1 year of age) and in some cases may be responsible for SIDS.

Genetics of Short QT Syndrome

Short QT syndrome (SQT) was initially shown by Gussak et al. (2000) to have autosomal dominant inheritance. Multiple families and sporadic cases have been reported (Bellocq et al., 2004; Brugada et al., 2004; Gaita et al., 2003; Gussak et al., 2000; Priori et al., 2004). Mutations in three ion channel genes (*SQT1*, *HERG/KCNH2*; *SQT2*, *KvLQT1/KCNQ1*; *SQT3*, *Kir 2.1/KCNJ2*; Table 29.5) have been reported to date (Bellocq et al., 2004; Brugada et al., 2004; Priori et al., 2005). In the case of SQT1, mutations dramatically increased I_{Kr}, leading to heterogeneous abbreviation of action potential duration and refractoriness and reduced the affinity of the channels to I_K blockers. Bellocq et al. performed functional studies on the *KCNQ1* mutations which revealed a pronounced shift of the half-activation kinetics which led to a gain-of function of I_{Ks} and repolarization shortening. In the case of the SQT3-causing mutations in *KCNJ2*, whole-cell patch-clamp studies of this inward rectifying potassium channel Kir2.1 (I_{K1}) demonstrated a larger outward I_{K1} than wild type and shifting of peak current, as well as acceleration of the final phase of repolarization leading to shortened action potential duration. Clinically, this leads to tall and asymmetric T waves, a finding that appears to be distinct for SQT3 (Priori et al., 2005).

TABLE 29.5 Short QT syndromes

Gene	Locus	Syndrome	Protein and subunit	Function and abnormality
KCNH2	7q35	SQTS1	K_V11.1 α	*IKr↑HERG*
KCNQ1	11p15.5	SQTS2	K_V7.1 α	*IKs↑KvLQT1*
KCNJ2	17q23	SQTS3	Kir2.1 α	*IK1↑*

Therapy in SQT Syndrome

Since sudden death occurs commonly in this disorder, defibrillator implantation is indicated. Gaita et al. (2004) evaluated the role of medical therapy, testing QTc interval response to flecainide, sotalol, ibutilide, and quinidine with only quinidine producing QTc prolongation (normalization). However, no definitive treatment approaches have been described to date. Wolpert et al. (2005) evaluated the role of quinidine in SQTS due to HERG mutations as well.

FAMILIAL VT/CPVT

Clinical Features

This inherited form of VT was first described in detail by Rubin et al. (1992) and is another form of VT that occurs in the absence of structural heart disease. Clinically, these patients present with frequent runs of nonsustained VT at rest, which may decrease or extinguish with exercise (Rubin et al., 1992; Sacks et al., 1978; Vlay, 1987) and may occur during childhood with episodes of Adams Strokes syndrome (Coumel et al., 1978; Leenhardt et al., 1995). In some cases, salvos of nonsustained VT may be incessant, causing palpitations, dizziness or syncope. The ECG is usually normal during sinus rhythm, with a normal QT interval. Electrophysiologic evaluation may not consistently show inducible VT in these patients, suggesting that the mechanism for this dysrhythmia is not reentry, as with more common forms of VT. Instead, it is believed that this form of inherited dysrhythmia is due to enhanced automaticity (i.e., an increased rate of electrical firing of a ventricular myocyte) or triggering (i.e., results from secondary depolarizations that occur during or immediately after repolarization).

Sumitomo et al. (2003) described the ECG characteristics of CPVT and clinical features in detail. The initial manifestations included young age of presentation (10.3 years, mean age of onset), syncope (79%), cardiac arrest (7%), and family history (14%). ECGs demonstrated sinus bradycardia and normal QTc. The CPVT morphology included polymorphic (62%), polymorphic and bidirectional (21%), bidirectional (10%) or polymorphic with VF (7%). Cohen et al. (1989) demonstrated a sudden death risk with familial bidirectional ventricular tachycardia.

Genetics and Management

This inherited tachyarrhythmia is transmitted as an autosomal dominant trait (Tables 29.2 and 29.4; Glikson et al., 1991; Wren et al., 1990). Mutations in the ryanodine receptor (RyR2) have been identified as causative in CPVT (Laitinen et al., 2001; Marks et al., 2002; Priori et al., 2001, 2002b; Swan et al., 1999). In addition, mutations of the calsequestrin gene have been found in CPVT (Eldar et al., 2003; Lahat et al., 2002; Postma et al., 2002) appears to be relatively uncommon but the incidence and prevalence are not known. The therapy of this disorder appears to rely on the use of β-adrenergic blocking agents, which effectively suppresses this rhythm disturbance. However, Sumitomo showed complete β-blocker control in only 31% of cases (2003). In addition, the use of class I and class III antiarrhythmics have reportedly been successful in treating this disorder. Sumitomo and colleagues also showed calcium channel antagonists partially suppressed CPVT in some familial autosomal dominant cases. The use of ICDs have not been reported in this disorder but this is certainly a likely option in malignant families or difficult cases.

PRIMARY CONDUCTION ABNORMALITIES

Lev-Lenegre Progressive Cardiac Conduction Disease

Clinical Features

This syndrome, also known as progressive cardiac conduction defect (PCCD), is amongst the most common cardiac conduction disorders in the world and represents the major cause for pacemaker implantation worldwide at 0.15 pacemaker implantations per 1000 inhabitants per year in developed countries (Schott et al., 1999). PCCD is characterized by progressive worsening of cardiac conduction through the His-Purkinje system, resulting in right or left bundle branch block and widening of the QRS complex (Lenegre and Moreau, 1963; Lev, 1964; Lev et al., 1970). Patients with this disorder ultimately develop complete AV block (Figure 29.6) and in many instances present with syncope or sudden death. Etiologically, PCCD has been considered to be a primary degenerative disease or an exaggerated age-related process in which sclerosis of the conduction system occurs (Gault et al., 1972; Gazes et al., 1965).

Genetics

The first gene for PCCD was mapped to chromosome 19q13.3 (Table 29.6) in families with autosomal dominantly transmitted disease (Brink et al., 1995; Connor et al., 1959; De Meeus et al., 1995; Waxman et al., 1975). Genetic heterogeneity was identified when Schott et al. (1999) demonstrated linkage to chromosome 3p21 and identified mutations in the cardiac sodium channel gene *SCN5A* (Tables 29.3 and 29.6; Figure 29.7). This gene had previously been shown to cause a form of long QT syndrome (LQT3) as well as Brugada syndrome (Chen et al., 1998; Wang et al., 1995a, b). Other supportive evidence of *SCN5A* mutations causing AV block have been reported. Miura et al. (2003) identified *SCN5A* mutations in congenital LQT and 2:1 AV block while Shirai et al. (2002) showed that mutations in this gene may cause overlapping features of Brugada syndrome and cardiac conduction disease. Tan and colleagues showed that mutations in *SCN5A* results in isolated conduction system disease as well, and this was confirmed by Wang et al. (2002). In addition, Viswanathan et al. (2003) showed that a common *SCN5A* polymorphism (H558R) mitigates the effects of mutations on channel function and its clinical response. Mechanisms responsible for these varying clinical features, patient sudden death, have been speculated to be based on the biophysical responses to the mutations in SCN5A (Towbin, 2001a, b).

Management

Patients with PCCD require pacemaker therapy once bradycardia occurs. In cases where syncope precedes diagnosis, immediate implantation is appropriate. Prior to pacemaker implantation, serial ECGs and Holter monitor studies are necessary for close monitoring.

Sinus Node Dysfunction/Sick Sinus Syndrome

Clinical Features

Congenital absence of sinus rhythm (Surawicz and Hariman, 1988) or SND (Caralis and Varghese, 1976; Spellberg, 1971) occurs in familial forms. The clinical presentation is generally due to symptomatic bradycardia, but it also may be associated with paroxymal AF and other atrial tachydysrhythmias. Histologic data are limited, but one report of a single member of a family who suffered sudden death showed mononuclear cell infiltration of the

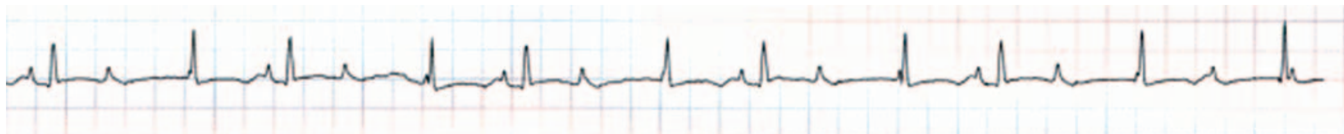


Figure 29.6 Complete AV block in this individual with bradycardia.

TABLE 29.6 Atrial arrhythmia syndromes

Gene	Locus	Syndrome	Protein and subunit	Function and abnormality
KCNQ1	11p15.5	ATFB1, LQTS1	$K_V7.1$ α	$I_{Ks}\uparrow$ KvLQT1
KCNQ1	11p15.5	ATFB1, SQTS2	$K_V7.1$ α	$I_{Ks}\uparrow$ KvLQT1
KCNE2	21q22.1	ATFB1	MiRP1 β	$I_{Ks}\uparrow$
KCNE1	21q22.1	AF	minK β	$I_{Ks}\downarrow$
KCNA5	12p13	AF	$K_V1.5$ α	$I_{Kur}\downarrow$
KCNJ2	17q23	ATFB1	Kir2.1 α	$I_{K1}\uparrow$
ANKβ	4q25	AF, Bradycardia, LQTS4	Ankyrin-B	$I_{Na,K}\downarrow$ $I_{NCX}\downarrow$
KCNH2	7q35	AF, SQTS2	$K_V11.1$ α	$I_{Kr}\uparrow$ HERG
SCN5A	3p21	Congenital SSS	$Na_V1.5$ α	$I_{Na}\downarrow$
GJA5	1q21.1	ATFB4	Connexin 40	Coupling ↓
SCN5A GJA5 *GJA5*	3p21 1q21.1	Atrial standstill (Coinheritance)	$Na_V1.5$ α Connexin 40	$I_{Na}\downarrow$ Coupling ↓
RyR2	1q42	AT, CPVT1	RyR2 α	SR Ca^{2+} leak ↑
HCN4	15q24	SSS, AF, Bradycardia	HCN4 α	$I_f\downarrow$

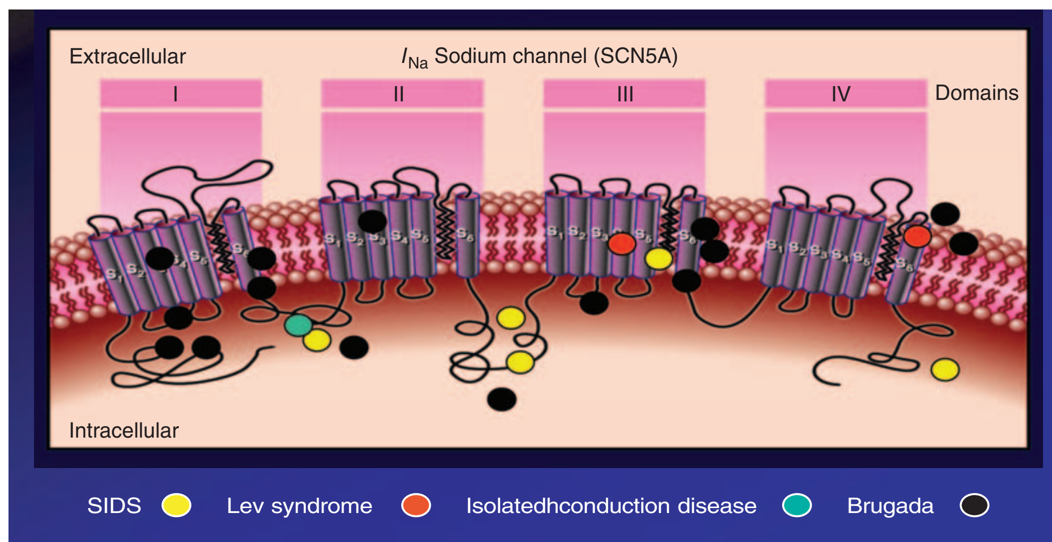


Figure 29.7 Mutations in SCN5A in patients with Brugada syndrome (black).

sinus node; fibrosis of the sinus, AV nodes, and atrial tissue; and atrophy of the right and left bundle branches (Bharati et al., 1992).

Genetics and Management

The inheritance pattern of familial congenital absence of sinus node function is autosomal dominant with a high degree of penetrance (Table 29.6). The gene for this disorder remains unknown. However, Benson et al. (2003) identified homozygous mutations in the cardiac sodium channel gene *SCN5A* (Figure 29.7) in individuals with SSS.

Treatment consists of permanent pacing for symptomatic bradycardia and, unlike persistent atrial standstill, in which the atria are unable to be paced due to inability to produce electrical excitation, atrial or dual-chamber pacing may be used. If AF is a prominent feature, antiarrhythmic medications and/or chronic anticoagulation may be indicated.

Atrial Fibrillation

Clinical Aspects

Atrial fibrillation (AF), first described in 1906, is a cardiac rhythm disorder characterized by rapid, irregular activation of the atria, resulting in loss of coordinated atrial contraction and therefore reduced ventricular filling and stasis of blood in the atria (Fye, 2006). This predisposes to heart failure and thromboembolic stroke (Albers, 1994; Wang et al., 2003; Wolf et al., 1991). It is currently believed to be the most common sustained cardiac

arrhythmia worldwide, accounting for 15% of all strokes and over 30% of all strokes in subjects 65 years of age or older (Wolf et al., 1991). The prevalence of AF reportedly has a age dependence, ranging from <1% in young adults to approximately 10% in individuals older than 80 years of age (Kannel et al., 1998). In the United States, more than 3 million people are estimated to have persistent AF (Feinberg et al., 1995; Go et al., 2001). AF is an independent risk factor for death, as well (Benjamin et al., 1998).

AF may be classified as paroxysmal, persistent or permanent (Boston Area Anticoagulation Trial, 1990). The paroxysmal form of AF accounts for 35–40% of all AF, and there is a 30–50% chance of paroxysmal AF converting to a permanent form (Boston Area Anticoagulation Trial, 1990). It is frequently observed as a complication of a variety of cardiovascular and systemic disorders, such as hypertension, coronary artery disease, valvular heart disease, various forms of congenital heart disease, as well as cardiomyopathies. Thus, AF is typically considered to be a sporadic, non-genetic form of disease. In approximately 10–20% of cases, however, no underlying associated defect is identified and, therefore, it is termed "idiopathic" or "lone" AF (Brand et al., 1985). In rare occasion, fetal onset has been identified (Tikanoja et al., 1998).

Electrophysiology of AF

Atrial conduction depends on the connections between cells, the structural properties of the atrial wall, and ion flux. Cardiomyocytes link in an end-to-end fashion by intercalated disks and therefore form a syncytium of elongated branching fibers. The structural components of atria are required to function properly for normal electrical activity. Atrial electrophysiology is a complex process which requires coordinated interactions between multiple ionic and structural factors in order to generate normal impulse formation and propagation (Weiss et al., 2005).

Genetics of AF

The first case of apparent inherited AF was described in 1936 and subsequently a small number of families have been published in the literature (Arnar et al., 2006; Bedi et al., 2006; Beyer et al., 1993; Darbar et al., 2003; Ellinor et al., 2005; Fox et al., 2004; Orgain et al., 1936). Over the past decade, it has become apparent that familial aggregation of AF is relatively common and, for this reason, genetic analyses were able to be pursued. The first genetic locus for AF was reported in 1997 by Brugada and colleagues, identifying chromosome 10q22-q24 linkage. A second locus was later identified by Oberti et al. (2004) on chromosome 5p13 in a family with neonatal-onset autosomal recessive AF. Later, Chen et al. (2003) identified linkage in a single family with AF on chromosome 11p15 and, using a candidate gene screening approach, identified a mutation in *KCNQ1*. This was subsequently confirmed in a small number of subjects and appears to be an uncommon cause of AF (Ellinor et al., 2004, 2006; Otway et al., 2007). Another uncommon cause of AF includes mutations in the β-subunits of the cardiac I_{Ks} channel, including KCNE2 and possibly KCNE3 (Yang et al., 2004; Zhang et al., 2005). The *KCNJ2* gene located on chromosome 17q23-q24, which encodes the Kir2.1 protein that forms the α-subunit of the cardiac I_{KI}, has also been identified in a small number of subjects (Xia et al., 2005). A novel and promising gene, *KCNA5*, an atrial-specific gene located on chromosome 12p13, has been reported to be mutated and causative of AF in a single family with AF completes the genes identified to date that cause predominant AF (Olson et al., 2006). An additional group of genes have been found to carry mutations in subjects with combined atrial and ventricular phenotypes, and these include *KCNH2*, *SCN5A*, and *LMNA*. In the case of *KCNH2*, an N588K missense mutation was identified in one family with an overlapping phenotype including AF and SQT syndrome (Hong et al., 2005). This same mutation was identified in two other families only presenting with SQTS alone (Brugada et al., 2004). Similarly, another ion channel-encoding gene that is well known as a cause of arrhythmia disorders, SCN5A, has been shown to cause complex phenotypes that include AF, in particular those associated with dilated cardiomyopathy and conduction system disease (McNair et al., 2004; Olson et al., 2005). Finally, another pleiomorphic gene, *lamin A/C* (*LMNA*), has also been shown to cause AF in patients and families with dilated cardiomyopathy and conduction disease (Fatkin et al., 1999; Sebillon et al., 2003).

Another group of mutations that have been described in subjects are so-called "somatic mutations." This group of abnormalities differs from the germline mutations described above. Typically, genetic-based disorders are inherited in families by transmission of gene mutations in the germ cells. Mutations can also arise *de novo* in discrete cell lineages during embryonic development or postnatally. Germ cell *de novo* mutations can be transmitted to future generations while somatic mutations cannot, but in this case can give rise to mixed populations of mutant and normal cells within specific tissues, known as mosaicism. Mosaic mutations may have variable functional consequences and hence the clinical phenotype that develops may be unpredictable. In AF, somatic mutations have been associated in a small number of studies. Gollob et al. (2006) identified mutations in the *GJA5* gene, encoding the cardiac gap junction protein connexin40, in DNA from atrial tissue from 4 subjects; however, none of these variants were found in DNA extracted from blood. In one other subject, a mutation was identified in DNA extracted from both atrial tissue and blood. Functional studies was performed on the mutations and showed reduction of gap junction formation and/or coupling properties.

Genetic Mechanisms in AF

The majority of mutations in subjects with AF have occurred in genes encoding potassium channels. Mutations in *KCNQ1* and *KCNE2* appear to increase I_{Ks} channel activation while mutations in *KCNJ2* and *KCNH2* lead to increased activation of cardiac I_{KI} and I_{Kr} currents, respectively. These gain-of-function mutations all result in shortening of the action potential duration and effective refractory period, thereby promoting AF by creating an electrical substrate for reentry.

2009 UPDATE

During the past year, several new discoveries have occurred in understanding the genetics of arrhythmias. Two new LQTS-causing genes have been identified, *AKAP9* (LQT11) and *SNTA1* (LQT12). Chen et al. (2007) reported mutations in *AKAP9* (LQT11) that affect the KCNQ1 binding domain and result in functional disturbance of the KCNQ1 potassium channel. Mutations in *SNTA1* (LQT12) were described by Ueda et al. (2008), who reported an A390V variant that affects the SNTA1 link to the neuronal nitric oxide synthase (nNOS) inhibitor plasma membrane Ca-ATPase subtype 4b (PMCA4b), causing the release of inhibition of nNOS. This leads to *S*-nitrosylation of $Na_v1.5$, resulting in an increase in late sodium current. In addition, Wu et al. (2008) identified an A257G variant in *SNTA1* that led to increased peak sodium currents, which negatively shifted the onset and peak currents of the current-voltage relationships and shifted steady-state activation leftward; no change in the voltage-dependent inactivation kinetics and late sodium currents were observed.

In addition to the standard candidate gene approach, novel genetic linkage studies aimed at identifying new disease-causing genes have included genetic association studies (GAS) and genome-wide association studies (GWAS). In the case of GAS, these studies are a form of candidate gene screen that aims to assess the association between disease status and genetic variants (polymorphisms, single nucleotide polymorphisms or SNPs) in a population, while GWAS is an unbiased, hypothesis-free approach that involves large scans of the genome using a dense set of SNPs (up to 500,000 or more) aimed at identifying causal variants (Donohue and Allen, 2005). GWAS studies generate vast datasets, leading to challenges both in primary analysis and meta-analysis (Hirschhorn and Daly, 2005; Visscher et al., 2008; Zintzaras and Lau, 2008).

Over the years, GAS and GWAS studies have come under scrutiny due to the lack of replication of data by independent studies. Major methodological issues that have surfaced include small sample size and insufficient power to detect minor contributing roles of one or more alleles. Other confounding factors may include population stratification and experimental design flaws have been problematic. More recently, many of these challenges have been overcomed and new genes and pathways have been identified using GWAS. However, effect sizes are small in these studies, with relative genotype risks typically <1.5. In the case of quantitative traits, the individual effect sizes are consistent with $<1\%$ of the phenotypic variance being explained by a single polymorphism. Hence, while challenges exist for this approach, successes have been forthcoming and excitement appears to exist in regards to the future impact of these methods.

Reliable information concerning the clinical, ethnic, and genetic background of the subjects under study is essential when performing association studies. In this regard, the Framingham Heart Study represents an extraordinary effort to collect family members throughout multiple generations and includes the recruitment of 5209 individuals from the first generation in 1948, 5124 from the second generation in 1971, and 4095 from the third generation in 2002. Recently, the GWAS approach was applied to the Framingham study population in an attempt to identify genetic factors associated with arrhythmias. Larson et al. (2007) evaluated the Framingham study population and identified linkage to chromosome 9p21 for subjects with major cardiovascular disease, as well as associating linkage of this region with AF. Gudbjartsson et al. (2007) identified variants conferring risk of AF on chromosome 4q25. These and other studies (Ellinor et al., 2008) are quickly focusing new studies for new targets responsible for these phenotypes. More recently, Newton-Cheh et al. (2007) showed, using an Affymetrix 100K GeneChip on 1345 related Framingham Heart Study Original and Offspring cohort participants, that common variants in NOS1AP are associated with prolongation of the QT interval and heart rate variation.

In the case of atrial fibrillation, SNPs in several cardiac ion channel genes have been linked with the disease. In several reports, I_{Ks} channel components have been linked with atrial fibrillation, including KCNE1, KCNE2, and KCNE5 (Chen et al., 2007; Ehrlich et al., 2005; Fatini et al., 2007; Lai et al., 2002; Ravn et al., 2005; Zeng et al., 2007). In addition, SNPs in genes that alter regulation of ion channel function (Bedi et al., 2006; Firouzi et al., 2007; Nyberg et al., 2007; Schreieck et al., 2004), intracellular handling of calcium (Firouzi et al., 2007), gap junction formation (Juang et al., 2007; Yamashita et al., 1997), and activation of the renin-angiotensin system (Fatini et al., 2007; Tsai et al., 2004) have been reported. These associated genes include GNB3 and NOS3, which regulate channel function, and GJA5, which is involved in the formation of gap junctions, amongst others.

2009 UPDATE REFERENCES

Bedi, M., McNamara, D., London, B. and Schwartzman, D. (2006). Genetic susceptibility to atrial fibrillation in patients with congestive heart failure. *Heart Rhythm* 6, 808–812.

Chen, L., Marquardt, M.L., Tester, D.J., Sampson, K.J., Ackerman, M.J. and Kass, R.S. (2007). Mutation of an A-kinase-anchoring protein causes long-QT syndrome. *Proc Natl Acad Sci U S A* 104, 20990–20995.

Donahue, M.P. and Allen, A.S. (2005). Genetic association studies in cardiology. *Am Heart J* 149, 964–970.

Ehrlich, J.R., Zicha, S., Coutu, P., Hebert, T.E. and Nattel, S. (2005). Atrial fibrillation-associated mink38G/S polymorphism modulates delayed rectifi er current and membrane localization. *Cardiovasc Res* 67, 520–528.

Ellinor, P.T., Yi, B.A. and MacRae, C.A. (2008). Genetics of atrial fibrillation. *Med Clin North Am* 92, 41–51.

Fatini, C., Sticchi, E., Gensini, F., Gori, A.M., Marchucci, R., Lenti, M., Michelucci, A., Genuardi, M., Abbate, R. and Gensini, G.F. (2007). Lone and secondary nonvalvular atrial fibrillation: Role of a genetic susceptibility. *Int J Cardiol* 120(1), 59–65.

Firouzi, M., Ramanna, H., Kok, B., Jongsma, H.J., Koelman, B.P.C., Doevendans, P.A., Groenewegen, W.A. and Hauer, R.N.W. (2004). Association of human connexin 40 gene polymorphisms with atrial vulnerability as a risk factor for idiopathic atrial fibrillation. *Circ Res* 95, e29–e33.

Gudbjartsson, D.F., Arnar, D.O., Helgadottir, A., Gretarsdottir, S., Holm, H., Sigurdsson, A., Jonasdottir, A., Baker, A., Thorleifsson, G., Kristjansson, K. et al. (2007). *Nature* 448(7151), 353–357.

Hirschhorn, J.W. and Daly, M.J. (2005). Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet*, 65–108.

Juang, J.M., Chern, Y.R., Tsai, C.T., Chiang, F.T., Lin, J.L., Hwang, J.J., Hsu, K.L., Tseng, C.D., Tseng, Y.Z. and Lai, L.P. (2007). The association of human connexin 40 genetic polymorphisms with atrial fibrillation. *Int J Cardiol* 116, 107–112.

Lai, L.P., Su, M.J., Yeh, H.M., Lin, J.L., Chiang, F.T., Hwang , J.J., Hsu, K.L., Tseng, C.D., Lien, W.P., Tseng, Y.Z. et al. (2002). Association of the human mink gene 38G allele with atrial fibrillation: Evidence of possible genetic control on the pathogenesis of atrial fibrillation. *Am Heart J* 144, 485–490.

Larson, M.G., Atwood, L.D., Benjamin, E.J., Cupples, L.A., D'Agostino, R.B. Sr, Fox, C.S., Govindaraju, D.R., Guo, C.Y., Heard-Costa, N.L., Hwang, J.J. et al. (2007). Framingham Heart study 100K project: Genomewide associations for cardiovascular disease outcomes. *BMC Med Genet* 8(Suppl 1), S5.

Newton-Cheh, C., Guo, C.Y., Wang, T.J., O'donnell, C.J., Levy, D. and Larson, M.G. (2007). Genome-wide association study of electrocardiographic and heart rate variability traits: The Framingham Heart Study. *BMC Med Genet* 8(Suppl 1), S7.

Nyberg, M.T., Stoevring, B., Behr, E.R., Ravn, L.S., McKenna, W.J. and Christiansen, M. (2007). The variation of the sarcolipin gene (SLN) in atrial fibrillation, long QT syndrome and sudden arrhythmic death syndrome. *Clin Chim Acta* 375, 87–91.

Ravn, L.S., Hoffman-Bang, J., Dixen, U., Larsen, S.O., Jenesen, G., Haunso, S., Svendsen, J.H. and Christiansen, M. (2005). Relation of 97T polymorphism in KCNE5 to risk of atrial fibrillation. *Am J Cardiol* 96, 405–407.

Schreieck, J., Dostel, S., von Beckerath, N., Wacker, A., Flory, N., Weyerbrock, S., Koch, W., Schomig, A. and Schmitt, C. (2004). C825T polymorphism of the G-protein beta-3 subunit gene and atrial fibrillation: association of the TT genotype with a reduced risk for atrial fibrillation. *Am Heart J* 148, 545–550.

Tsai, C.T., Lai, L.P., Lin, J.L., Chiang, F.T., Hwang, J.J., Ritchie, M.D., Moore, J.H., Hsu, K.L., Tseng, C.D., Liau, C.S. et al. (2004). Renin-angiotensin system gene polymorphisms and atrial fibrillation. *Circulation* 109, 1640–1646.

Visscher, P.M., Andrew, T. and Nyholt, D.R. (2008). Genome-wide association studies of quantitative traits with related individuals: Little (power) lost but much to be gained. *Eur J Hum Genet* 16(3), 387–390.

Ueda, K., Valdivia, C., Medeiros-Domingo, A., Tester, D.J., Vatta, M., Farrugia, G., Ackerman, M.J. and Makielski, J.C. (2008). Syntrophin mutation associated with long QT syndrome through activation of the nNOS-SCN5A macromolecular complex. *Proc Natl Acad Sci U S A* 105, 9355–9360.

Wu, G., Ai, T., Kim, J.J., Mohapatra, B., Xi, Y., Li, Z., Abbasi, S., Purevjav, E., Smani, K., Ackerman, M.J. et al. (2008). Alpha-1-syntrophin mutation and the long QT syndrome: A disease of sodium channel disruption. *Circ Arrhythm Electrophysiol* 1, 193–201.

Yamashita, T., Hayami, N., Ajiki, K., Oikawa, N., Sezaki, K., Inoue, M., Omata, M. and Murakawa, Y. (1997). Is ACE gene polymorphism associated with lone atrial fibrillation? *Jpn Heart J* 38, 637–641.

Zeng, Z., Tan, C., Teng, S., Chen, J., Su, S., Zhou, X., Wang, F., Zhang, S., Gu, D., Makielski, J.C. et al. (2007). The single nucleotide polymorphisms of I (Ks) potassium channel genes and their association with atrial fibrillation in a Chinese population. *Cardiology* 108, 97–103.

Zintzaras, E. and Lau, J. (2008). Trends in meta-analysis of genetic association studies. *J Hum Genet* 53, 1–9.

REFERENCES

Abbott, G.W., Sesti, F., Splawski, I. et al. (1999). *MiRP1* forms I_{Kr} potassium channels with *HERG* and is associated with cardiac arrhythmia. *Cell* 97, 175–187.

Ackerman, M.J., Tester, D.J., Porter, C.J. and Edwards, W.D. (1999). Molecular diagnosis of the inherited long-QT syndrome in a woman who died after near-drowning. *N Engl J Med* 341, 1115–1121.

Ackerman, M.J., Siu, B., Sturner, W.Q., Tester, D.J., Valdivia, C.R., Makielski, J.C. and Towbin, J.A. (2001). Postmortem molecular analysis of SCN5A defects in sudden infant death syndrome. *JAMA* 286, 2264–2269.

Aihara, N., Ohe, T., Kamakura, S. et al. (1990). Clinical and electrophysiologic characteristics of idiopathic ventricular fibrillation. *Shinzo* 22(Suppl. 2), 80–86.

Albers, G.W. (1994). Atrial fibrillation and stroke: Three new studies, three remaining questions. *Arch Intern Med* 154, 1443–1448.

An, R.H., Wang, X.L., Kerem, B. et al. (1998). Novel *LQT-3* mutation affects Na+ channel activity through interactions between alpha- and beta 1-subunits. *Circ Res* 83, 141–146.

Andelfinger, G., Tapper, A.R., Welch, R.C., Vanoye, C.G., George, A.L. Jr. and Benson, D.W. (2002). KCNJ2 mutation results in Andersen syndrome with sex-specific cardiac and skeletal muscle phenotypes. *Am J Hum Genet* 71, 663–668.

Andersen, K.D., Krasilnikoff, P.A. and Overvad, H. (1971). Intermittent muscular weakness, extrasystoles, and multiple developmental anomalies: A new syndrome? *Acta Paediat Scand* 60, 559–564.

Antzelevitch, C. (1999). Ion channels and ventricular arrhythmias: Cellular and ionic mechanisms underlying the Brugada syndrome. *Curr Opin Cardiol* 14, 274–279.

Antzelevitch, C., Sicouri, S., Luteas, A. et al. (1995). Regional differences in the electrophysiology of ventricular cells: Physiological and clinical implications. In *Cardiac Electrophysiology: From Cell to Bedside* (D.P. Zipes and J. Jalife, eds), WB Saunders, Philadelphia, pp. 228–245.

Antzelevitch, C., Brugada, P., Brugada, J., Brugada, R., Towbin, J.A. and Nademanee, K. (2003). Brugada syndrome: 1992–2002: A Historical Perspective. *J Am Coll Cardiol* 41, 1665–1671.

Antzelevitch, C., Brugada, P., Borggrefe, M., Brugada, J., Brugada, R., Corrado, D., Gussak, I., LeMarec, H., Nademanee, K., Perez Riera, A.R. et al. (2005). Brugada syndrome: Report of the second consensus conference: Endorsed by the Heart Rhythm Society and the European Heart Rhythm Association. *Circulation* 111, 659–670.

Arcangeli, A., Rosati, B., Cherubini, A. et al. (1997). *HERG*- and IRK-like inward rectifier currents are sequentially expressed during neuronal crest cells and their derivatives. *Eur J Neurosci* 9, 256–2604.

Arnar, D.O., Thorvaldsson, S., Manolio, T.A., Thorgeirsson, G., Kristjansson, K., Hakonarson, H. and Stefansson, K. (2006). Familial aggregation of atrial fibrillation in Iceland. *Eur Heart J* 27, 708–712.

Asano, Y., Davidenko, J., Bater, W.T. et al. (1997). Optical mapping of drug-induced polymorphic arrhythmias and *torsades de pointes* in the isolated rabbit heart. *J Am Coll Cardiol* 29, 831–842.

Barhanin, J., Lesage, F., Guillemare, E., Finc, M., Lazdunski, M. and Romey, G. (1996). *KVLQT1* and IsK (*minK*) proteins associate to form the I_{Ks} cardiac potassium current. *Nature* 384, 78–80.

Bedi, M., McNamara, D., London, B. and Schwartzman, D. (2006). Genetic susceptibility to atrial fibrillation in patients with congestive heart failure. *Heart Rhythm* 6, 808–812.

Bellavere, F., Ferri, M., Guarini, L., Bax, G., Piccoli, A., Cardone, C. and Fedele, D. (1988). Prolonged QT period in diabetic autonomic neuropathy: A possible role in sudden cardiac death. *Br Heart J* 59, 379–383.

Bellocq, C., van Ginneken, A.C., Bezzina, C.R., Alders, M., Escando, D., Mannens, M.M., Baro, I. and Wilde, A.A. (2004). Mutation in the KCNQ1 gene leading to the short QT-interval syndrome. *Circulation* 109, 2394–2397.

Bendahhou, S., Donaldson, M.R., Plaster, N.M., Tristani-Firouzi, M., Fu, Y.H. and Ptacek, L.J. (2003). Defective potassium channel Kir2.1 trafficking underlies Andersen-Tawil syndrome. *J Biol Chem* 278, 51779–51785.

Benhorin, J., Kalman, Y.M., Medina, A., Towbi, J., Rave-Harel, N., Dyer, T.D., Blangero, J., MacCluer, J.W. and Kerem, B.S. (1993). Evidence of genetic heterogeneity in the long QT syndrome. *Science* 260, 1960–1962.

Benhorin, J., Goldmit, M., MacCluer, J.W., Blangero, J., Goffen, R., Leibovitch, A., Rahat, A., Wang, Q., Medina, A., Towbin, J. et al. (1998). Identification of a new SCN5A mutation, D1840G, associated with the long QT syndrome. *Hum Mut* 12, 72.

Benhorin, J., Taub, R., Goldmit, M., Kerem, B., Kass, R.S., Windman, I. and Medina, A. (2000). Effects of flecainide in patients with new SCN5A mutation: Mutation-specific therapy for long-QT syndrome? *Circulation* 101, 1698–1706.

Benjamin, E.J., Wolf, P.A., D'Agostino, R.B., Silbershatz, H., Kannel, W.B. and Levy, D. (1998). Impact of atrial fibrillation on the risk of death: The Framingham Heart Study. *Circulation* 98, 946–952.

Bennett, P.B., Yazawa, K., Makita, N. and George, A.L. Jr. (1995). Molecular mechanism for an inherited cardiac arrhythmia. *Nature* 376, 683–685.

Bennett, V., and Baines, A.J. (2001). Spectrin and ankyrin-based pathways: Metazon inventions for integrating cells into tissues. *Physiol Rev* 81, 1353–1392.

Benson, D.W., Wang, D.W., Dyment, M., Knilans, T.K., Fish, F.A., Strieper, M.J., Rhodes, T.H. and George, A.L. Jr. (2003). Congenital sick sinus syndrome caused by recessive mutations in the cardiac sodium channel gene (*SCN5A*). *J Clin Invest* 112, 1019–1028.

Beyer, F., Paul, T., Luhmer, I., Bertram, H. and Kallfelz, H.C. (1993). Familial idiopathic atrial fibrillation with bradyarrhythmia. *Zeitschrift fuer Kardiologie* 83, 674–677.

Bharati, S., Surawicz, B., Vidaillet, H.J. et al. (1992). Familial congenital sinus rhythm abnormalities: Clinical and pathological correlations. *PACE* 15, 1720–1729.

Bianchi, L., Wible, B., Arcangeli, A. et al. (1998). *HERG* encodes a K+ current highly conserved in tumors of different histogenesis: A selective advantage for cancer cells? *Canc Res* 58, 815–822.

Bianchi, L., Shen, Z., Dennis, A.T. et al. (1999). Cellular dysfunction of *LQT5-minK* mutants: Abnormalities of I_{Ks}, I_{Kr} and trafficking in long QT syndrome. *Hum Mol Genet* 8, 1499–1507.

Bowles, N.E., Bowles, K.R. and Towbin, J.A. (2000). The "final common pathway" hypothesis and inherited cardiovascular disease: The role of cytoskeletal proteins in dilated cardiomyopathy. *Herz* 25, 168–175.

Brand, F.N., Abbott, R.D., Kannel, W.B. and Wolf, P.A. (1985). Characteristics and prognosis of lone atrial fibrillation. *JAMA* 254, 3449–3453.

Brink, P.A., Fereira, A., Moolman-Weymar, H.W., van der Merwe, P.L. and Corfield, V.A. (1995). Gene for progressive familial heart block type I maps to chromosome 19q13. *Circulation* 91, 1633–1640.

Brugada, J. and Brugada, P. (1997). Further characterization of the syndrome of right bundle branch block, ST segment elevation, and sudden death. *J Cardiovasc Electrophysiol* 8, 325–331.

Brugada, P. and Brugada, J. (1992). Right bundle branch block, persistent ST segment elevation and sudden cardiac death: A distinct clinical and electrocardiographic syndrome. A multi center report. *J Am Coll Cardiol* 20, 1391–1396.

Brugada, J., Brugada, R., Antzelevitch, C., Towbin, J., Nademanee, K. and Brugada, P. (2002). Long-term follow-up of individuals with the electrocardiographic pattern of right-bundle-branch block and ST-segment elevation in precordial leads V_1 to V_3. *Circulation* 105, 73–78.

Brugada, J., Brugada, P. and Brugada, R. (2003). The ajmaline challenge in Brugada syndrome. *Eur Heart J* 24, 1085–1086.

Brugada, P. and Brugada, J. (1991). A distinct clinical and electrocardiographic syndrome: Right bundle-branch block, persistent ST segment elevation with normal QT interval and sudden cardiac death. *PACE* 14, 746.

Brugada, P., Brugada, R., Antzelevitch, C. and Brugada, J. (2005). The Brugada syndrome. *Arch Mal Coeur Vaiss* 98, 115–122.

Brugada, R., Tapscott, T., Czernusewicz, G.Z., Marian, A.J., Iglesias, A., Mont, L., Brugada, J., Girona, J., Domingo, A., Bachinski, L.L. et al. (1997). Identification of a genetic locus for familial atrial fibrillation. *N Engl J Med* 336, 905–911.

Brugada, R., Brugada, J., Antzelevitch, C. et al. (2000). Sodium channel blockers identify risk for sudden death in patients with ST-segment elevation and right bundle branch block but structurally normal hearts. *Circulation* 101, 510–515.

Brugada, R., Hong, K., Dumaine, R., Cordeiro, J., Gaita, F., Borggrefe, M., Menendez, T.M., Brugada, J., Pollevick, G.D., Wolpert, C. et al. (2004). Sudden death associated with short-QT syndrome linked to mutations in HERG. *Circulation* 109, 30–35.

Caralis, D.G. and Varghese, P.J. (1976). Familial sinoatrial node dysfunction: Increased vagal tone a possible aetiology. *Br Heart J* 38, 951–956.

Chatrath, R., Bell, C.M. and Ackerman, M.J. (2004). Beta-blocker therapy failures in symptomatic probands with genotyped long-QT syndrome. *Pediatr Cardiol* 25, 459–465.

Chauhan, V.S., Tuvia, S., Buhusi, M., Bennett, V. and Grant, A.D. (2000). Abnormal cardiac Na+ channel properties and QT heart rate

adaptation in neonatal ankyrin-β knockout mice. *Circ Res* 86, 441–447.

Chen, Q., Kirsch, G.E., Zhang, D. et al. (1998). Genetic basis and molecular mechanism for idiopathic ventricular fibrillation. *Nature* 392, 293–296.

Chen, Q., Zhang, D., Gingell, R.L. et al. (1999). Homozygous deletion in *KVLQTI* associated with Jervell and Lange-Nielsen syndrome. *Circulation* 99, 1344–1347.

Chen, Y.H., Xu, S.J., Bendahhou, S., Wang, X.L., Wang, Y., Xu, W.Y., Jin, H.W., Sun, H., Su, X.Y., Zhuang, Q.N. et al. (2003). KCNQ1 gain-of-function mutation in familial atrial fibrillation. *Science* 299, 251–254.

Choi, G., Kopplin, L.J., Tester, D.J., Will, M.L., Haglund, C.M. and Ackerman, M.J. (2004). Spectrum and frequency of cardiac channel defects in swimming-triggered arrhythmia syndromes. *Circulation* 110, 2119–2124.

Chouabe, C., Neyroud, N., Guicheney, P., Lazdunski, M., Romey, G. and Barhanin, J. (1997). Properties of *KVLQTI* K^+ channel mutations in Romano-Ward and Jervell and Lange-Nielsen inherited cardiac arrhythmias. *EMBO* 16, 5472–5479.

Chouabe, C., Neyroud, N., Richard, P.K. et al. (2000). Novel mutations in KVLQT1 that affect I_{Ks} activation through interactions with IsK. *Cardiovasc Res* 45, 971–980.

Choy, A.M., Lang, C.C., Chomsky, D.M., Rayos, G.H., Wilson, J.R. and Roden, D.M. (1997). Normalization of acquired QT prolongation in humans by intravenous potassium. *Circulation* 96, 2149–2154.

Christiansen, M., Tonder, N., Larsen, L.A., Andersen, P.S., Simonsen, H., Oyen, N., Kanters, J.K., Jacobsen, J.R., Fosdal, I., Wettrell, G. et al. (2005). Mutations in the HERG K+-ion channel: A novel link between long QT syndrome and sudden infant death syndrome. *Am J Cardiol* 95, 433–434.

Cohen, T.J., Liem, B.L. and Hancock, W. (1989). Association of bidirectional ventricular tachycardia with familial sudden death syndrome. *Am J Cardiol* 64, 1078–1079.

Compton, S.J., Lux, R.L., Ramsey, M.R., Strelich, K.R., Sanguinetti, M.C., Green, L.S., Keating, M.T. and Mason, J.W. (1996). Genetically defined therapy of inherited long-QT syndrome. Correction of abnormal repolarization by potassium. *Circulation* 94, 1018–1022.

Connor, A.C., McFadden, J.F., Houston, B.J. et al. (1959). Familial congenital complete heart block: Case report and review of the literature. *Am J Obstet Gynecol* 78, 75–79.

Coumel, P., Fidelle, J., Lucet, V., Attuel, P. and Bouvrain, Y. (1978). Catecholamine-induced severe ventricular arrhythmias with Adams-Stokes syndrome in children: Report of four cases. *Br Heart J* 40(Suppl), 28–37.

Curran, M.E., Atkinson, D., Timothy, K. et al. (1993). Locus heterogeneity of autosomal dominant long QT syndrome. *J Clin Invest* 92, 799–803.

Curran, M.E., Splawski, I., Timothy, K.W., Vincent, G.M., Green, E.D. and Keating, M.T. (1995). A molecular basis for cardiac arrhythmia: *HERG* mutations cause long QT syndrome. *Cell* 80, 795–803.

Darbar, D., Herron, K.J., Ballew, J.D., Jahangir, A., Gersh, B.J., Shen, W.K., Hammill, S.C., Packer, D.L. and Olson, T.M. (2003). Familial atrial fibrillation is a genetically heterogeneous disease. *J Am Coll Cardiol* 41, 2185–2192.

De Meeus, A., Stephan, E., Debrus, S. et al. (1995). An isolated cardiac conduction disease maps to chromosome 19q. *Circ Res* 77, 735–740.

Demolombe, S., Baro, I., Pereon, Y., Blick, J., Mohammad-Panah, R., Pollard, H., Morid, S., Mannens, M., Wilde, A., Barhanin, J. et al. (1998). A dominant negative isoform of the long QT syndrome 1 gene product. *J Biol Chem* 273, 6837–6843.

Donaldson, M.R., Jensen, J.L., Tristani-Firouzi, M., Tawil, R., Bendahhou, S., Suarez, W.A., Cobo, A.M., Poza, J.J., Behr, E., Wagstaff, J. et al. (2003). PIP2 binding residues of Kir2.1 are common targets of mutations causing Andersen syndrome. *Neurology* 60, 1811–1816.

Donaldson, M.R., Yoon, G., Fu, Y.H. and Ptacek, L.J. (2004). Andersen-Tawil syndrome: A model of clinical variability, pleiotropy, and genetic heterogeneity. *Ann Med* 36(Suppl 1), 92–97.

Duggal, P., Vesely, M.R., Wattanasirichaigoon, D., Villafane, J., Kaushik, V. and Beggs, A.H. (1998). Mutation of the gene for IsK associated with Jervell and Lange-Nielsen and Romano-Ward forms of long QT syndrome. *Circulation* 97, 142–146.

Dumaine, R., Wang, Q., Keating, M.T. et al. (1996). Multiple mechanisms of sodium channel-linked long QT syndrome. *Circ Res* 78, 916–920.

Edeiken, J. (1954). Elevation of RS-T segment, apparent or real in right precordial leads as probable normal variant. *Am Heart J* 48, 331–339.

Eldar, M., Griffin, J.C., Abbott, J.A., Benditt, D., Bhandari, A., Herre, J.M., Benson, D.W. and Scheinman, M.M. (1987). Permanent cardiac pacing in patients with the long QT syndrome. *J Am Coll Cardiol* 10, 600–607.

Eldar, M., Pras, E. and Lahat, H. (2003). A missense mutation in the CASQ2 gene is associated with autosomal-recessive catecholamine-induced polymorphic ventricular tachycardia. *Trends Cardiovasc Med* 13, 148–151.

Ellinor, P.T., Shin, J.R., Moore, R.K., Yoerger, D.M. and MacRae, C.A. (2003). Locus for atrial fibrillation maps to chromosome 6q14-16. *Circulation* 107, 2880.

Ellinor, P.T., Moore, R.K., Patton, K.K., Ruskin, J.N., Pollak, M.R. and Macrae, C.A. (2004). Mutations in the long QT gene, KCNQ1, are an uncommon cause of atrial fibrillation. *Heart* 90, 1487–1488.

Ellinor, P.T., Yoerger, D.M., Ruskin, J.N. and MacRae, C.A. (2005). Familial aggregation in lone atrial fibrillation. *Hum Genet* 118, 175–184.

El-Sherif, N., Caref, E.B., Yin, H. and Restivo, M. (1996). The electrophysiological mechanism of ventricular arrhythmias Emanuel R, O'Brien K, in the long QT syndrome: Tridimensional mapping of activation and recovery patterns. *Circ Res* 77, 474–492.

El-Sherif, N., Chinushi, M., Caref, E.B. et al. (1997). Electrophysiological mechanism of the characteristic electrocardiographic morphology of *torsade de pointes* tachyarrhythmias in the long-QT syndrome: Detailed analysis of ventricular tridimensional activation patterns. *Circulation* 96, 4392–4399.

Etheridge, S.P., Compton, S.J., Tristani-Firouzi, M. and Mason, J.W. (2003). A new oral therapy for long QT syndrome: Long-term oral potassium improves repolarization in patients with HERG mutations. *J Am Coll Cardiol* 42, 1777–1782.

Fatkin, D., MacRae, C., Sasaki, T., Wolff, M.R., Porcu, M., Frenneaux, M., Atherton, J., Vidaillet, H.J., Spudich, S., De Girolami, U. et al. (1999). Missense mutations in the rod domain of the lamin A/C gene as causes of dilated cardiomyopathy and conduction system disease. *N Engl J Med* 341, 1172–1715.

Feinberg, W.M., Blackshear, J.L., Laupacis, A., Kronmal, R. and Hart, R.G. (1995). Prevalence, age distribution, and gender of patients with atrial fibrillation. *Arch Intern Med* 155, 469–473.

Fox, C.S., Parise, H., D'Agostino, R.B., Lloyd-Jones, D.M., Vasan, R.S., Wang, T.J., Levy, D., Wolf, P.A. and Benjamin, E.J. (2004). Parental

atrial fibrillation as a risk factor for atrial fibrillation in offspring. *JAMA* 291, 2851–2855.

Franqueza, L., Lin, M., Splawsky, I., Keating, M.T. and Sanguinetti, M.C. (1999). Long QT syndrome-associated mutations in the S4-S5 linker of KVLQT1 potassium channels modify gating and interaction with minK subunits. *J Biol Chem* 274, 21063–21070.

Furutani, M., Trudeau, M.C., Hagiwara, N. et al. (1999). Novel mechanism associated with an inherited cardiac arrhythmia. Defective protein trafficking by the mutant HERG (G601S) potassium channel. *Circulation* 99, 2290–2294.

Fye, W.B. (2006). Tracing atrial fibrillation: 100 years. *N Engl J Med* 355, 1412–1414.

Gaita, F., Giustetto, C., Bianchi, F., Wolpert, C., Schimpf, R., Riccardi, R., Grossi, S., Richiardi, E. and Borggrefe, M. (2003). Short QT syndrome. *Circulation* 108, 965.

Gaita, F., Giustetto, C., Bianchi, F., Schimpf, R., Haissaguerre, M., Calo, L., Brugada, R., Antzelevitch, C., Borggrefe, M. and Wolpert, C. (2004). Short QT syndrome: Pharmacological treatment. *J Am Coll Cardiol* 43, 1494–1499.

Gault, J.H., Cantwell, J., Lev, M. et al. (1972). Fatal familial cardiac arrhythmias: Histologic observations on the cardiac conduction system. *Am J Cardiol* 29, 548–553.

Gazes, P.C., Culter, R.M., Taber, E. et al. (1965). Congenital familial cardiac conduction defects. *Circulation* 32, 32–34.

Gellens, M., George, A.L., Chen, L. et al. (1992). Primary structure and functional expression of the human cardiac tetrodotoxin-insensitive voltage-dependent sodium channel. *Proc Natl Acad Sci USA* 89, 554–558.

zGlikson, M., Constantini, N., Grafstein, Y. et al. (1991). Familial bidirectional ventricular tachycardia. *Eur Heart J* 12, 741–745.

Go, A.S. et al. (2001). Prevalence of diagnosed atrial fibrillation in adults: National implications for rhythm management and stroke prevention: The AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) Study. *JAMA* 285, 2370–2375.

Gollob, M.H., Jones, D.L., Krahn, A.D., Danis, L., Gong, X.Q., Shao, Q., Liu, X., Veinot, J.P., Tang, A.S.L., Stewart, A.F.R. et al. (2006). Somatic mutations in the connexin40 gene (GJA5) in atrial fibrillation. *N Engl J Med* 354, 2677–2688.

Gotoh, K. (1976). A histopathological study on the conduction system of the so-called Pokkuri disease (sudden unexpected cardiac death of unknown origin in Japan). *Jpn Circ J* 40, 753–768.

Groh, W.J., Silka, M.J., Oliver, R.P., Halperin, B.D., McAnulty, J.H. and Kron, J. (1996). Use of implantable cardioverter-defibrillators in the congenital long QT syndromes. *Am J Cardiol* 78, 703–706.

Gussak, I., Brugada, P., Brugada, J., Wright, R.S., Kopecky, S.L., Chaitman, B.R. and Bjerregaard, P. (2000). Idiopathic short QT interval: A new clinical syndrome? *Cardiology* 94, 99–102.

Hartmann, H.A., Colom, L.V., Sutherland, M.L. and Noebels, J.L. (1999). Selective localization of cardiac *SCN5A* sodium channels in limbic regions of rat brain. *Nat Neurosci* 2, 593–595.

Hayashi, M., Murata, M., Satoh, M. et al. (1985). Sudden nocturnal death in young males from ventricular flutter. *Jpn Heart J* 26, 585–591.

Hong, K., Brugada, J., Oliva, A., Berruezo-Sanchez, A., Potenza, D., Pollevick, G.D., Guerchicoff, A., Matsuo, K., Burashnikov, E., Dumaine, R. et al. (2004). Value of electrocardiographic parameters and ajmaline test in the diagnosis of Brugada syndrome cause by SCN5A. *Circulation* 110, 3023–3027.

Hong, K., Bjerregaard, P., Gussak, I. and Brugada, R. (2005). Short QT syndrome and atrial fibrillation caused by mutation in KCNH2. *J Cardiovasc Electrophysiol* 16, 394–396.

Honore, E., Attali, B., Heurteaux, C. et al. (1991). Cloning, expression, pharmacology and regulation of a delayed rectifier K^+ channel in mouse heart. *EMBO* 10, 2805–2811.

Jackman, W.M., Wang, X., Friday, K.J. et al. (1991). Catheter ablation of accessory atrioventricular pathways (Wolff-Parkinson-White syndrome) by radiofrequency current. *N Engl J Med* 324, 1605–1611.

January, C.T. and Riddle, J.M. (1989). Early afterdepolarizations: Mechanism of induction and block: A role for L-type Ca^{2+} current. *Circ Res* 64, 977–990.

Jervell, A. and Lange-Nielsen, (1957). Congenital deaf-mutism, function heart disease with prolongation of the Q-T interval and sudden death. *Am Heart J* 54, 59–68.

Jiang, C., Atkinson, D., Towbin, J.A. et al. (1994). Two long QT syndrome loci map to chromosome 3 and 7 with evidence for further heterogeneity. *Nat Genet* 8, 141–147.

Kaltman, J.R., Ro, P.S., Stephens, P., McBride, M.G., Cohen, M.I., Tanel, R.E., Vetter, V.L. and Rhodes, L.A. (2003). Effects of beta-adrenergic antagonists on the QT measurements from exercise stress tests in pediatric with long QT syndrome. *J Pediatr Cardiol* 24, 553–558.

Kannel, W.B., Wolf, P.A., Benjamin, E.J. and Levy, D. (1998). Prevalence, incidence, prognosis, and predisposing conditions for atrial fibrillation: Population-based estimates. *Am J Cardiol* 82, 2N–9N.

Keating, M.T., Atkinson, D., Dunn, C., Timothy, K., Vincent, G.M. and Leppert, M. (1991a). Linkage of a cardiac arrhythmia, the long QT syndrome, and the Harvey ras-1 gene. *Science* 252, 704–706.

Keating, M.T., Atkinson, D., Dunn, C., Timothy, K., Vincent, G.M. and Leppert, M. (1991b). Consistent linkage of the long QT syndrome to the Harvey ras-1 locus on chromosome 11. *Am J Hum Genet* 49, 1335–1339.

Laitinen, P.T. et al. (2001). Mutations in the cardiac ryanodine receptor (RyR2) gene in familial polymorphic ventricular tachycardia. *Circulation* 103, 485–490.

Lange, P.S., Er, F., Gassanov, N. and Hoppe, U.C. (2003). Andersen mutations of KCNJ2 suppress the native inward rectifier current IK1 in a dominant-negative fashion. *Cardiovasc Res* 59, 321–327.

Leenhardt, A., Lucet, V., Denjoy, I., Grau, F., Ngoc, D.D. and Coumel, P. (1995). Catecholaminergic polymorphic ventricular tachycardia in children. *Circulation* 91, 1512–1519.

Lenegre, J. and Moreau, P.H. (1963). Le block auriculo-ventriculaire chronique. Etude anatomique, clinique et histologieque. *Arch Mal Coeur* 56, 867–888.

Lev, M. (1964). Anatomic basis for atrioventricular block. *Am J Cardiol* 37, 742–748.

Lev, M., Kinare, S.G. and Pick, A. (1970). The pathogenesis of atrioventricular block in coronary disease. *Circulation* 42, 409–425.

Levine, H.D., Wanzer, S.H. and Merrill, J.P. (1956). Dialyzable currents of injury in potassium intoxication resembling acute myocardial infarction or pericarditis. *Circulation* 13, 29–36.

Li, H., Chen, Q., Moss, A.J. et al. (1998). New mutations in the *KVLQT1* potassium channel that cause long QT syndrome. *Circulation* 97, 1264–1269.

Lupoglazoff, J.M., Denjoy, I., Villain, E., Fressart, V., Simon, F., Bozio, A., Berthet, M., Benammar, N., Hainque, B. and Guicheney, P. (2004). Long QT syndrome in neonates: Conduction disorders associated with HERG mutations and sinus bradycardia with KCNQ1 mutations. *J Am Coll Cardiol* 43, 826–830.

Madeiros-Domingo, A., Kaku, T., Tester, D.J., Iturralde-Torres, P., Itty, A., Ye, B., Valdivia, C., Ueda, K., Canizales-Quinteros, S., Tusie-Luna, M.T. et al. (2007). SCN4B-encoded sodium channel β4 subunit in congenital long-QT syndrome. *Circulation* 116, 134–142.

Marks, A.R., Priori, S., Memmi, M., Kontula, K. and Laitinen, P.J. (2002). Involvement of the cardiac ryanodine receptor/calcium release channel in catecholaminergic polymorphic ventricular tachycardia. *J Cell Physiol* 190, 1–6.

Marks, M.L., Trippel, D.L. and Keating, M.T. (1995). Long QT syndrome associated with syndactyly identified in females. *Am J Cardiol* 10, 744–745.

Martini, B., Nava, A., Thiene, G. et al. (1989). Ventricular fibrillation without apparent heart disease. Description of six cases. *Am Heart J* 118, 1203–1209.

McDonald, T.V., Yu, Z., Ming, Z. et al. (1997). A *MinK-HERG* complex regulates the cardiac potassium current I_{Kr}. *Nature* 388, 289–292.

McNair, W.P., Ku, L., Taylor, M.R., Fain, P.R., Dao, D., Wolfel, E. and Mestroni, L. (2004). SCN5A mutation associated with dilated cardiomyopathy, conduction disorder, and arrhythmia. *Circulation* 110, 2163–2167.

Miura, M., Yamagishi, H., Morikawa, Y. and Matsuoka, R. (2003). Congenital long QT syndrome and 2:1 atrioventricular block with a mutation of the SCN5A gene. *Pediatr Cardiol* 24, 70–72.

Mohammad-Panah, R., Demolombe, S., Neyroud, N. et al. (1999). Mutations in a dominant-negative isoform correlate with phenotype in inherited cardiac arrhythmias. *Am J Hum Genet* 64, 1015–1023.

Mohler, P.J., Schott, J.J., Gramolini, A.O. et al. (2003). Ankyrin-β mutation causes type 4 long-QT cardiac arrhythmia and sudden cardiac death. *Nature* 421, 634–639.

Moss, A.J. (1998). Management of patients with the hereditary long QT syndrome. *J Cardiovasc Electrophysiol* 9, 668–674.

Moss, A.J. and McDonald, J. (1970). Unilateral cervicothoracic sympathetic ganglionectomy for the treatment of long QT interval syndrome. *N Engl J Med* 285, 903–904.

Moss, A.J., Liu, J.E., Gottlieb, S., Locati, E., Schwartz, P.J. and Robinson, J.L. (1991). Efficacy of permanent pacing in the management of high-risk patients with long-QT syndrome. *Circulation* 84, 1524–1529.

Moss, A.J., Zareba, W., Benhorin, J. et al. (1995). ECG T-wave patterns in genetically distinct forms of the hereditary long QT syndrome. *Circulation* 92, 2929–2934.

Moss, A.J., Hall, W.J., Cannom, D.S. et al. (1996). For the MADIT investigators: Improved survival with an implantable defibrillator in patients with coronary disease at high-risk of ventricular arrhythmias. *N Engl J Med* 335, 1194–1933.

Moss, A.J., Robinson, J.L., Gessman, L. et al. (1999). Comparison of clinical and genetic variables of cardiac events associated with loud noise vs swimming among subjects with the long QT syndrome. *Am J Cardiol* 84, 876–879.

Moss, A.J., Zareba, W., Hall, W.J. et al. (2000). Effectiveness and limitations of β-blocker therapy in congenital long-QT syndrome. *Circulation* 101, 616–623.

Myerburg, R.J. (1997). Sudden cardiac death in persons with normal (or near normal) heart. *Am J Cardiol* 79, 3–9.

Nademanee, K., Veerakul, G., Nimmannit, S. et al. (1997). Arrhythmogenic marker for the sudden unexplained death syndrome in Thai men. *Circulation* 96, 2595–2600.

Neyroud, N., Tesson, F., Denjoy, I. et al. (1997). A novel mutation on the potassium channel gene *KVLQT1* causes the Jervell and Lange-Nielsen cardioauditory syndrome. *Nat Genet* 15, 186–189.

Oberti, C., Wang, L., Li, L., Dong, J., Rao, S., Du, W. and Wang, Q. (2004). Genome-wide linkage scan identifies a novel genetic locus on chromosome 5p13 for neonatal atrial fibrillation associated with sudden death and variable cardiomyopathy. *Circulation* 110, 3753–3759.

Olson, T.M., Michels, V.V., Ballew, J.D., Reyna, S.P., Karst, M.L., Herron, K.J., Horton, S.C., Rodeheffer, R.J. and Anderson, J.L. (2005). Sodium channel mutations and susceptibility to heart failure and atrial fibrillation. *JAMA* 293, 447–454.

Olson, T.M., Alekseev, A.E., Liu, X.K., Park, S., Zingman, L.V., Bienengraeber, M., Sattiraju, S., Ballew, J.D., Jahangir, A. and Terzic, A. (2006). Kv1.5 channelopathy due to KCNA5 loss-of-function mutation causes human atrial fibrillation. *Hum Mol Genet* 15, 2185–2191.

Orgain, E.S., Wolff, L. and White, P.D. (1936). Uncomplicated auricular fibrillation and flutter: frequent occurrence and good prognosis in patients without other evidence of cardiac disease. *Arch Int Med* 57, 493–513.

Osher, H.L. and Wolff, L. (1953). Electrocardiographic pattern simulating acute myocardial injury. *Am J Med Sci* 226, 541–545.

Otway, R., Vandenberg, J.I., Guo, G., Varghese, A., Castro, M.L., Liu, J., Zhao, J.T., Bursill, J.A., Wyse, K.R., Crotty, H. et al. (2007). Stretch-sensitive KCNQ1 mutation: a link between genetic and environmental factors in the pathogenesis of atrial fibrillation? *J Am Coll Cardiol* 49, 578–586.

Pennefather, P.S., Zhou, W. and Decoursey, T.E. (1998). Idiosyncratic gating of *HERG*-like K+ channels in microglia. *J Gen Physiol* 111, 795–805.

Plaster, N.M., Tawil, R., Tristani-Firouzi, M., Canun, S., Bendahhou, S., Tsunoda, A., Donaldson, M.R., Iannoccone, S.T., Brunt, E., Barohn, R. et al. (2001). Mutations in Kir2.1 cause the developmental and episodic electrical phenotyhpes of Andersen's syndrome. *Cell* 105, 511–519.

Platia, E.V., Griffith, L.S.C., Watkins, L., Mirowski, M., Mower, M.M. and Reid, P.R. (1985). Management of the prolonged QT syndrome and recurrent ventricular fibrillation with an implantable defibrillator. *Clin Cardiol* 8, 490–493.

Poret, P., Mabo, P., Deplace, C. et al. (1996). Isolated atrial fibrillation genetically determined? Apropos of a familial history. *Arch Mal du Coeur Vaiss* 89, 1197–1203.

Postma, A.V., Denjoy, I., Hoorntje, T.M., Lupoglazoff, J.M., Da Costa, A., Sebillon, P., Mannens, M.M., Wilde, A.A. and Guicheney, P. (2002). Absence of calsequestrin 2 causes severe forms of catecholaminergic polymorphic ventricular tachycardia. *Circ Res* 91, e21–e26.

Priori, S.G., Schwartz, P.J., Napolitano, C. et al. (1998). A recessive variant of the Romano-Ward long-QT syndrome. *Circulation* 97, 2420–2425.

Priori, S.G., Barhanin, J., Hauer, R.N.W. et al. (1999a). Genetic and molecular basis of cardiac arrhythmias. Impact on clinical management. Parts I & II. *Circulation* 99, 518–528.

Priori, S.G., Barhanin, J., Hauer, R.N. et al. (1999b). Genetic and molecular basis of cardiac arrhythmias: Impact on clinical management. Part III. *Circulation* 99, 674–681.

Priori, S.G., Napolitano, C. and Schwartz, P.J. (1999c). Low penetrance in the long-QT syndrome. Clinical impact. *Circulation* 99, 529–533.

Priori, S.G. et al. (2001). Mutations in the cardiac ryanodine receptor gene (hRyR2) underlie catecholaminergic polymorphic ventricular tachycardia. *Circulation* 103, 196–200.

Priori, S.G., Napolitano, C., Gasparini, M., Pappone, C., Bella, P.D., Giordano, U., Bloise, R., Giustetto, C., De Nardis, R., Grillo, M. et al. (2002a). Natural history of Brugada syndrome. *Circulation* 105, 1342.

Priori, S.G., Napolitano, C., Memmi, M., Colombi, B., Drago, F., Gasparini, M., DeSimone, L., Coltorti, F., Bloise, R., Keegan, R. et al. (2002b). Clinical and molecular characterization of patients with catecholaminergic polymorphic ventricular tachycardia. *Circulation* 106, 69–74.

Priori, S.G., Napolitano, C., Schwartz, P.J., Grillo, M., Bloise, R., Ronchetti, E., Moncalvo, C., Tulipani, C., Veia, A., Bottelli, G. et al. (2004). Association of long QT syndrome loci and cardiac events among patients treated with beta-blockers. *JAMA* 292, 1341–1344.

Priori, S.G., Pandit, S.V., Rivolta, I., Berenfeld, O., Ronchetti, E., Dhamoon, A., Napolitano, C., Anumonwo, J., Raffaele di Barletta, M., Gudapakkam, S. et al. (2005). A novel form of short QT syndrome (SQT3) is caused by a mutatioins in the KCNJ2 gene. *Circ Res Mar* 10. [Epub ahead of print].

Rolf, S., Bruns, H.J., Wichter, T., Kirchhof, P., Ribbing, M., Wasmer, K., Paul, M., Breithardt, G., Haverkamp, W. and Eckhardt, L. (2003). The ajmaline challenge in Brugada syndrome: Diagnostic impact, safety, and recommended protocol. *Eur Heart J* 24, 1104–1112.

Romano, C., Gemme, G. and Pongiglione, R. (1963). Antmie cardiache rare in eta pediatrica. *Clin Pediatr* 45, 656–683.

Rosenbaum, M.B., Elizari, M.V. and Lazzari, J.O. (1969). The mechanisms of bidirectional tachycardia. *Am Heart J* 78, 4–12.

Rosero, S.Z., Zareba, W., Moss, A.J., Robinson, J.L., Hajj, A.R.H., Locati, E.H., Benhorin, J. and Andrews, M.L. (1999). Asthma and the risk of cardiac events in the long QT syndrome. *Am J Cardiol* 84, 1406–1411.

Rubin, D.A., O'Keefe, A., Kay, R.H., McAllister, A. and Mendelson, D.M. (1992). Autosomal dominant inherited ventricular tachycardia. *Am Heart J* 123, 1082–1108.

Sacks, H.S., Matisonn, R. and Kennelly, B.M. (1978). Familial paroxysmal ventricular tachycardia of unknown etiology. *Proc Royal Soc Med* 68, 305–306.

Sanguinetti, M.C., Jiang, C., Curran, M.E. and Keating, M.T. (1995). A mechanistic link between an inherited and an acquired cardiac arrhythmia: *HERG* encodes the I_{Kr} potassium channel. *Cell* 81, 299–307.

Sanguinetti, M.C., Curran, M.E., Zou, A. et al. (1996a). Coassembly of *KvLQT1* and *minK* (IsK) proteins to form cardiac I_{Ks} potassium channel. *Nature* 384, 80–83.

Sanguinetti, M.C., Curran, M.E., Spector, P.S. and Keating, M.T. (1996b). Spectrum of *HERG* K^+-channel dysfunction in an inherited cardiac arrhythmia. *Proc Natl Acad Sci USA* 93, 2208–2212.

Sangwatanaroj, S., Yanatasneejit, P., Sunsaneewitayakul, B. and Sitthisook, S. (2002). Linkage analyses and SCN5A mutations screening in five sudden unexplained death syndrome (Lai-tai) families. *J Med Assoc Thai* 85(Suppl 1), S54–S61.

Schott, J-J., Alshinawi, C., Kyndt, F. et al. (1999). Cardiac conduction defects associated with mutations in SCN5A. *Nat Genet* 23, 20–21.

Schott, J., Charpentier, F., Peltier, S. et al. (1995). Mapping of a gene for long QT syndrome to chromosome 4q25-27. *Am Hum Genet* 57, 1114–1122.

Schulze-Bahr, E., Wang, Q., Wedekind, H. et al. (1997). *KCNE1* mutations cause Jervell and Lange-Nielsen syndrome. *Nat Genet* 17, 267–268.

Schwartz, P.J. (1985). Idiopathic long QT syndrome. Progress and questions. *Am Heart J* 109, 399–411.

Schwartz, P.J., Locati, E., Moss, A.J., Crampton, R.S., Trazzi, R. and Ruberti, U. (1991). Left cardiac sympathetic denervation in the therapy of the congenital long-QT syndrome: A worldwide report. *Circulation* 84, 503–511.

Schwartz, P.J., Moss, A.J., Vincent, G.M. and Crampton, R.S. (1993). Diagnostic criteria for the long QT syndrome. An update. *Circulation* 88, 782–784.

Schwartz, P.J., Priori, S.G., Locati, E.H., Napolitano, C., Cantu, F., Towbin, J.A., Keating, M.T., Hammoude, H., Brown, A.M. and Chen, L.S. (1995). Long QT syndrome patients with mutations of the SCN5A and HERG genes have differential responses to Na+ channel blockade and to increases in heart rate. Implications for gene-specific therapy. *Circulation* 92, 3381–3386.

Schwartz, P.J., Stramba-Badiale, M., Segantini, A. et al. (1998). Prolongation of the QT interval and the sudden infant death syndrome. *N Engl J Med* 338, 1709–1714.

Schwartz, P.J., Priori, S.G. and Napolitano, C. (2000a). Long QT syndrome. In *Cardiac Electrophysiology: From Cell to Bedside* (D.P. Zipes and J. Jalife, eds), *3rd Edition*. Saunders, Philadelphia, PA, pp. 597–615.

Schwartz, P.J., Priori, S.G., Dumaine, R. et al. (2000b). A molecular link between the sudden infant death syndrome and the long-QT syndrome. *N Engl J Med* 343, 262–267.

Schwartz, P.J., Priori, S.G., Cerrone, M., Spazzolini, C., Odero, A., Napolitano, C., Bloise, R., De Ferrari, G.M., Klersy, C., Moss, A.J. et al. (2004). Left cardiac sympathetic denervation in the management of high-risk patients affected by the long-QT syndrome. *Circulation* 109, 1826–1833.

Sebillon, P., Boucher, C., Bidot, L.D., Bonne, G., Ahamed, K., Charron, P., Drouin-Garraud, V., Millaire, A., Desrumeaux, G., Benaiche, A. et al. (2003). Expanding the phenotype of LMNA mutations in dilated cardiomyopathy and functional consequences of these mutations. *J Med Genet* 40, 560–567.

Shalaby, F.Y., Levesque, P.C., Yang, W.P., Little, W.A., Conder, M.L., Jenkins-West, T. and Blanar, M.A. (1997). Dominant-negative KvLQT1 mutations underlie the LQT1 form of long QT syndrome. *Circulation* 96, 1733–1736.

Shimizu, W., Kurita, T., Matsuo, K., Suyama, K., Aihara, N., Kamakura, S., Towbin, J.A. and Shimomura, K. (1998). Improvement of repolarization abnormalities by a K+ channel opener in the LQT1 form of congenital long-QT syndrome. *Circulation* 97, 1581–1588.

Shimizu, W., Horie, M., Ohno, S., Takenaka, K., Yamaguchi, M., Shimizu, M., Washizuka, T., Aizawa, Y., Nakamura, K., Ohe, T. et al. (2004). Mutation site-specific differences in arrhythmic risk and sensitivity to sympathetic stimulation in the LQT1 form of congenital long QT syndrome: Multicenter study in Japan. *J Am Coll Cardiol* 44, 117–125.

Shirai, N., Makita, N., Sasaki, K., Yokoi, H., Sakuma, I., Sakurada, H., Akai, J., Kimura, A., Hiraoka, M. and Kitabatake, A. (2002). A mutant cardiac sodium channel with multiple biophysical defects associated with overlapping clinical features of Brugada syndrome and cardiac conduction disease. *Cardiovasc Res* 53, 348–354.

Sopontammarak, S., Khongphatthanayothin, A. and Sa-Nguanchua, P. (2003). Prevalence of idiopathic long QT syndrome in congenital sensori-neural hearing loss students of Songkhla School for the Deaf. *J Med Assoc Thai* 86, 1149–1155.

Spellberg, R.D. (1971). Familial sinus node disease. *Chest* 60, 246–251.

Splawski, I., Shen, J., Timothy, K.W., Lehman, M.H., Priori, S., Robinson, J.L., Moss, A.J., Schwartz, P.J., Towbin, J.A., Vincent, G.M. and Keating, M.T. (2000). Spectrum of mutations in long-QT syndrome genes KVLQT1, HERG, SCN5A, KCNE1, and KCNE2. *Circulation* 102, 1178–1185.

Splawski, I., Timothy, K.W., Vincent, G.M., Atkinson, D.L. and Keating, M.T. (1997a). Brief report: Molecular basis of the long-QT syndrome associated with deafness. *N Engl J Med* 336, 1562–1567.

Splawski, I., Tristani-Firouzi, M., Lehmann, M.H., Sanguinetti, M.C. and Keating, M.T. (1997b). Mutations in the *minK* gene cause long QT syndrome and suppress I_{Ks} function. *Nat Genet* 17, 338–340.

Sumitomo, N., Harada, K., Nagashima, M., Yasuda, T., Nakamura, Y., Aragaki, Y., Saito, A., Kurosaki, K., Jouo, K., Koujiro, M. et al. (2003). Catecholaminergic polymorphic ventricular tachycardia: Electrocardiographic characteristics and optimal therapeutic strategies to prevent sudden death. *Heart* 89, 66–70.

Surawicz, B. and Hariman, R.J. (1988). Follow-up of the family with congenital absence of sinus rhythm. *Am J Cardiol* 61, 467–469.

Swan, H., Piippo, K., Viitasalo, M. et al. (1999). Arrhythmic disorder mapped to chromosome 1q42-q43 causes malignant polymorphic ventricular tachycardia in structurally normal hearts. *J Am Coll Cardiol* 34, 2035–2042.

Ten Harkel, A.D., Witsenburg, M., de Jong, P.L., Jordaens, L., Wijman, M. and Wilde, A.A. (2005). Efficacy of an implantable cardioverter-defibrillator in a neonate with LQT3 associated arrhythmias. *Europace* 7, 77–84.

The Boston Area Anticoagulation trial for Atrial Fibrillation Investigators (1990). The effect of low-dose warfarin on the risk of stroke in patients with nonrheumatic atrial fibrillation. *N Engl J Med* 323, 1505–1511.

Tikanoja, T., Kirkinen, P., Nikolajev, K., Eresmaa, L. and Haring, P. (1998). Familial atrial fibrillation with fetal onset. *Heart* 79, 195–197.

Towbin, J.A. (1998). The role of cytoskeletal proteins in cardiomyopathies. *Curr Opin Cell Biol* 10, 131–139.

Towbin, J.A. (2000). Cardiac arrhythmias: The genetic connection. *J Cardiovasc Electrophysiol* 11, 601–602.

Towbin, J.A. (2001a). Early repolarization syndrome and the Brugada syndrome: Forme frustes?. *Eur Heart J* 22, 448–449.

Towbin, J.A. (2001b). Ventricular tachycardia or conduction disease: What is the mechanism of death associated with SCN5A? *J Cardiovasc Electrophysiol* 12, 637–638.

Towbin, J.A. (2006). Mutation screening for the genes causing cardiac arrhythmias. *Methods Mol Med* 126, 57–59.

Towbin, J.A. and Ackerman, M.J. (2001). Cardiac sodium channel gene mutations and sudden infant death syndrome Confirmation of froof of concept? *Circulation* 104, 1092–1093.

Towbin, J.A., Pagotto, L., Siu, B. et al. (1992). Romano-Ward long QT syndrome (RWLQTS): Evidence of genetic heterogeneity. *Pediatr Res* 31, 23A.

Towbin, J.A., Li, H., Taggart, T. et al. (1994). Evidence of genetic heterogeneity in Romano-Ward long QT syndrome: Analysis of 23 families. *Circulation* 90, 2635–2644.

Tristani-Firouzi, M., Jensen, J.L., Donaldson, M.R., Sansone, V., Meola, G., Hahn, A., Bendahhou, S., Kwiecinski, H., Fidzianska, A., Plaster, N. et al. (2002). Functional and clinical characterization of KCNJ2 mutations associated with LQT7 (Andersen syndrome). *J Clin Invest* 222, 381–388.

Trudeau, M.C., Warmke, J., Ganetzky, B. and Robertson, G. (1995). *HERG*, a human inward rectifier in the voltage-gated potassium channel family. *Science* 269, 92–95.

Tyson, J., Transbjaerg, L., Bellman, S. et al. (1997). IsK and KVLQT1: Mutation in either of the two subunits of the slow component or the delayed rectifier potassium channel can cause Jervell and Lange-Nielsen syndrome. *Hum Mol Genet* 6, 2179–2185.

Vatta, M. and Towbin, J.A. (2006). Mutations in KCNE1 in long QT syndrome (LQTS): Insights into mechanism of LQTS and drug sensitivity? *Heart Rhythm* 3, 1041–1043.

Vatta, M., Li, H. and Towbin, J.A. (2000). Molecular biology of arrhythmic syndromes. *Curr Opin Cardiol* 15, 12–22.

Vatta, M., Ackerman, M.J., Ye, B., Makielski, J.C., Ughanze, E.E., Taylor, E.W., Tester, D.J., Balijepalli, R.C., Foell, J.D., Li, Z. et al. (2006). Mutant caveolin-3 induces persistent late sodium current and is associated with long-QT syndrome. *Circulation* 114, 2104–2112.

Verduyn, S.C., Vos, M.A., Vanderzande, J. et al. (1997). Further observations to elucidate the role of interventricular dispersion of repolarization and early after depolarizations in the genesis of acquired *torsades de pointes* arrhythmias. A comparison between almokalant and d-sotalol using the dog as its own control. *J Am Coll Cardiol* 30, 1575–1584.

Vetter, D.E., Mann, J.R., Wangemann, P. et al. (1996). Inner ear defects induced by null mutation of the IsK gene. *Neuron* 17, 1251–1264.

Villain, E., Denjoy, I., Lupoglazoff, J.M., Guicheney, P., Hainque, B., Lucet, V. and Bonnet, D. (2004). Low incidence of cardiac events with beta-blocking therapy in children with long QT syndrome. *Eur Heart J* 25, 1405–1411.

Viskin, S. (2000). Cardiac pacing in the long QT syndrome: review of available data and practical recommendations. *J Cardiovasc Electrophysiol* 11, 593–600.

Viskin, S., Alla, S.R., Barron, H.V. et al. (1996). Mode of onset of *torsade de pointes* in congenital long QT syndrome. *J Am Coll Cardiol* 28, 1262–1268.

Viswanathan, P.C., Benson, D.W. and Balser, J.R. (2003). A common *SCN5A* polymorphism modulates the biophysical effects of an *SCN5A* mutation. *J Clin Invest* 111, 341–346.

Vlay, S.C. (1987). Catecholamine-sensitive ventricular tachycardia. *Am Heart J* 114, 455–461.

Wang, D.W., Viswanathan, P.C., Balser, J.R., George, A.L. Jr. and Benson, D.W. (2002). Clinical, genetic, and biophysical characterization of *SCN5A* mutations associated with atrioventricular conduction block. *Circulation* 105, 341.

Wang, T.J., Larson, M.G., Levy, D., Vasan, R.S., Leip, E.P., Wolf, P.A., D'Agostino, R.B., Murabito, J.M., Kannel, W.B. and Benjamin, E.J. (2003). Temporal relations of atrial fibrillation and congestive heart failure and their joint influence on mortality: The Framingham Heart Study. *Circulation* 107, 2920–2925.

Wang, Q., Shen, J., Splawski, I. et al. (1995a). *SCN5A* mutations associated with an inherited cardiac arrhythmia, long QT syndrome. *Cell* 80, 805–811.

Wang, Q., Shen, J., Li, Z. et al. (1995b). Cardiac sodium channel mutations in patients with long QT syndrome, an inherited cardiac arrhythmia. *Hum Mol Genet* 4, 1603–1607.

Wang, Q., Curran, M.E., Splawski, I. et al. (1996). Positional cloning of a novel potassium channel gene: *KVLQT1* mutations cause cardiac arrhythmias. *Nat Genet* 12, 17–23.

Ward, D.E., Ho, S.Y. and Shinebourne, E.A. (1984). Familial atrial standstill and inexcitability in childhood. *Am J Cardiol* 53, 965–967.

Ward, O.C. (1964). A new familial cardiac syndrome in children. *J Irish Med Assoc* 54, 103–106.

Warmke, J.E. and Ganetzky, B. (1994). A family of potassium channel genes related to *eag* in *Drosophila* and mammals. *Proc Natl Acad Sci USA* 91, 3438–3442.

Waxman, M.B., Catching, J.D., Felderhof, C.H., Downar, E., Silver, M.D. and Abbott, M.M. (1975). Familial atrioventricular heart block; an autosomal dominant trait. *Circulation* 51, 226–233.

Weiss, J.N., Qu, Z., Chen, P.S., Lin, S.F., Karagueuzian, H.S., Hayashi, H., Garfinkel, A. and Karma, A. (2005). The dynamics of cardiac fibrillation. *Circulation* 112, 1232–1240.

Westenskow, P., Splawski, I., Timothy, K.W., Keating, M.T. and Sanguinetti, M.C. (2004). Compound mutations: A common cause of severe long QT. *Circulation* 109, 1834–1841.

Wilde, A.A., Jongbloood, R.J., Doevendans, P.A. et al. (1999). Auditory stimuli as a trigger for arrhythmic events differentiate HERG-related (LQTS2) patients from KVLQT1-related patients (LQTS1). *J Am Coll Cardiol* 33, 327–332.

Wilde, A.A., Antzelevitch, C., Borggrefe, M., Brugada, J., Brugada, R., Brugada, P., Corrado, D., Hauer, R.N., Kass, R.S., Nademanee, K. et al. (2002a). Study group on the molecular basis of arrhythmias of the European Society of Cardiology. *Eur Heart J* 24, 2072.

Wilde, A.A.M., Antzelevitch, C., Borggrefe, M., Brugada, J., Brugada, R., Brugada, P., Corrado, D., Hauer, R.N.W., Kass, R.S., Nademanee, K. et al. (2002b). Proposed diagnosic criteria for the Brugada syndrome. *Circulation* 106, 2514.

Wilde, A.A., Antzelevitch, C., Borggrefe, M., Brugada, J., Brugada, R., Brugada, P., Corrado, D., Hauer, R.N.W., Kass, R.S., Nademanee, K., Priori, S.G., Towbin, J.A.: Study Group on the Molecular Basis of Arrhythmias of the European Society of Cardiology (2003). Proposed diagnostic criteria for the Brugada syndrome. *Euro Heart J* 23, 1648–1654.

Wolff, L. (1943). Familial auricular fibrillation. *N Engl J Med* 229, 396–397.

Wolf, P.A., Abbott, R.D. and Kannel, W.B. (1991). Atrial fibrillation as an independent risk factor for stroke: The Framingham Heart Study. *Stroke* 22, 983–988.

Wollnick, B., Schreeder, B.C., Kubish, C., Esperer, H.D., Wieacker, P. and Jensch, T.J. (1997). Pathophysiological mechanisms of dominant and recessive *KVLQTI* K^+ channel mutations found in inherited cardiac arrhythmias. *Hum Mol Genet* 6, 1943–1949.

Wolpert, C., Schimpf, R., Giustetto, C., Antzelevitch, C., Cordeiro, J., Dumaine, R., Brugada, R., Hong, K., Bauersfeld, U., GAita, F. et al. (2005). Further insights into the effect of quinidine in short QT syndrome caused by a mutation in HERG. *J Cardiovasc Electrophysiol* 16, 54–58.

Wren, C., Rowland, E., Burn, J. and Campbell, R.W.F. (1990). Familial ventricular tachycardia: A report of four families. *Br Heart J* 63, 169–174.

Xia, M., Jin, Q., Bendahhou, S., He, Y., Larroque, M.M., Chen, Y., Zhou, Q., Yang, X., Liu, Y., Liu, B. et al. (2005). A Kir2.1 gain-of-function mutation underlies familial atrial fibrillation. *Biochem Biophys Res Comm* 332, 1012–1019.

Yang, Y., Xia, M., Jin, Q., Bendahhou, S., Shi, J., Chen, Y., Liang, B., Lin, J., Liu, Y., Liu, B. et al. (2004). Identification of a KCNE2 gain-of-function mutation in patients with familial atrial fibrillation. *Am J Hum Genet* 75, 899–905.

Zareba, W., Moss, A.J., Schwartz, P.J. et al. (1998). Influence of the genotype on the clinical course of the long-QT syndrome. *N Engl J Med* 339, 960–965.

Zareba, W., Moss, A.J., Sheu, G., Kaufman, E.S., Priori, S., Vincent, G.M., Towbin, J.A., Benjorin, J., Schwartz, P.J., Napolitano, C. et al. (2003a). Location of mutation in the KCNQ1 and phenotypic presentation of long QT syndrome. *J Cardiovasc Electrophysiol* 14, 1149–1153.

Zareba, W., Moss, A.J., Locati, E.H., Lehmann, M.H., Peterson, D.R., Hall, W.J., Schwartz, P.J., Vincent, G.M., Priori, S.G., Benhorin, J. et al. (2003b). International Long QT Syndrome Registry. Modulating effects of age and gender on the clinical course of long QT syndrome by genotype. *J Am Coll Cardiol* 42, 103–109.

Zareba, W., Moss, A.J., Daubert, J.P., Hlal, W.J., Robinson, J.L. and Andrews, M. (2003c). Implantable cardioverter defibrillator in high-risk long QT syndrome patients. *J Cardiovasc Electrophysiol* 14, 337–341.

Zhang, D., Liang, B., Lin, J., Liu, B., Zhou, Q.S. and Yang, Y.Q. (2005). KCNE3 R53H substitution in familial atrial fibrillation. *Chin Med J* 118, 1735–1738.

Zhang, L., Vincent, G.M., Baralle, M., Baralle, F.E., Anson, B.D., Benson, D.W., Whiting, B., Timothy, K.W., Carlquist, J., January, C.T. et al. (2004). An intronic mutation causes long QT syndrome. *J Am Coll Cardiol* 44, 1283–1291.

Zhou, Z., Gong, Q., Miles, L. et al. (1998). *HERG* channel dysfunction in human long QT syndrome. *J Biol Chem* 263, 21061–22106.

CHAPTER

Genetics and Genomics in the Management of Hemostasis and Thrombosis

Richard C. Becker and Felicita Andreotti

INTRODUCTION

Blood coagulation is a cell surface, biochemical event designed not only to stem the loss of blood following vascular injury (hemostasis), but also to provide the necessary molecular, cellular, and protein constituents for growth and repair as well. In addition, coagulation occurring within medium-sized arteries and veins can have detrimental effects, ranging from end-organ damage to death. Conceptually, blood coagulation represents a complex, yet well-coordinated series of events that involve tissue factor-bearing cells and platelets. The initiation and propagation phases of coagulation, under biological conditions are catalyzed by thrombin in small (nM) and large concentrations, respectively (Figures 30.1 and 30.2) (Monroe and Hoffnan, 2006).

The complex catalysts that participate in tissue factor-mediated thrombin generation each consist of a serine protease interacting with a receptor and/or cofactor protein, which collectively are anchored to a specific cellular surface. One must recognize, however, that coagulation is regulated tightly by stoichiometric and dynamic systems of inhibition. Specifically, the tissue factor concentration threshold for thrombin generation is steep, and the resulting product is dependent largely on the concentration of plasma proteins and surface inhibition, including tissue factor pathway inhibitor and antithrombin III. Thus, the blood coagulation proteome can be studied in biochemically quantifiable terms (Mann et al., 2006).

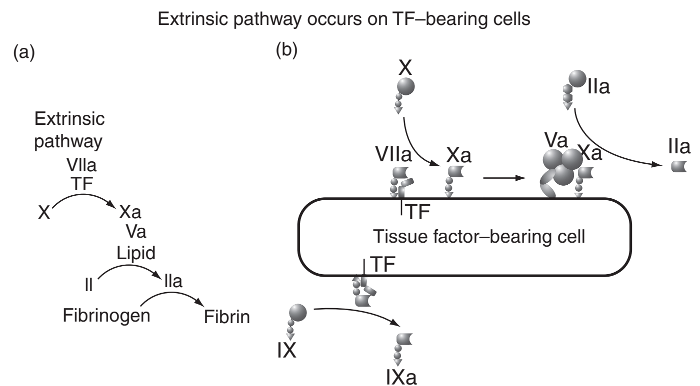


Figure 30.1 Traditional (a) and contemporary (b) paradigms of coagulation highlighting both the biochemical and cell-based components required for thrombin generation. From Monroe and Hoffman (2006).

GENETICS OF COAGULATION

The complex network of integrated biochemical events regulating mammalian coagulation comprises, in essence, five proteases

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

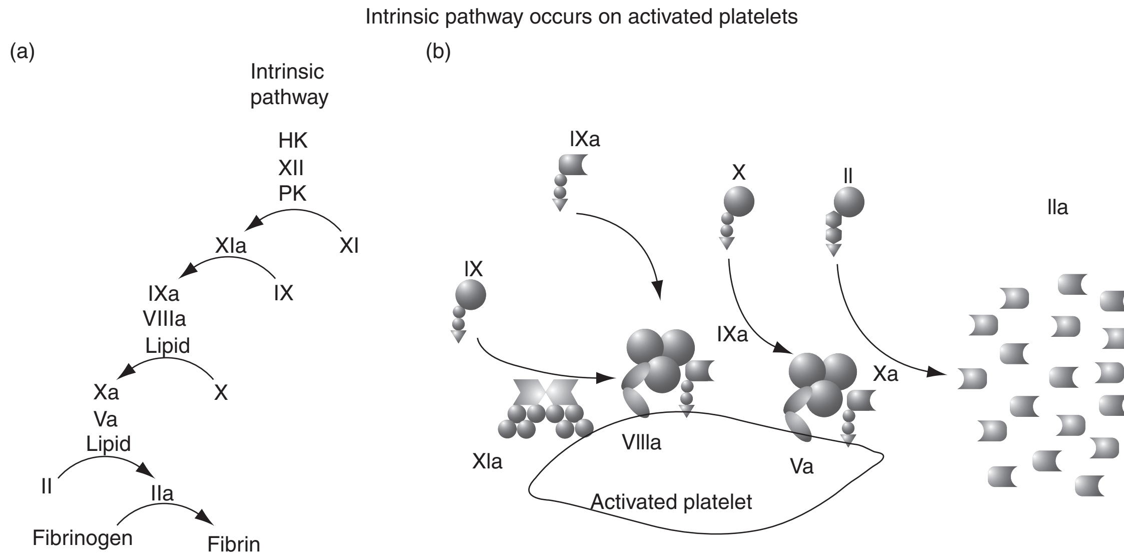


Figure 30.2 The assembly of coagulation proteins on activated platelets leads to a burst of thrombin (IIa) generation – a prerequisite event in thrombus growth and propagations, and fibrin formation. From Monroe and Hoffman (2006).

(fII or prothrombin, f VII, f IX, f X, and protein C) that interact with five cofactors (tissue factor, f VIII, f V, thrombomodulin, and membrane proteins) to generate fibrin (Davidson et al., 2003b). Although each component of the network has unique functional properties, data derived from gene organizations, protein structure, and sequence analysis suggest that coagulation-hemostatic regulatory proteins may have emerged more than 400 million years ago from duplication and diversification of only two gene structures: a vitamin K–dependent serine protease, composed of a γ-carboxylated glutamic acid-epidermal growth factor (EGF)–like domain structure (common to fVII, fIX, fX, and protein C), and the A1-A2-B-A3-C1-C2 domain structure common to f V and f VIII. Prothrombin is also a vitamin K–dependent serine protease; however, it contains kringle domains rather than EGF domains, suggesting a replacement during gene duplication and exon shuffling. Analysis of active-site amino acid residues reveals distinguishing characteristics of thrombin from other serine proteases, supporting its position as the ancestral blood enzyme (Figures 30.3 and 30.4) (Davidson et al., 2003a, b; Krem and Di Cera, 2002; McLysaght et al., 2002; Van Hylckama Vlieg et al., 2003).

There is evidence that large regional or genome duplications have contributed to the overarching structure of mammalian coagulation genomes. Similarly, local duplication and translocation could have contributed to the evaluation of multigene families on separate chromosomal regions (Abi-Rached et al., 2002).

The evolution of complex mammalian coagulation pathways from invertebrate and early vertebrate species is perhaps best illustrated in Zebrafish, whose cDNA/gene orthologues for major coagulant, anticoagulant and fibrinolytic proteins bare striking homology to mammalian sequences (Hanumanthaiah et al., 2002). The only difference, which requires further investigation, is a f VII-like gene in Zebrafish that clusters with f VII and f X genes and functions as a inhibitor (f VIIi) of coagulation. Interestingly, the fVII, fVIIi, and f X gene cluster is homologous to a Drosophilia trypsin gene cluster, supporting a rapid path of evolution from invertebrates to vertebrates, and ultimately, mammals.

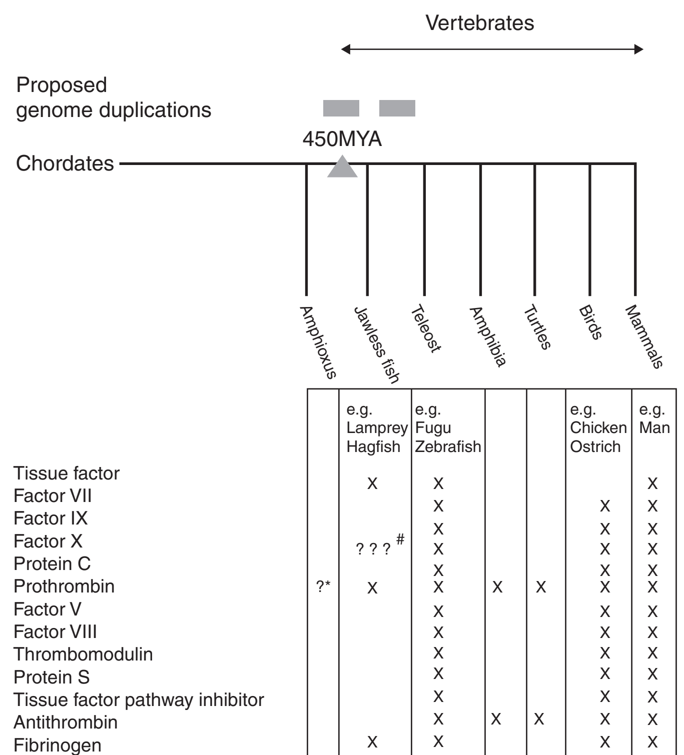


Figure 30.3 Chordate phylogenetic tree illustrating proposed genome duplication and the complement of coagulation proteins identified to date. Reprinted, with permission, from Davidson et al. (2003a, b).

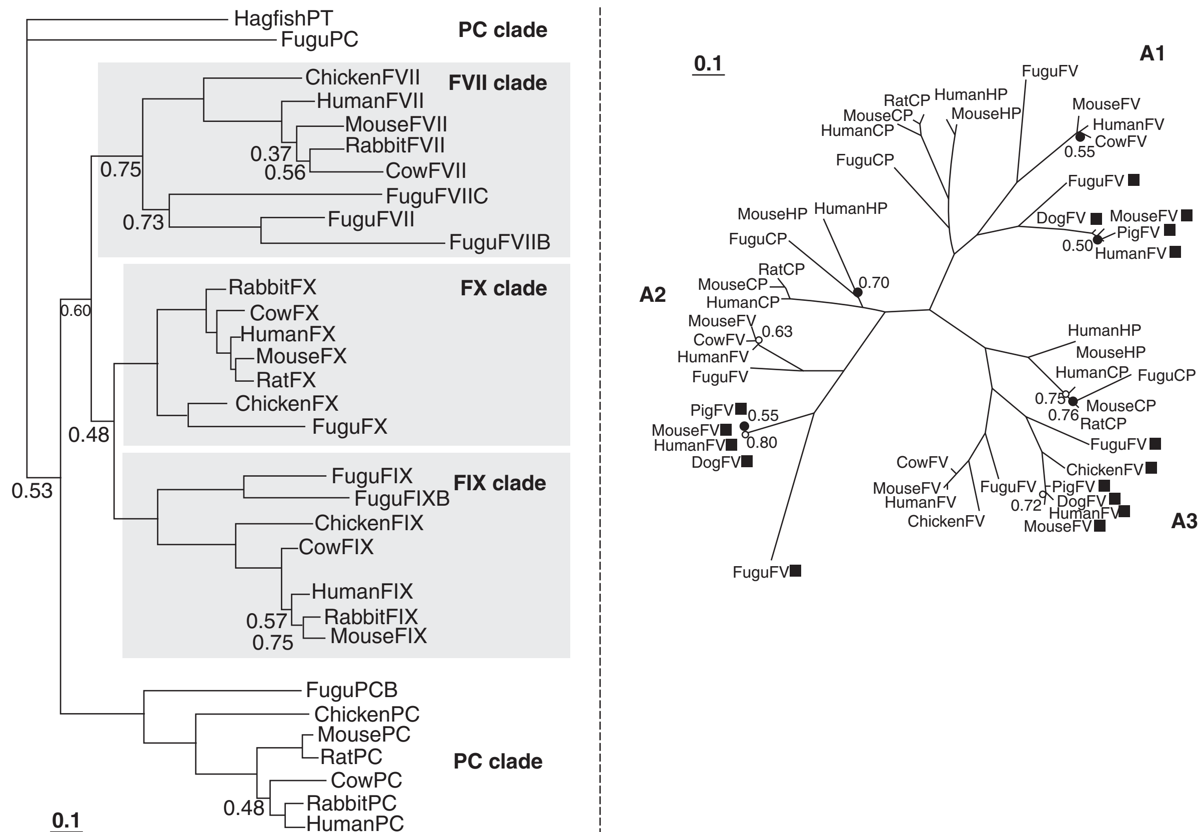


Figure 30.4 Consensus tree illustrating the phylogenetic relationship of vitamin K–dependent proteases and A domain–containing proteins. The number at each internal node represents the posterior probability that the taxa in the corresponding subtree form a clade in the recovered consensus tree. Reprinted, with permission, from Davidson et al. (2003a, b).

HUMAN HEMOSTATIC VARIABILITY

de Lange and colleagues (de Lange et al., 2001) performed a class twin study, including 1002 female twins, 149 pairs of monozygotic and 352 pairs of dizygotic twins. Quantitative genetic model fitting revealed that genetic factors were responsible for 41–75% of the variation in fibrinogen, factor VII, factor VIII, plasminogen activator, factor XIII A-subunit and B-subunit, and von Willebrand factor (vWF). Factor XIII activity showed higher (82%) and factor XII lower (38%) heritability. A higher monozygotic than dizygotic twin correlation was seen for all factors, supporting further the influence of genetics on hemostatic proteins.

Because the genetic contribution to cardiovascular disease and related thrombotic phenotypes decreases with advancing age, the heritability of hemostatic proteins was determined in 130 monozygotic and 155 dizygotic same-sex twin pairs (ages 73–94 years) participating in the Longitudinal Study of Aging of Danish Twins. Genetic factors accounted for 33% (D-dimer) to 71% (Thrombin Activatable Fibrinolysis Inhibitor). In a linage analysis, polymorphisms explained a very small proportion of genetic variations in hemostatic variables (Bladbjerg et al., 2006). Thus, age has a modest effect on hemostatic proteins, explaining anywhere from 1.5% to 14.5% of the variance in plasma concentrations (de Lange et al., 2001).

GENOTYPE–PHENOTYPE INFLUENCES

The work of Rosendaal and colleagues (Van Hylckama Vlieg et al., 2003) underscores the diversity of hemostasis as a biologic system. The structural homologies among coagulation factors suggests similarities in their biosynthesis (transcription, posttranslational processing) and in the mechanisms that govern their plasma levels and clearance. In a group of healthy subjects participating in the Leiden Thrombophilia Study (LETS) (Van Hylckama Vlieg et al., 2003), clustering was found among the plasma concentrations of vitamin K–dependent factors (II, VII, IX, and X) and those of factors XI and XII. Factors V and VIII clustered with fibrinogen and D-dimer (a measure of fibrin formation and its subsequent degradation). The anticoagulant factors (protein C, protein S, and antithrombin III) clustered together, whereas f XIII remained independent. The identification of several independent clusters within the group of procoagulant and anticoagulant factors suggests that the basis for individual fluctuations in plasma levels may lie outside the genes coding for these factors. This example highlights the complexity

of linking genotypes to phenotypes and the relevance of understanding the regulation of gene expression (in particular, the role played by environmental factors) and posttranslational protein modifications (Romualdi et al., 2003).

Completion of the human genome sequence, although one of the most important achievements in science, medicine, and human biology, highlights the importance of proteins in determining biological function and disease (reviewed in reference [Loscalzo, 2003]). Indeed, gene sequences cannot fully predict the complex quaternary structure of proteins (nor their function or dysfunction). The regulation of proteins, in fact, escapes genetic control through posttranslational modifications and protein–protein interactions. Proteomics is defined as the identification and functional characterization of complete sets of proteins expressed by complex biological systems.

Posttranslational modifications of proteins, including proteolysis, oxidation, phosphorylation, nitrosylation, glycation, and sulfuration may well participate in the development and clinical expression of thrombophilias. Many of these modifications are the consequence of environmental influences on gene expression (reviewed in reference [Loscalzo, 2003]). Examples include the correlation between the oxidative state of plasma proteins (determined by measurement of carbonyl content) and several markers of thrombin generation and activity (De Cristofaro et al., 2002); oxidation of thrombomodulin at methionine 388, which reduces its capacity to down-regulate coagulation (through activated protein C), but not its capacity to inhibit fibrinolysis [through Thrombin Activatable Fibrinolysis Inhibitor (Nesheim, 2001)]; oxidation of low-density lipoprotein, a process that involves modification of the amino acid side chain of apoprotein B, which changes the protein moiety's identity, rendering it prothrombotic (Vlassara et al., 1995); nonenzymatic glycation of protein and lipid macromolecules and their condensation to advanced glycation end products (AGEs), which promote atherothrombosis (Loscalzo, 2003); and, the mixed disulfide-bond formation between cysteinyl side chains of proteins and homocysteine, which may induce protein modifications that favor vascular injury and thrombosis (Van Hylckama Vlieg et al., 2003).

The Framingham investigators (Wang et al., 2006) measured 10 biomarkers in 3209 participants attending a routine examination cycle of the Framingham Heart Study. During a median follow-up of 7.4 years, 207 participants died and 169 had a first major cardiovascular event. Neither fibrinogen, D-dimer or plasminogen activator inhibitor type 1 strongly predicted death or major cardiovascular events by Cox proportional hazards models adjusted for conventional risk factors, suggesting that protein modifications, to include their metabolic byproducts, may represent more accurately, pathobiological events and provide more direct links between genotype and phenotype.

TABLE 30.1 Variants of fibrinogen and factor XIII

Variants	Function	Relation to disease
Splice variants		
Fibrinogen γ′	Binds thrombin and f XIII; reduces fiber diameter	Increases risk for MI
Fibrinogen $A\alpha_E C$	Unknown	None?
Noncoding polymorphisms		
Fibrinogen Aα Taq1	–	–
Fibrinogen Bβ Bcl1	Increases fibrinogen level	More common in CAD
Fibrinogen Bβ-148C/T	–	–
Fibrinogen Bβ-448G/A	Alters nuclear protein binding	More common in CAD
Coding polymorphisms		
Fibrinogen Aα Thr312Ala	Changes fibrin structure/ function and FXIII cross-linking	Atrial fibrillation/ pulmonary embolism
Fibrinogen BβArg448Lys	Changes fibrin structure/function	Macrovascular disease
Factor XIIIA Val34Leu	Changes FXIII activation rate and fibrin structure/function	MI, Deep vein thrombosis (DVT)
Factor XIIIB His95Arg	Unknown	MI

From Scott et al. (2004).
MI = myocardial infarction; CAD = coronary artery disease.

GENE-ENVIRONMENT INFLUENCES ON HEMOSTASIS

Fibrin, as summarized previously, is the predominant protein constituent of blood clots formed from fibrinogen, a large glycoprotein present within the circulation. Because fibrin clot architecture plays an important role in hemostasis and vascular repair, variations have important clinical implications. In turn, variability in fibrin strand width, branch points, mass-to-length ratio, density and cross-linking among healthy individuals, as well as those with atherosclerosis, metabolic disorders, and other prothrombotic disease states, support both genetic and environmental influences (Table 30.1) (Scott et al., 2004). Alterations in gene expression and coding function, splice variants, and posttranslational modifications each influence fibrin structure and functionality.

Vascular Bed–Specific Hemostasis and Thrombosis

A well-recognized feature of thrombophilias, whether inherited or acquired (or both), is the focal and vascular bed–specific

TABLE 30.2 Vascular bed–specific thrombosis

	Organ bed				
	Heart	**Lung**	**Spleen**	**Kidney**	**Brain**
Clinical disorders/ associated syndromes	Myocardial infarction Acute coronary syndromes	Pulmonary hypertension Pulmonary embolism/infarction	Splenic infarction	Nephropathy Renal thrombosis	Stroke
Clinically associated factors	Activated factor IX peptide, factor XIIa Fibrinogen Plasminogen activator inhibitor-1 Tissue-type plasminogen activator antigen von Willebrand factor	Fibrinopeptide A Hereditary spherocytosis Soluble thrombomodulin	Protein C Sickle cell anemia	Factor V Leiden Homocysteine Plasminogen activator inhibitor-1 4G/4G	Antithrombin III Factor V Leiden Homocysteinuria

From Edelberg et al. (2001) with permission.

expression of thrombosis. This distinct feature suggests that local regulatory pathways and intrinsic vessel-determined responses to prothrombotic stimuli are distinguishing factors (Mackman, 2005). Moreover, from the clinical perspective, it argues strongly in favor of a patient/vascular bed–specific approach to screening, diagnostic testing, and management rather than a more generalized strategy (Table 30.2) (Edelberg et al., 2001).

The regulation of platelet–vessel wall interactions, coagulation proteases, and fibrinolytic factors takes place on the endothelial surface, suggesting strongly that overall hemostatic regulation and the heterogeneity of clinical expression observed under conditions favoring thrombosis, is based on site-specific differences in endothelial cell structure, functionality, and molecular responsiveness to both biologic and rheologic conditions (Bavendiek et al., 2002; Braddock et al., 1998; Chien et al., 1998; Lin et al., 1997; Sokabe et al., 2004).

Endothelial cells derived from venous, arterial, and microvascular beds exhibit distinct phenotypes in mitosis rates (Beekhuizen and van Furth, 1994), growth responses (Rupnick et al., 1988), signaling pathways (Chang et al., 2000), and expression of nitric oxide synthase (Guillot et al., 2000), vWF (Edelberg et al., 1998), and tissue-type plasminogen activator (Christie, 1999; Rosenberg, 1999). In addition, the functional relationship between vascular endothelial cells and surrounding tissues (cortex, myocardium, myometrium, connective tissue/skeletal muscle) and hemostatic regulation, may be particularly important in "site-specific" responses to prothrombotic stimuli (Christie et al., 1999; Le Flem et al., 1999; Nishida et al., 1993; Tabrizi et al., 1999). An area of considerable interest and relevance to the subject of thrombophilias and vascular bed–specific thrombotic potential is endothelial cell repair and death (apoptosis). An ability to restore endothelial integrity (structural and functional) following injury and the prompt regulation of apoptotic cell–mediated prothrombotic activity may, in fact, be the most critical part of the biologic equation (Figure 30.5).

The clear distinction between vascular beds in response to injury, regulation of coagulation and perhaps thromboprophylaxis with antithrombotic drugs has implications for investigation of medical genomics in general, and the impact of disease predisposition attributable to traditional risk factors in particular. Indeed, the major risk factors for clinical atherosclerosis, including dyslipidemia, hypertension, glucose intolerance, adiposity, cigarette smoking, sedentary life style, inflammatory markers and, to a lesser degree, hemostatic factors accounted for 85% of the cardiovascular disease arising within the Framingham population (Kannel and Wolf, 2006).

CIRCULATING CELLULAR AND PROTEIN INFLUENCES ON HEMOSTASIS AND THROMBOSIS

A traditionally held view that thrombosis, particularly involving the arterial circulatory system, is governed solely by factors intrinsic to the vessel wall at sites of injury requires reconsideration. Tissue factor is found in high concentrations within atherosclerotic plaques, activated endothelial cells, fibroblasts, macrophages, and vascular smooth muscle cells; however, tissue factor antigen is also present within the circulating blood of patients with coronary artery disease or with hematologic disorders characterized by heightened thrombogenicity (Falciani et al., 1998). Tissue factor–containing neutrophils, monocytes, and microparticles (Giesen et al., 1999), circulating in peripheral blood, can be delivered to sites of vascular injury where they contribute directly to both the initiation of thrombus formation and its subsequent propagation. This evolving construct emphasizes an important interface between leukocytes and activated endothelial cells, activated platelets, and the vessel wall, where essential substrate for thrombin generation already exists (Giesen et al., 1999; McEver, 2001). Adherent leukocytes enhance fibrin deposition by

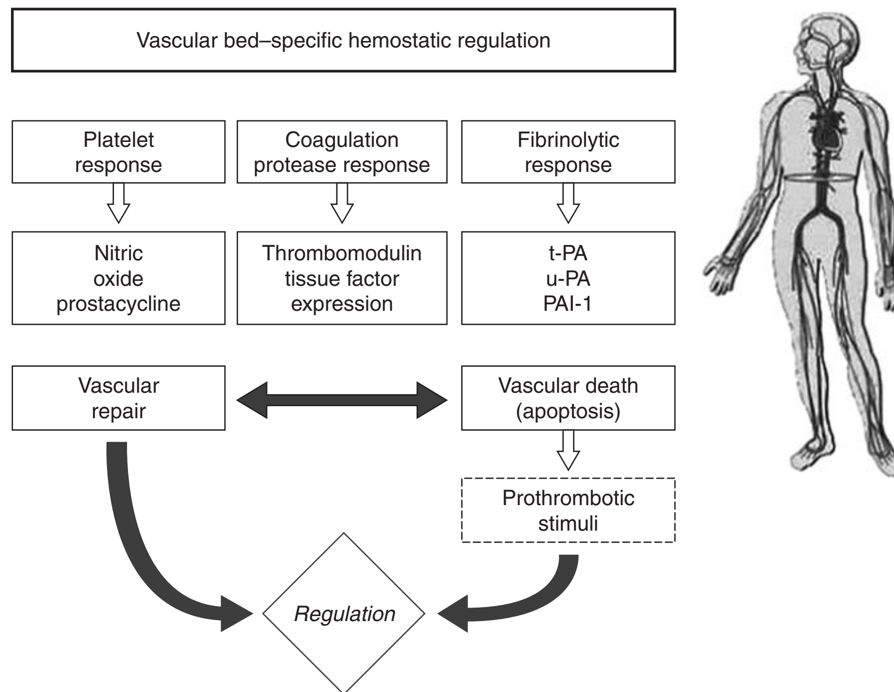


Figure 30.5 Vascular bed–specific hemostasis and thrombosis is regulated by endothelial cell-based vascular responses to prothrombotic stimuli involving platelets, coagulation proteases, and fibrin. Cell–cell and cell–protein interactions, coupled with programmed cellular events, play a particularly important role in vascular repair.

CD18-dependent capture of fibrin protofibrils that are flowing in plasma, and by f XII–dependent thrombin generation. They also activate platelets, providing a fully functional platform for fibrin formation under flow conditions (Goel and Diamond, 2001).

A blood-borne propensity for thrombosis, not entirely dependent on vascular pathology, provides a biology-based explanation for the observed disparities between "degree" or "extent" of atherosclerosis and risk of thrombotic events of the arterial vascular bed (Karnicki et al., 2002), as well as for the propensity toward venous thrombosis with malignancy and following trauma or major surgery.

LINKAGE STUDIES IN THROMBOSIS

Linkage studies differ from association studies by their dependence on transmission from parents to offspring (of a gene marker) and functional genetic variant. Accordingly, linkage investigation requires sibling pairs, nuclear families or extended families (Souto, 2003). Unlike association studies that are prone to false positive and false negative (failure to identify the candidate gene despite the identification of polymorphisms [unrelated to the phenotype in question]) results, linkage studies are more technically challenging but critical to uncovering new genes that causally influence the observed phenotype.

In the Genetic Analysis of Idiopathic Thrombosis (GAIT) family-based study, an additive genetic heritability of 60% for thrombosis was estimated, suggesting the gene mutations would represent the single largest causal mechanism in the pathobiology of disease (Blangero et al., 2003).

Genome-wide scans were carried out using 325 pedigrees with 1144 individuals participating in the Framingham Heart Study (Lin et al., 2007). Using variance-component linkage methods, heritabilities were estimated at greater than 50% for red blood cell count, mean corpuscular volume and mean corpuscular hemoglobin. For red blood cell count, a maximum LOD score of 3.2 on chromosome 19 was identified in close proximity to several genes known to influence cellular differentiation and possibly thrombogenicity-erythropoietin and erythroid Kruppel-like factor.

A variance-component linkage analysis of C-reactive protein (CRP), IL-6, MCP-1, and soluble ICAM was performed in 304 extended families from the Framingham Heart Study (Dupuis et al., 2005). Heritability estimates ranged from 14% to 44%. A significant linkage to MCP-1 was found on chromosome 1 (LOD 4.27), in a region containing several candidate genes, including E-selectin, P-selectin, and CRP. The findings suggest that genes on chromosome 1 may influence inflammation and may have a potential role in atherothrombosis.

Based on available information derived from investigations among individuals with thrombotic disorders, a preferred approach is to utilize linkage studies for gene identification, followed by association studies to further examine polymorphisms in the positional candidate genes (Broeckel et al., 2002; Cooper et al., 2002). If prior linkage studies have not been performed,

candidate genes, at the very least, should be based on known functionality.

ASSOCIATION STUDIES IN THROMBOSIS

The study of complex diseases to include venous and, perhaps to an even greater extent, arterial thrombosis has taken several unique paths. Association studies, conducted among unrelated subjects are undertaken with a basic premise that genetic variants are either related directly to a predetermined phenotype or closely linked to a causative variant. Using either a cohort or case-control approach, the association between genotype and phenotype can be examined. Although population admixture is a potential confounder in genetic association studies, careful selection of a "control" population may address the problem (Wacholder et al., 2000). In contrast, the generalizability and discovery of intermediate phenotypes that can then be used to identify candidate gene or tested in linkage studies is a potential strength of association studies.

A linkage disequilibrium-based genetic approach was utilized to investigate common gene sequence variance in five thrombosis-related genes and related plasma hemostatic proteins among 1811 unrelated Framingham Heart Study participants (Kathiresan et al., 2006). Forty-one tag single nucleotide polymorphisms (SNPs) were genotyped and revealed associations between a fibrinogen-beta SNPs with circulating fibrinogen levels and 7 fVII SNPs and plasma fVII levels. In a step-wise analysis, a single fibrinogen-beta variant explained 1% of the residual variance in fibrinogen levels, and 2 fVII SNPs together explained 10% of the residual variance in fVII levels.

The availability of genome-wide surveys of genetic variants and decreasing cost of genotyping provides an ideal environment for using association studies to unravel complex human conditions. Programs in biostatistical methods and informatics will undoubtedly play a critical role in realizing the full potential of this technique (Lin et al., 2006).

HERITABILITY AND THROMBOSIS: EXISTING COMPLEXITIES

Heritability, defined as the proportion of the total phenotypic variation within a population, is attributed to genetic variance – also referred to as shared variance. The challenge in complex diseases, including thrombotic disorders, is driven by a lack of independence for genetic and environmental components. In many cases, there is an interaction between the two, with one or more environmental factors modifying the genetic effect.

A common illustration is an individual with an inherited thrombophilia who experiences trauma or undergoes major surgery. Under these circumstances, an existing predisposition, coupled with an acquired risk may culminate in a thrombotic event.

There are several subtypes of genetic variance: additive (variance resulting from individual alleles and best represented in parent–offspring studies); dominance (variance stemming from pairs of homologous alleles as determined in twin studies); and epistasis (variance from genes that affect the expression of other genes). Accordingly, heritability is itself complex and may be most suitable for detecting the presence (or absence) of genetic variance (van Asselt et al., 2006).

A PERSONALIZED APPROACH TO HEMOSTASIS AND THROMBOSIS

A recognized priority area for thrombosis-related investigation is identifying patients who will ultimately experience the clinical phenotype. This knowledge allows a clear distinction between individuals in whom treatment is required to prevent life-altering or life-threatening events, and those who potentially incur undue risk from antithrombotics with little likelihood of benefit.

Numerous studies have demonstrated the ability of gene-expression profiles to identify subtle distinctions that define important clinical phenotypes. In a study performed at Duke University Medical Center, Potti and colleagues (Potti et al., 2006) identified gene-expression profiles that predicted thrombotic events among patients with anticardiolipin antibodies (Figures 30.6 and 30.7). An analysis of 50 genes who expression patterns provided the power to discriminate and predict thrombosis included APOE, coagulation factor X, thomboxane, matrix metalloproteinase 19, interleukin 22 receptor, and hematopoietic progenitor cell antigen (CD34) precursor.

An ability to identify polymophisms in high-porosity genes may, in turn, lead to novel diagnostic tools for determining patients at risk, and intervening prior to a clinical event.

PATIENT SCREENING: A TRADITIONAL PARADIGM

Whom to Investigate

In patients with arterial thrombosis, the search for an underlying thrombophilic condition is justified in the presence of at least one of the following: (1) recurrent thromboembolic event; (2) young age (≤50 years if male, ≤55 years if female); (3) lack of significant arterial stenosis at angiography; (4) age ≤55 years if male or ≤60 years if female and no apparent cause (i.e., lack of traditional cardiovascular risk factors, systemic illnesses, malignancies, offending drugs); or (5) age ≤55 years if male or ≤60 years if female and strong family history of thrombosis (Table 30.3) (Figure 30.8) (Andreotti and Becker, 2005). A venous thromboembolic (VTE) event should prompt further evaluation in the following individuals: age <50 years, thrombosis at unusual sites (cerebral, mesenteric, hepatic, portal veins), recurrent venous thrombosis, venous thrombosis and a strong family history of thrombotic disease, women with recurring miscarriages and/or puerperal complications. A targeted approach to testing

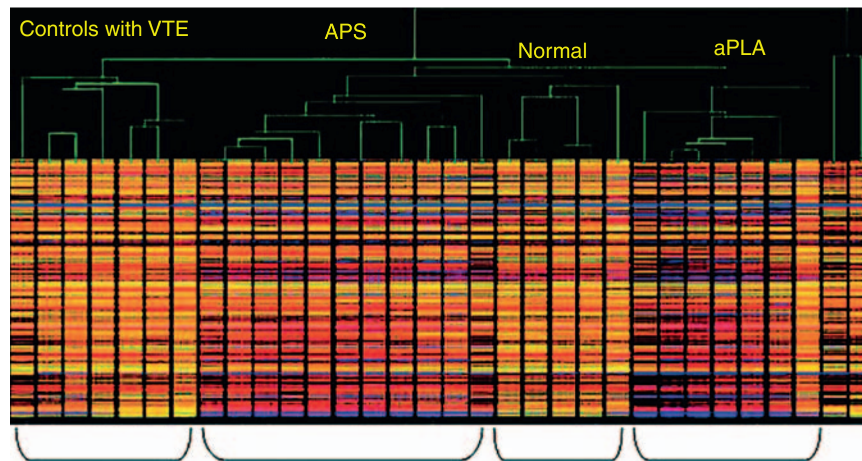


Figure 30.6 Patterns of gene expression that characterize clinical phenotypes. Hierarchal clustering of the initial patient samples based on gene-expression patterns. Each gene is represented by a single row, and each sample is represented by a single column. The color heat map represents genes in a graded fashion along a spectrum of activation, extending from strongly upregulated genes in red to the down-regulated in blue. VTE, venous thromboembolism; APS, antiphospholipid antibody syndromes; aPLA, antiphospholipid antibody. From Potti et al. (2006) with permission.

influences strongly the likelihood of detecting an inherited mutation with biological linkage (Table 30.4) (Seligsohn and Lubetsky, 2001). The importance of gathering a family history, a simple and frequently underutilized tool available to all and useful in assessing the risk for common complex diseases, has been stressed by the Centers for Disease Control and Prevention, Office of Genomics and Disease Prevention (Yoon et al., 2003). Selective testing for common inherited thrombophilias is more cost-effective than universal screening (Wu, 2006).

What and When to Investigate

The tests summarized in Figure 30.8 may be performed during hospitalization, even during the initial stages of a thrombotic episode. Evaluations for an acquired thrombophilia, including those associated with underlying malignancy, systemic disorders, and drug-induced prothrombotic states, should also begin without delay. A definitive diagnosis of myeloproliferative disorder or the search for occult malignancy may require serial office visits and carefully selected imaging studies.

Implications for Treatment

The approaches to inherited and acquired thrombophilias differ at several levels. The former raises questions of susceptibility to recurrent events, treatment duration, and whether to perform testing among related family members (who may themselves carry the trait) (Seligsohn and Lubetsky, 2001). The latter is based on concomitant illnesses and identification of offending drugs or conditions in which diagnosis and treatment of the predisposing disorder have a major impact on the overall thrombotic risk.

PATIENT SCREENING: A COMPREHENSIVE AND POPULATION-BASED APPROACH

Individuals who have experienced an initial VTE event incur a significant risk for recurrence. Traditional risk factors including stasis (from immobility and surgery), vascular injury and either inherited or acquired thrombophilias contribute to sustained risk. In the last 10–15 years it has become apparent that there are number of genetic mutations that are present variably within populations that predispose some individuals to recurring events. Accordingly, attention to personal and family history of thrombosis, with testing according to established guidelines is recommended.

A thrombosis risk intake system, modeled for use in a community setting, is summarized in Figure 30.9. The algorithm and targeted risk management recommendations ideally should be based on the ACMG and CAP consensus statements with respect to genetic testing for thrombophilias and expert opinion. There should be shared decision made before proceeding with the test, as well as education posttest. This particular approach is based on existing concern for patients with inherited thrombophilias. A recent study specifically investigated the knowledge and education needs of individuals with the factor V Leiden mutation (Hellmann et al., 2003). Several observations are worthy of careful consideration: 79% incorrectly estimated their thrombosis risk (2–8 years posttesting); 64% stated that they were given little information about factor V Leiden (pre-or posttesting); 68% had remaining questions at the time of the survey; 53% believed their health care providers to be inadequately informed about the test implications; and 88% preferred knowing the results (of genetic testing).

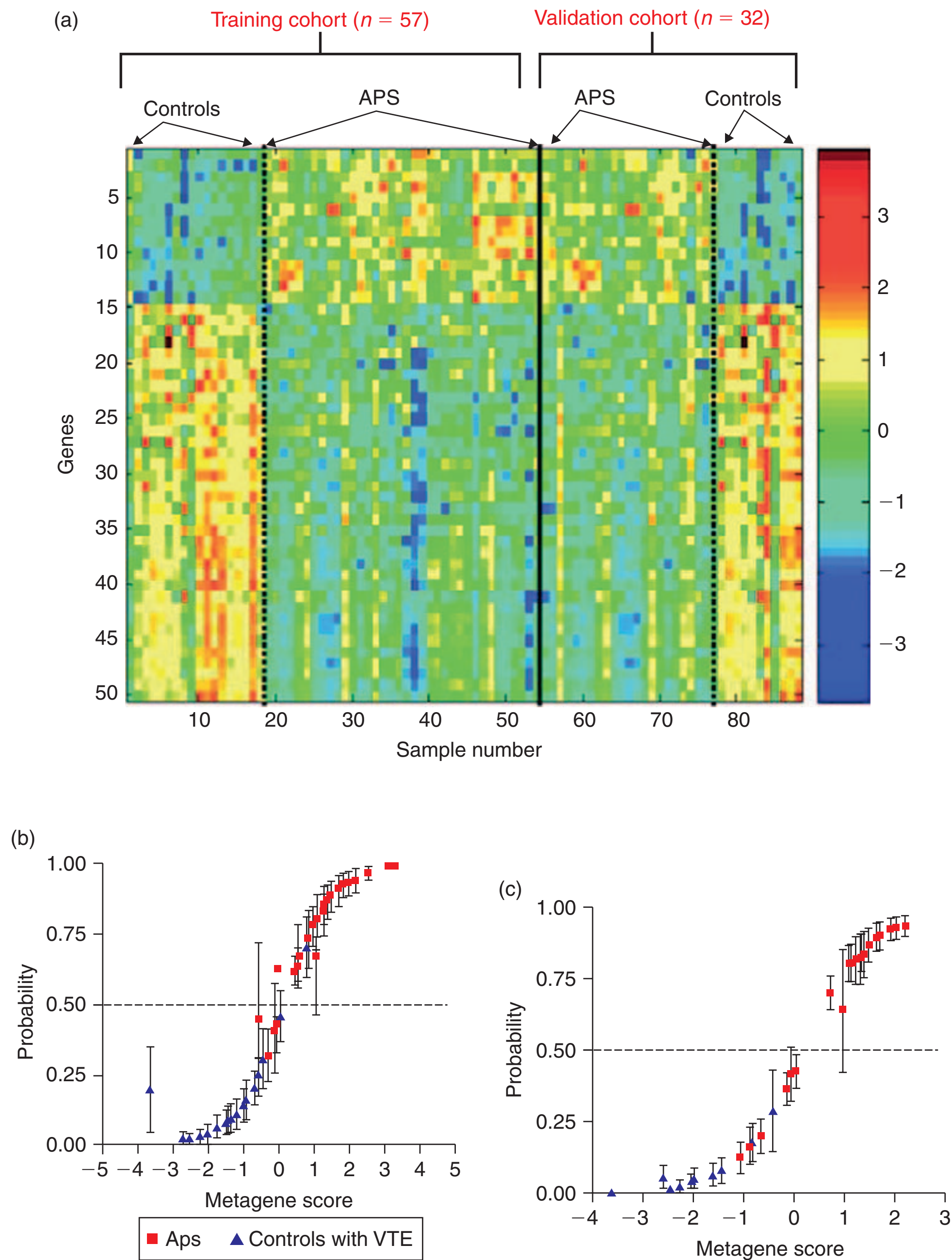


Figure 30.7 (a) Gene-expression profiles that classify and predict APS phenotype; (b) leave- one out cross validation probabilities in the training cohort; (c) Validation of binary regression model in a blinded clot. APS, antiphospholipid antibody syndrome; VTE, venous thromboembolism. From Potti et al. (2006).

PROGNOSTIC CONSIDERATIONS

The most prevalent phenotype of a thrombophilia is venous thromboembolism – a continuum from deep vein thrombosis (DVT) to pulmonary embolism (PE). The annual incidence of venous thromboembolism is approximately 1 in 1000 patients, similar to the annual incidence of stroke. The mortality of PE is particularly high with approximately one in six cases resulting in death. As previously described, individuals who have experienced an initial VTE event have a significant risk of recurrence even after an initial course of treatment. The declining risk over time suggests that concomitant illness and factors related to the thrombus itself contribute.

Recent data, based on patients with unprovoked (spontaneous or idiopathic) VTE, suggest a 10–15% rate of recurrence. The incidence is particularly high following a PE (Gonzalez-Porras

Table 30.3 Conditions that predispose to thromboembolism

Venous	Arterial
Inherited	*Inherited*
1. Gene polymorphisms of the hemostatic system: – factor V Leiden (G1691A) – G20210A prothrombin variant – gain-of-function variants of factor VIII, IX, XI	1. Gene polymorphisms of the hemostatic system:[b] – factor V Leiden (G1691A) – G20210A prothrombin variant – gain-of-function variants of fibrinogen, factor VII, plasminogen activator inhibitor type-1 – glycoprotein IIIa (Leu33Pro)
2. Deficiencis of antithrombin, protein C, protein S	2. Family history of arterial thrombosis[a]
3. Dysfibrinogenemia	3. Homocystinuria
4. Family history of venous thrombosis[a]	4. Congenital dyslipidemias
5. Homocystinuria and MTHF C677T	
6. Varicose veins	
Physiologic	*Physiologic*
1. Pregnancy, pueperium 2. Aging	1. Male gender 2. Aging
Environmental	*Environmental*
1. Surgery, trauma, immobilization 2. Oral contraceptives, HRT 3. Heparin-induced thrombocytopenia 4. Antifibrinolytic agents, prothrombin complex concentrates 5. Endotoxemia	1. Smoking, cocaine use 2. Oral contraceptives, HRT 3. Heparin-induced thrombocytopenia 4. Antifibrinolytic agents, prothrombin complex concentrates 5. Thienopyridine-related TTP
Other	*Other*
1. Previous venous thromboembolism 2. Obesity 3. Malignancies 4. Antiphospholipid antibodies 5. Polycythemia vera, essential thrombocytosis 6. Congestive heart failure 7. Nephrotic syndrome 8. Behcet's disease, other vasculitides 9. Paroxysmal nocturnal hemoglobinuria	1. Previous arterial thrombosis 2. Atherosclerosis, vasculitis 3. Hypercholesterolemia, metabolic syndrome[c] 4. Congestive heart failure, renal failure 5. Atrial fibrillation 6. Antiphospholipid antibodies, rheumatoid arthritis 7. Polycythemia vera, essential thrombocytosis 8. Sickle cell anemia, macroglobulinemia 9. Malignancies

Modified from Voetsch et al. (2004) and Levine et al. (2002) MTHF = methylenetetrahydrofolate reductase; HRT = hormone replacement therapy; TTP = thrombotic thrombocytopenic purpura.
[a]May be defined as at least 1 first-degree relative affected ≤50 years if male and ≤55 years if female.
[b]Relation between genotype and thrombosis variable in different studies.
[c]Components of the metabolic syndrome include hypertension, diabetes or glucose intolerance, obesity, reduced HDL-cholesterol, and hypertriglyceridemia.

et al., 2006). Although the presence of an inherited thrombophilia is associated with increased risk, common mutations such as factor V Leiden, prothrombin 20210, and methylene tetrahydrofolate reductase (MTHFR), pose the greatest concern when combined (double defects), or exist as homozygous or double heterozygous forms.

The case fatality rates for recurrent PE during anticoagulant therapy vary from 8% to as high as 26%. Recurrence after a course of therapy, on average for 3–6 months, is fatal in approximately 5% of cases (Gonzalez-Porras et al., 2006).

The risk of recurrent arterial thrombosis is determined by several factors, include age, vascular bed origin, coexisting disease in other vascular beds, concomitant risk factors, the presence of a defined thrombophilia, and the use and associated effectiveness of existing treatments (that may included percutaneous and/or surgical interventions).

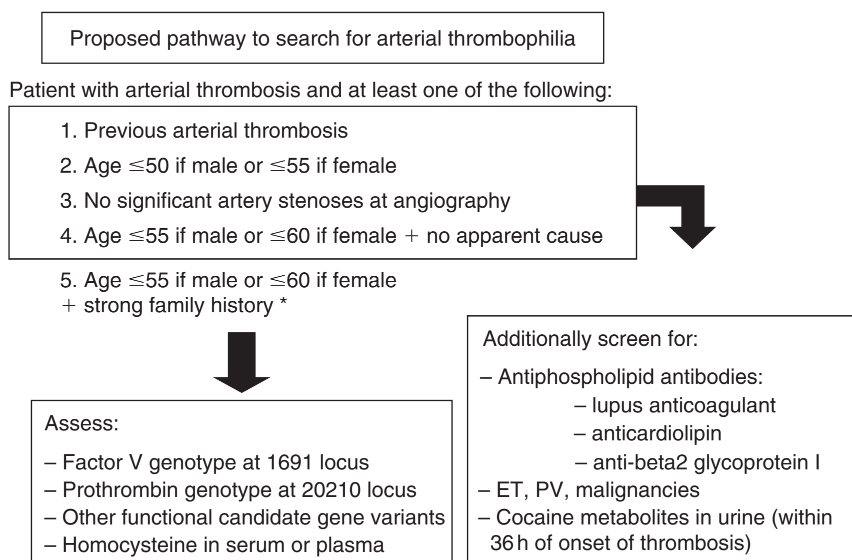


Figure 30.8 Proposed criteria for the selection of patients in whom it may be justified to perform a number of nonroutine tests for underlying thrombophilic conditions. Asterisk denotes at least 1 first-degree relative affected at age ≤50 years if male of ≤55 years if female. ET indicates essential thrombocytosis; PV, polycythemia vera. From Andreotti and Becker (2005).

Table 30.4 The prevalence of factor V Leiden and prothrombin gene mutations in healthy, unselected and selected populations with venous thrombosis

Genetic variant	Healthy subjects		Unselected patients with venous		Selected patients with venous thromboembolism	
	# Examined	% with variant	# Examined	% with variant	# Examined	% with variant
Factor V Leiden, heterozygote	16,150	4.8	1142	18.8	162	40
Prothrombin G20210A, heterozygote	11,932	2.7	2884	7.1	551	16

Selected patients, age less than 50, a family history of venous thrombosis, a history of recurrent events, and the absence of acquired risk factors except pregnancy of the use of oral contraceptives. Adapted from Seligsohn (2001).

EMERGING PLATFORM FOR HEMOSTASIS AND THROMBOSIS RESEARCH

Scientific advances have provided a strong knowledge base for research agendas in the fields of hemostasis and thrombosis. Coupled with great strides in molecular biology, the next decade promises many opportunities to translate fundamental science to patient care. Several emerging platforms will be instrumental and include: proteomic analysis capabilities for defining signaling cascades in platelets (Garcia, 2006), systems-based models of individual pro- and anticoagulant factor levels, and related *in silico* prediction models of thrombotic capacity (hemostatic proteome) (Brummel-Ziedins et al., 2005), expression profiles of circulating blood components (Potti et al., 2006) and thrombus modulation achieved through inhibition of cellular adhesive events (Jackson et al., 2005).

A pyrosequencing-based genotype platform that analyzes common prothrombotic, hemostatic, and treatment-modulating mutations may represent an optimized diagnostic strategy for analysis (Holmberg et al., 2005), with potential application in both epidemiologic studies and clinical diagnosis (Mooser et al., 2003).

Pharmacogenomics

Antithrombotic therapy has improved patient outcomes across a broad range of thrombotic disorders; however, not all patients in all clinical settings respond favorably. Accordingly, efforts must be undertaken to better distinguish "good responders" from "bad responders" or "non-responders." Pharmacogenomics represents an important part of an emerging paradigm for drug

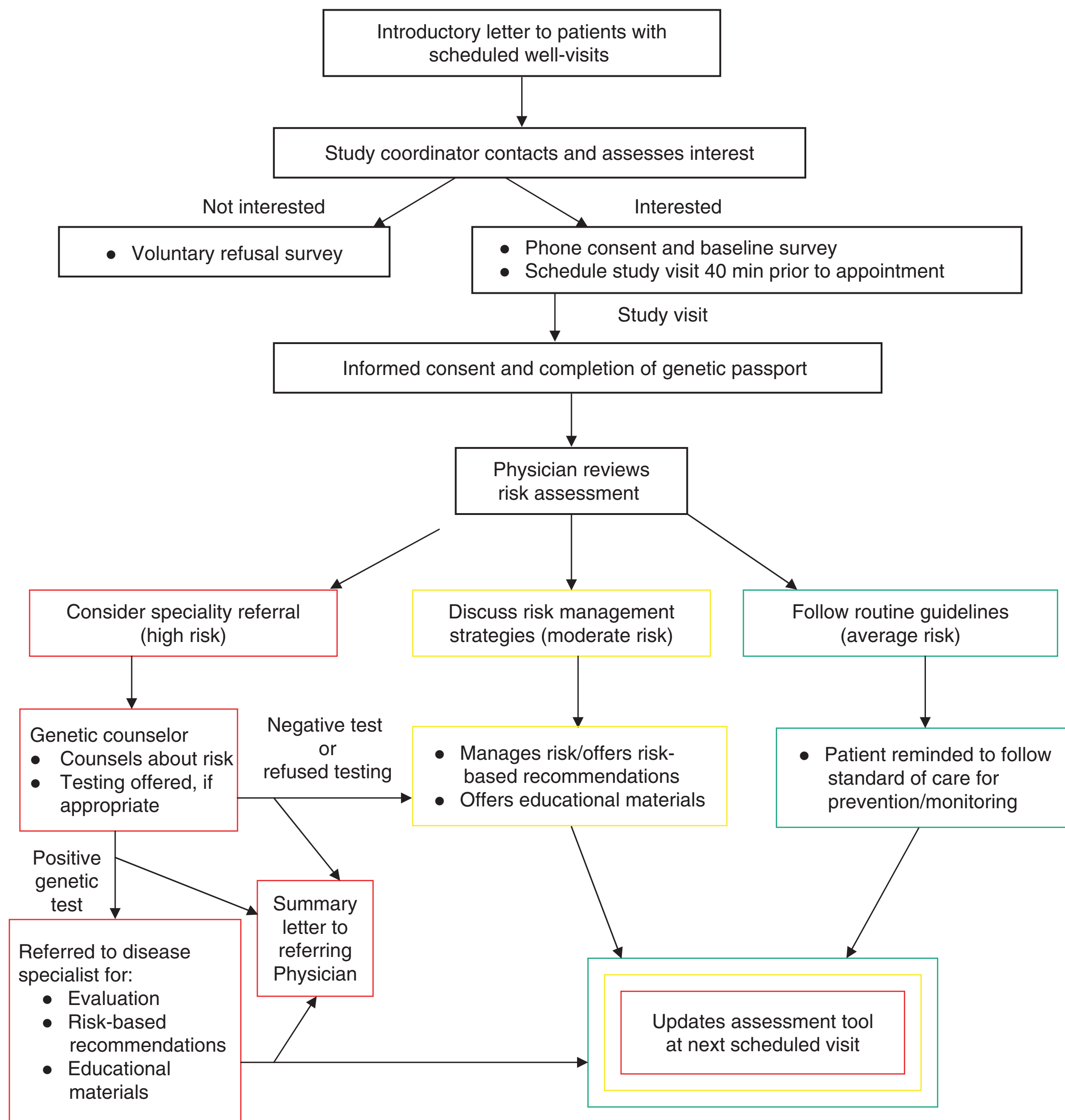


Figure 30.9 A population-based approach to thrombophilias emphasizes family and patient history of thrombotic events, carefully selected tests, genetic counseling when appropriate and educated disease management.

treatment given gene-based influences on enzymes, transporters, and receptors involved with several fundamental properties, including drug absorption, metabolism, excretion, mechanism of action, and toxicity (reviewed in reference [Palkimas et al., 2003]). An illustration of pharmacogenomics as a platform for patient care is well represented with the anticoagulant warfarin (a vitamin K antagonists used widely in the management of thrombotic disorders). Genetic variations in the hepatic metabolism of S-warfarin, the more potent isomer that accounts for 70% of the overall anticoagulant response, by P450 2C9 have important implications in drug dose requirements and risk of bleeding (reviewed in reference [Aithal et al., 1999]) (Table 30.5) (Dang et al., 2005; Higashi et al., 2002; Loebstein et al., 2001; Tabrizi et al., 2002; Taube et al., 2000). Although the prevalence of 2C9 polymorphisms varies among ethnic groups, this may not fully explain differences in warfarin dose requirements (Dang et al., 2005). One must acknowledge, despite the likely importance of pharmacogenomics in clinical practice, that the role of genetic testing in therapy decisions must be defined through carefully designed clinical trials.

The Bloodomics project at Cambridge University, financed by the European Community, has a primary objective to

Table 30.5 Genetics of warfarin metabolism

Author and *n*	Genetic variants	Effect on dosage requirements	Increased risks
Higashi et al. $n = 185$	*CYP2C9*1/*1*($n = 127$) *CYP2C9*1/*2*($n = 28$) *CYP2C9*1/*3*($n = 18$) *CYP2C9*2/*2*($n = 4$) *CYP2C9*2/*3*($n = 3$) *CYP2C9*3/*3*($n = 5$)	Mean dose is significantly related to genotype being highest in *CYP2C9*1/*1* (5.63 mg) and lowest for *CYP2C9*3/*3* (1.60 mg) ($p = 0.001$)	Increased risk of high INR, longer time to achieve stable dosing, and increase serious bleeding events
Tabrizi et al. $n = 153$	*CYP2C9*1/*1*($n = 107$) *CYP2C9*1/*2*($n = 22$) *CYP2C9*1/*3*($n = 21$) *CYP2C9*2/*2*($n = 1$) *CYP2C9*2/*3*($n = 1$) *CYP2C9*3/*3*($n = 1$)	Significant decrease in weekly dose in *CYP2C9*2* (28.3 mg) and *CYP2C9*3* (27.9 mg) compared to wild type allele (40.1 mg) ($p = 3860.0021$ and $p = 0.0016$, respectively)	Not available
Taube et al. $n = 561$	*CYP2C9*1/*1*($n = 392$) *CYP2C9*1/*2*($n = 107$) *CYP2C9*1/*3*($n = 53$) *CYP2C9*2/*2*($n = 3$) *CYP2C9*2/*3*($n = 6$) *CYP2C9*3/*3*($n = 0$)	Dose is related to genotype: homozygous CYP2C9*2 had the lowest maintenance dose (3.04 mg) compared to the wild type (5.01 mg) ($p = 0.001$)	No increased risks found
Loebstein et al. $n = 156$	Age <65 *CYP2C9*1*($n = 49$) *CYP2C9*2*($n = 15$) *CYP2C9*3*($n = 10$) Age > 65 *CYP2C9*1*($n = 59$) *CYP2C9*2*($n = 13$) *CYP2C9*3*($n = 10$)	The dose was lowest in patients >65 years old with the CYP2C9*3 variant	Not available
Aithal et al. $n = 188$	Low dose patients *CYP2C9*1/*1*($n = 7$) *CYP2C9*1/*2*($n = 12$) *CYP2C9*1/*3*($n = 10$) *CYP2C9*2/*2*($n = 2$) *CYP2C9*2/*3*($n = 5$) *CYP2C9*3/*3*($n = 0$) Random dose patients *CYP2C9*1/*1*($n = 32$) *CYP2C9*1/*2*($n = 9$) *CYP2C9*1/*3*($n = 10$) *CYP2C9*2/*2*($n = 1$) *CYP2C9*2/*3*($n = 0$) *CYP2C9*3/*3*($n = 0$)	A strong correlation between CYP2C9 and warfarin sensitivity was demonstrated. In addition, there were a higher percentage of variants in the low dose group	Difficulties at initiation of warfarin therapy due to supratherapeutic INR and increased episodes of major bleeding

Adapted from Palkimas, et al. (2003) with permission.

discover genetic markers within platelets that predict arterial thrombosis risk. An overarching plan is to identify and catalog SNPs in 300 genes, and establish a basis for population genetics (reviewed in reference [Nurden, 2006]).

Polymorphisms in platelet receptors, coagulation proteins and fibrin; coupled with concomitant genetic polymorphisms affecting the metabolism, disposition, transporter proteins, and target binding of antithrombotic drugs emphasizes the importance of ongoing pharmacogenomic research within clinical trials. Application of bioinformatics, genomics, and pharmacogenomics will play a critical role in the unraveling of complex interactions (Iqbal and Fareed, 2006).

2009 UPDATE

The field of genomics as it relates to coagulation and thrombosis is evolving rapidly in three major areas: genetic predisposition to venous thrombosis, genetic predisposition to drug-induced, malignancy-associated venous thromboembolism, and the pharmacogenomics of antithrombotic therapy for thrombotic disorders of the coronary arterial circulatory system.

While an inherited predisposition to venous thrombosis is detectable in upward of 50% of selected patients, the remaining patients lack an identifiable thrombotic mutation, even under conditions in which multiple family members or several generations have the clinical phenotype. The employment of large population data sets, including the LETS and multiple environmental and genetic assessment (MEGA) of Risk Factors for Venous Thromboembolism Study provided a powerful resource for assessing fundamental questions of thrombotic risk and common inherited variations (Bezemer et al., 2008).

Malignancy is a common acquired thrombophilia that poses a strong risk for venous thromboembolism. In some instances, chemotherapy itself may heighten the inherent risk, creating unique challenges for both clinicians and patients. Large-scale clinical trials and established networks have recently provided an opportunity to investigate genomic associations for thrombosis. Using patients and controls from the Medical Research Council (MRC) Myeloma IX, Hovon-50, and Eastern Cooperative Oncology Group (ECOG) studies, Johnson and Colleagues identified a set of SNPs, associated with thalidomide-related venous thromboembolism, that were enriched in genes and pathways important in drug transport/metabolism, DNA repair, and cytokine synthesis. (Johnson et al., 2008). The fundamental question that requires further investigation is whether the relationship exists at the tumor level, tumor-related microenvironment, or the vascular endothelium. Further studies will be required to identify those patients who are at increased risk for thromboembolism in the setting of malignancy.

The opportunity for translation of pharmacogenomics to patient care in coronary arterial thrombotic disorders is well illustrated in two recent studies (Mega et al., 2008; Simon et al., 2008). In the Trial to Assess Improvement in Therapeutic Outcomes by Optimizing Platelet Inhibition with Prasugrel (TRITON) Study (Mega et al., 2008), 1477 subjects with acute coronary syndromes treated with clopidogrel – who were carriers of at least one CYP2C19 reduced-function allele (27.1% of the study population) – were at significantly higher risk for death from cardiovascular causes, MI, or stroke than noncarriers (12.1% versus 8.0%; hazard ratio, 1.53; 95% CI 1.07–2.19; $p = 0.01$). The risk of stent thrombosis in carriers of a reduced-function allele was threefold higher than noncarriers (2.6% versus 0.08%; hazard ratio, 3.09; 95% CI, 1.19–8.00; $p = 0.02$). The presence of at least one copy of the *2 allele for CYP2C19 accounted for 95% of the subjects classified as carriers of a reduced-function allele. In a study of healthy subjects with the CYP gene variant, carriers had a relative reduction of 32.4% in plasma exposure to clopidogrel's active metabolite, and a diminished pharmacodynamic response, with an absolute change in maximal platelet aggregation in response to clopidogrel that was 9% points less than noncarriers (Mega et al., 2008).

The French Registry of Acute ST-elevation and non ST-elevation Myocardial Infarction (FAST-MI) investigators (Simon et al., 2008) assessed the relation of allelic variants in genes modulating clopidogrel absorption (ABCB1), metabolic activation (CYP3A5, CYP2C19), and biologic activity ($P2RY_{12}$, ITGB3) to the risk of death (all cause), MI, or nonfatal stroke during one year of follow-up in 2208 patients. Patients with two variant alleles of ABCB1 had a higher rate of cardiovascular events than those with the wild-type genotype (15.5% versus 10.7% adjusted hazard ratio 1.72; 95% CI, 1.20–2.47). Patients carrying any two CYP2C19 loss-of-function alleles (*2,*3, *4, or *5) had a higher event rate than patients without gene variants (21.5% versus 13.3%; adjusted hazard ratio 1.98, 95% CI 1.10–3.58). The event rates were particularly high among patients with CYP2C19 loss-of-function alleles who underwent PCI during initial hospitalization (3.58fold higher; 95% CI, 1.71–7.51). The presence of at least one ABCB1 variant allele and two CYP2C19 alleles was associated with the highest hazard for events (5.31; 95% CI, 1.71–7.51; $p = 0.005$).

The CYP2C19 gene variant was also a determinant for prognosis in patients less than 45 years of age who were receiving clopidogrel after treatment for a first acute myocardial infarction (Collet et al., 2008). Clearly, determination of CYP2C19 status will soon be part of the standard evaluation of patients who are candidates for clopidogrel in the near future.

2009 UPDATE REFERENCES

Bezemer, I.D., Bare, L.A. et al. (2008). Gene variants associated with deep vein thrombosis. *JAMA* 299(11), 1306–1314.

Collet, J.P., Hulot, J.S. et al. (2008). Cytochrome P450 2C19 polymorphism in young patients treated with clopidogrel after myocardial infarction: A cohort study. *Lancet* 373(9660), 309–317.

Johnson, D.C., Corthals, S. et al. (2008). Genetic associations with thalidomide mediated venous thrombotic events in myeloma identified using targeted genotyping. *Blood* 112(13), 4924–4934.

Mega, J.L., Close, S.L. et al. (2008). Cytochrome P-450 polymorphisms and response to clopidogrel. *N Engl J Med* 360(4), 354–362.

Simon, T., Verstuyft, C. et al. (2008). Genetic determinants of response to clopidogrel and cardiovascular events. *N Engl J Med* 360(4), 363–375.

REFERENCES

Abi-Rached, L., Gilles, A., Shiina, T., Pontarotti, P. and Inoko, H. (2002). Evidence of en bloc duplication in vertebrate genomes. *Nat Genet* 31, 100–105.

Aithal, G.P., Day, C.P., Kesteven, P.J. and Daly, A.K. (1999). Association of polymorphisms in the cytochrome P450 CYP2C9 with warfarin dose requirement and risk of bleeding complications. *Lancet* 353, 717–719.

Andreotti, F. and Becker, R.C. (2005). Atherothrombotic disorders: New insights from hematology. *Circulation* 111, 1855–1863.

Bavendiek, U., Libby, P., Kilbride, M., Reynolds, R., Mackman, N. and Schonbeck, U. (2002). Induction of tissue factor expression in human endothelial cells by CD40 ligand is mediated via activator protein 1, nuclear factor kappa B, and Egr-1. *J Biol Chem* 277, 25032–25039.

Beekhuizen, H. and van Furth, R. (1994). Growth characteristics of cultured human macrovascular venous and arterial and microvascular endothelial cells. *J Vasc Res* 31, 230–239.

Bladbjerg, E.M., de Maat, M.P., Christensen, K., Bathum, L., Jespersen, J. and Hjelmborg, J. (2006). Genetic influence on thrombotic risk markers in the elderly – a Danish twin study. *J Thromb Haemost* 4, 599–607.

Blangero, J., Williams, J.T. and Almasy, L. (2003). Novel family-based approaches to genetic risk in thrombosis. *J Thromb Haemost* 1, 1391–1397.

Braddock, M., Schwachtgen, J.L., Houston, P., Dickson, M.C., Lee, M.J. and Campbell, C.J. (1998). Fluid shear stress modulation of gene expression in endothelial cells. *News Physiol Sci* 13, 241–246.

Broeckel, U., Hengstenberg, C., Mayer, B., Holmer, S., Martin, L.J., Comuzzie, A.G., Blangero, J., Nurnberg, P., Reis, A. et al. (2002). A comprehensive linkage analysis for myocardial infarction and its related risk factors. *Nat Genet* 30, 210–214.

Brummel-Ziedins, K., Vossen, C.Y., Rosendaal, F.R., Umezaki, K. and Mann, K.G. (2005). The plasma hemostatic proteome: Thrombin generation in healthy individuals. *J Thromb Haemost* 3, 1472–1481.

Chang, Y.S., Munn, L.L., Hillsley, M.V., Dull, R.O., Yuan, J., Lakshminarayanan, S., Gardner, T.W., Jain, R.K. and Tarbell, J.M. (2000). Effect of vascular endothelial growth factor on cultured endothelial cell monolayer transport properties. *Microvasc Res* 59, 265–277.

Chien, S., Li, S. and Shyy, Y.J. (1998). Effects of mechanical forces on signal transduction and gene expression in endothelial cells. *Hypertension* 31, 162–169.

Christie, P.D., Edelberg, J.M., Picard, M.H., Foulkes, A.S., Mamuya, W., Weiler-Guettler, H., Rubin, R.H., Gilbert, P. and Rosenberg, R.D. (1999). A murine model of myocardial microvascular thrombosis. *J Clin Invest* 104, 533–539.

Cooper, D.N., Nussbaum, R.L. and Krawczak, M. (2002). Proposed guidelines for papers describing DNA polymorphism-disease associations. *Hum Genet* 110, 207–208.

Dang, M.T., Hambleton, J. and Kayser, S.R. (2005). The influence of ethnicity on warfarin dosage requirement. *Ann Pharmacother* 39, 1008–1012.

Davidson, C.J., Hirt, R.P., Lal, K., Snell, P., Elgar, G., Tuddenham, E.G. and McVey, J.H. (2003a). Molecular evolution of the vertebrate blood coagulation network. *Thromb Haemost* 89, 420–428.

Davidson, C.J., Tuddenham, E.G. and McVey, J.H. (2003b). 450 million years of hemostasis. *J Thromb Haemost* 1, 1487–1494.

De Cristofaro, R., Rocca, B., Marchioli, R. and Landolfi, R. (2002). Plasma protein oxidation is associated with an increase of procoagulant markers causing an imbalance between pro- and anticoagulant pathways in healthy subjects. *Thromb Haemost* 87, 58–67.

de Lange, M., Snieder, H., Ariens, R.A., Spector, T.D. and Grant, P.J. (2001). The genetics of haemostasis: A twin study. *Lancet* 357, 101–105.

Dupuis, J., Larson, M.G., Vasan, R.S., Massaro, J.M., Wilson, P.W., Lipinska, I., Corey, D., Vita, J.A., Keaney, J.F., Jr and Benjamin, E.J. (2005). Genome scan of systemic biomarkers of vascular inflammation in the Framingham Heart Study: Evidence for susceptibility loci on 1q. *Atherosclerosis* 182, 307–314.

Edelberg, J.M., Aird, W.C., Wu, W., Rayburn, H., Mamuya, W.S., Mercola, M. and Rosenberg, R.D. (1998). PDGF mediates cardiac microvascular communication. *J Clin Invest* 102, 837–843.

Edelberg, J.M., Christie, P.D. and Rosenberg, R.D. (2001). Regulation of vascular bed-specific prothrombotic potential. *Circ Res* 89, 117–124.

Falciani, M., Gori, A.M., Fedi, S., Chiarugi, L., Simonetti, I., Dabizzi, R.P., Prisco, D., Pepe, G., Abbate, R., Gensini, G.F. et al. (1998). Elevated tissue factor and tissue factor pathway inhibitor circulating levels in ischaemic heart disease patients. *Thromb Haemost* 79, 495–499.

Garcia, A. (2006). Proteome analysis of signaling cascades in human platelets. *Blood Cells Mol Dis* 36, 152–156.

Giesen, P.L., Rauch, U., Bohrmann, B., Kling, D., Roque, M., Fallon, J.T., Badimon, J.J., Himber, J., Riederer, M.A. and Nemerson, Y. (1999). Blood-borne tissue factor: Another view of thrombosis. *Proc Natl Acad Sci USA* 96, 2311–2315.

Goel, M.S. and Diamond, S.L. (2001). Neutrophil enhancement of fibrin deposition under flow through platelet-dependent and -independent mechanisms. *Arterioscler Thromb Vasc Biol* 21, 2093–2098.

Gonzalez-Porras, J.R., Garcia-Sanz, R., Alberca, I., Lopez, M.L., Balanzategui, A., Gutierrez, O., Lozano, F. and San Miguel, J. (2006). Risk of recurrent venous thrombosis in patients with G20210A mutation in the prothrombin gene or factor V Leiden mutation. *Blood Coagul Fibrinolysis* 17, 23–28.

Guillot, P.V., Liu, L., Kuivenhoven, J.A., Guan, J., Rosenberg, R.D. and Aird, W.C. (2000). Targeting of human eNOS promoter to the Hprt locus of mice leads to tissue-restricted transgene expression. *Physiol Genomics* 2, 77–83.

Hanumanthaiah, R., Day, K. and Jagadeeswaran, P. (2002). Comprehensive analysis of blood coagulation pathways in teleostei: Evolution of coagulation factor genes and identification of zebrafish factor VIIi. *Blood Cells Mol Dis* 29, 57–68.

Hellmann, E.A., Leslie, N.D. and Moll, S. (2003). Knowledge and educational needs of individuals with the factor V Leiden mutation. *J Thromb Haemost* 1, 2335–2339.

Higashi, M.K., Veenstra, D.L., Kondo, L.M., Wittkowsky, A. K., Srinouanprachanh, S.L., Farin, F.M. and Rettie, A.E. (2002). Association between CYP2C9 genetic variants and anticoagulation-related outcomes during warfarin therapy. *JAMA* 287, 1690–1698.

Holmberg, K., Persson, M.L., Uhlen, M. and Odeberg, J. (2005). Pyrosequencing analysis of thrombosis-associated risk markers. *Clin Chem* 51, 1549–1552.

Iqbal, O. and Fareed, J. (2006). Clinical applications of bioinformatics, genomics, and pharmacogenomics. *Methods Mol Biol* 316, 159–177.

Jackson, S.P., Schoenwaelder, S.M., Goncalves, I., Nesbitt, W.S., Yap, C.L., Wright, C.E., Kenche, V., Anderson, K.E., Dopheide, S.M.,

Yuan, Y. et al. (2005). PI 3-kinase p110beta: A new target for antithrombotic therapy. *Nat Med* 11, 507–514.

Kannel, W.B. and Wolf, P.A. (2006). Peripheral and cerebral atherothrombosis and cardiovascular events in different vascular territories: Insights from the Framingham Study. *Curr Atheroscler Rep* 8, 317–323.

Karnicki, K., Owen, W.G., Miller, R.S. and McBane, R.D., 2nd (2002). Factors contributing to individual propensity for arterial thrombosis. *Arterioscler Thromb Vasc Biol* 22, 1495–1499.

Kathiresan, S., Yang, Q., Larson, M.G., Camargo, A.L., Tofler, G.H., Hirschhorn, J.N., Gabriel, S.B. and O'Donnell, C.J. (2006). Common genetic variation in five thrombosis genes and relations to plasma hemostatic protein level and cardiovascular disease risk. *Arterioscler Thromb Vasc Biol* 26, 1405–1412.

Krem, M.M. and Di Cera, E. (2002). Evolution of enzyme cascades from embryonic development to blood coagulation. *Trends Biochem Sci* 27, 67–74.

Le Flem, L., Picard, V., Emmerich, J., Gandrille, S., Fiessinger, J.N., Aiach, M. and Alhenc-Gelas, M. (1999). Mutations in promoter region of thrombomodulin and venous thromboembolic disease. *Arterioscler Thromb Vasc Biol* 19, 1098–1104.

Lin, J.P., O'Donnell C, J., Jin, L., Fox, C., Yang, Q. and Cupples, L.A. (2007). Evidence for linkage of red blood cell size and count: Genome-wide scans in the Framingham Heart Study. *Am J Hematol* 82, 605–610.

Lin, M.C., Almus-Jacobs, F., Chen, H.H., Parry, G.C., Mackman, N., Shyy, J.Y. and Chien, S. (1997). Shear stress induction of the tissue factor gene. *J Clin Invest* 99, 737–744.

Loebstein, R., Yonath, H., Peleg, D., Almog, S., Rotenberg, M., Lubetsky, A., Roitelman, J., Harats, D., Halkin, H. and Ezra, D. (2001). Interindividual variability in sensitivity to warfarin – Nature or nurture? *Clin Pharmacol Ther* 70, 159–164.

Loscalzo, J. (2003). Proteomics in cardiovascular biology and medicine. *Circulation* 108, 380–383.

Mackman, N. (2005). Tissue-specific hemostasis in mice. *Arterioscler Thromb Vasc Biol* 25, 2273–2281.

Mann, K.G., Brummel-Ziedins, K., Orfeo, T. and Butenas, S. (2006). Models of blood coagulation. *Blood Cells Mol Dis* 36, 108–117.

McEver, R.P. (2001). Adhesive interactions of leukocytes, platelets, and the vessel wall during hemostasis and inflammation. *Thromb Haemost* 86, 746–756.

McLysaght, A., Hokamp, K. and Wolfe, K.H. (2002). Extensive genomic duplication during early chordate evolution. *Nat Genet* 31, 200–204.

Monroe, D.M. and Hoffman, M. (2006). What does it take to make the perfect clot? *Arterioscler Thromb Vasc Biol* 26, 41–48.

Mooser, V., Waterworth, D.M., Isenhour, T. and Middleton, L. (2003). Cardiovascular pharmacogenetics in the SNP era. *J Thromb Haemost* 1, 1398–1402.

Nesheim, M. (2001). Myocardial infarction and the balance between fibrin deposition and removal. *Ital Heart J* 2, 641–645.

Nishida, M., Springhorn, J.P., Kelly, R.A. and Smith, T.W. (1993). Cell-cell signaling between adult rat ventricular myocytes and cardiac microvascular endothelial cells in heterotypic primary culture. *J Clin Invest* 91, 1934–1941.

Nurden, A.T. (2006). Polymorphisms and platelet genotyping: The shape of things to come. *J Thromb Haemost* 4, 1194–1196.

Palkimas, M.P., Jr, Skinner, H.M., Gandhi, P.J. and Gardner, A.J. (2003). Polymorphism induced sensitivity to warfarin: A review of the literature. *J Thromb Thrombolysis* 15, 205–212.

Potti, A., Bild, A., Dressman, H.K., Lewis, D.A., Nevins, J.R. and Ortel, T.L. (2006). Gene-expression patterns predict phenotypes of immune-mediated thrombosis. *Blood* 107, 1391–1396.

Romualdi, C., Trevisan, S., Celegato, B., Costa, G. and Lanfranchi, G. (2003). Improved detection of differentially expressed genes in microarray experiments through multiple scanning and image integration. *Nucleic Acids Res* 31, e149.

Rupnick, M.A., Carey, A. and Williams, S.K. (1988). Phenotypic diversity in cultured cerebral microvascular endothelial cells. *In Vitro Cell Dev Biol* 24, 435–444.

Scott, E.M., Ariens, R.A. and Grant, P.J. (2004). Genetic and environmental determinants of fibrin structure and function: Relevance to clinical disease. *Arterioscler Thromb Vasc Biol* 24, 1558–1566.

Seligsohn, U. and Lubetsky, A. (2001). Genetic susceptibility to venous thrombosis. *N Engl J Med* 344, 1222–1231.

Sokabe, T., Yamamoto, K., Ohura, N., Nakatsuka, H., Qin, K., Obi, S., Kamiya, A. and Ando, J. (2004). Differential regulation of urokinase-type plasminogen activator expression by fluid shear stress in human coronary artery endothelial cells. *Am J Physiol Heart Circ Physiol* 287, H2027–H2034.

Souto, J.C. (2003). Genetic studies in complex disease: The case pro linkage studies. *J Thromb Haemost* 1, 1676–1678.

Tabrizi, A.R., Zehnbauer, B.A., Borecki, I.B., McGrath, S.D., Buchman, T.G. and Freeman, B.D. (2002). The frequency and effects of cytochrome P450 (CYP) 2C9 polymorphisms in patients receiving warfarin. *J Am Coll Surg* 194, 267–273.

Tabrizi, P., Wang, L., Seeds, N., McComb, J.G., Yamada, S., Griffin, J.H., Carmeliet, P., Weiss, M.H. and Zlokovic, B.V. (1999). Tissue plasminogen activator (tPA) deficiency exacerbates cerebrovascular fibrin deposition and brain injury in a murine stroke model: Studies in tPA-deficient mice and wild-type mice on a matched genetic background. *Arterioscler Thromb Vasc Biol* 19, 2801–2806.

Taube, J., Halsall, D. and Baglin, T. (2000). Influence of cytochrome P-450 CYP2C9 polymorphisms on warfarin sensitivity and risk of over-anticoagulation in patients on long-term treatment. *Blood* 96, 1816–1819.

van Asselt, K.M., Kok, H.S., van der Schouw, Y.T., Peeters, P.H., Pearson, P.L. and Grobbee, D.E. (2006). Role of genetic analyses in cardiology: Part II: Heritability estimation for gene searching in multifactorial diseases. *Circulation* 113, 1136–1139.

Van Hylckama Vlieg, A., Callas, P.W., Cushman, M., Bertina, R.M. and Rosendaal, F.R. (2003). Inter-relation of coagulation factors and d-dimer levels in healthy individuals. *J Thromb Haemost* 1, 516–522.

Vlassara, H., Fuh, H., Donnelly, T. and Cybulsky, M. (1995). Advanced glycation endproducts promote adhesion molecule (VCAM-1, ICAM-1) expression and atheroma formation in normal rabbits. *Mol Med* 1, 447–456.

Wacholder, S., Rothman, N. and Caporaso, N. (2000). Population stratification in epidemiologic studies of common genetic variants and cancer: Quantification of bias. *J Natl Cancer Inst* 92, 1151–1158.

Wang, T.J., Gona, P., Larson, M.G., Tofler, G.H., Levy, D., Newton-Cheh, C., Jacques, P.F., Rifai, N., Selhub, J., Robins, S.J. et al. (2006). Multiple biomarkers for the prediction of first major cardiovascular events and death. *N Engl J Med* 355, 2631–2639.

Yoon, P.W., Scheuner, M.T. and Khoury, M.J. (2003). Research priorities for evaluating family history in the prevention of common chronic diseases. *Am J Prev Med* 24, 128–135.

CHAPTER

Genomics of Congenital Heart Disease

Jessie H. Conta and Roger E. Breitbart

INTRODUCTION

Congenital heart disease (CHD) refers to structural malformations of the heart and great vessels that result from abnormal embryonic and fetal development. They comprise a broad spectrum of anatomic and clinical severity, ranging from relatively simple lesions such as atrial and ventricular septal defects to highly complex malformations of the cardiac chambers and their venous and arterial connections. As a group, cardiovascular malformations are the most common form of major birth defect, occurring at a frequency of 8 per 1000 live births (excluding bicuspid aortic valve, which has an estimated prevalence of 2% to 3%). They are also among the most frequent causes of death in infants beyond the neonatal period.

The genetic basis of CHD has been the subject of intense investigation and increasing understanding in the past decade. CHD has long been recognized as a key feature of certain genetic syndromes, a prime example being Down syndrome (trisomy 21). In contrast, the importance of genetic mutation in non-syndromic CHD has emerged relatively recently through the elucidation of single gene defects in certain families with inherited CHD. This work, undertaken largely before the completion of the Human Genome Project, has relied principally on conventional genetic approaches, that is linkage analysis and positional cloning, as well as candidate gene sequencing. As it happens, however, families with mendelian transmission of CHD are actually quite rare. Non-mendelian patterns with variable penetrance and expressivity are more the rule in families with manifestly inherited CHD. Furthermore, most cases of CHD are, or at least appear to be, sporadic, occurring in patients with no known family history of CHD. On this basis there is growing expectation that the bulk of CHD will ultimately be attributed to more complex genetics involving polymorphisms, modifying loci, and gene-gene and gene-environment interactions. Ongoing investigation of these mechanisms relies necessarily on genomic approaches.

This chapter reviews the current knowledge of the molecular genetics and genomics of CHD. Future directions for this research, and the associated challenges, are also considered. Recommendations for genetic evaluation of patients and families with CHD are offered, including screening, molecular diagnosis, and counseling. Of note, the genetics of cardiac arrhythmias and of the cardiomyopathies, both also affecting children, are beyond the scope of this chapter but are addressed elsewhere in this volume (see Chapters 28 and 29).

CHD GENE DISCOVERY BY CONVENTIONAL GENETICS

Salient human genes with established roles in the etiology of CHD are summarized in Table 31.1. This list includes the most well studied and illustrative examples of CHD genes but is by no means exhaustive. Indeed, a search of the Online Mendelian Inheritance in Man database (www.ncbi.nlm.nih.gov/omim/) yields more than 900 additional genes or loci linked to a range of CHD at this writing. CHD genes have been identified largely through conventional genetic approaches including linkage

Copyright © 2010, Elsevier Inc.
All rights reserved.

TABLE 31.1 Salient human congenital heart disease genes

Gene	Type[a]	Syndrome	Cardiac Lesions[b]	References
TBX5	TF	Holt-Oram	ASD, others	Basson et al. (1997); Li et al. (1997b); Bruneau et al. (2001)
TBX1	TF	DiGeorge	Conotruncal defects	Merscher et al. (2001); Lindsay et al. (2001); Jerome and Papaioannou (2001); Yagi et al. (2003)
NKX2-5	TF	Non-syndromic	ASD, VSD, others	Schott et al. (1998); Benson et al. (1999); McElhinney et al. (2003); Elliott et al. (2003)
GATA-4	TF	Non-syndromic	ASD, VSD, others	Garg et al. (2003)
ZFMP2/FOG2	TF	Non-syndromic	TOF	Pizzuti et al. (2003)
SALL4	TF	Okihiro	ASD, VSD, TOF	Al-Baradie et al. (2002); Kohlhase et al. (2002)
ZIC3	TF	Heterotaxy	Complex malformations	Casey et al. (1993); Gebbia et al. (1997); Ware et al. (2004)
TFAP2B	TF	Char	PDA	Satoda et al. (1999, 2000); Zhao et al. (2001)
JAG1	CS	Alagille	PS, TOF	Oda et al. (1997); Li et al. (1997a); Eldadah et al. (2001)
NOTCH2	CS	Alagille	PS, TOF	McDaniell et al. (2006)
NOTCH1	CS	Non-syndromic	BAV	Garg et al. (2005)
PTPN11	CS	Noonan	PS, HCM	Jamieson et al. (1994); Tartaglia et al. (2001, 2002)
FIBRILLIN-1	EM	Marfan	MVP, AD	Kainulainen et al. (1990); Magenis et al. (1991); Dietz et al. (1991a, b); Lee et al. (1991)
ELASTIN	EM	Williams	SVAS, others	Ewart et al. (1993)

[a]TF, transcription factor; CS, cell signaling factor; EM, extracellular matrix protein.

[b]ASD, atrial septal defect; VSD, ventricular septal defect; TOF, tetralogy of Fallot; PDA, patent ductus arteriosus; PS, pulmonary stenosis; HCM, hypertrophic cardiomyopathy; MVP, mitral valve prolapse; AD, aortic dilation; SVAS, supravalvar aortic stenosis.

analysis, positional cloning, and candidate gene sequencing. Some were discovered in the context of clinical genetic syndromes that include CHD phenotypes while others have been found in families with non-syndromic CHD. Re-sequencing of candidate genes has also demonstrated mutations or polymorphisms associated with sporadic CHD.

Transcription Factor Genes

Among the fourteen CHD genes in Table 31.1, eight encode nuclear transcription factors. Five of these, Tbx5, *Nkx2-5*, *GATA4*, *Zfpm2/Fog2*, and *Sall4*, have been demonstrated to interact cooperatively with each other to regulate gene expression during cardiac embryogenesis.

TBX5 (Holt-Oram Syndrome)

Holt-Oram syndrome is an autosomal dominant disorder characterized by skeletal malformations of the upper extremities and CHD, most commonly secundum atrial septal defects but also ventricular septal defects and tetralogy of Fallot. *TBX5*, encoding a T-box nuclear transcription factor, was identified as the disease gene using linkage analysis and detailed positional cloning, and several missense and nonsense mutations were identified in multiple affected families (Basson et al., 1997; Li et al., 1997b). *TBX5* was found expressed in embryonic human heart and limb, and targeted deletion of *Tbx5* in the mouse resulted in defective cardiac development, supporting the conclusion that human *TBX5* mutation is responsible for the CHD in Holt-Oram patients (Basson et al., 1997; Bruneau et al., 2001; Li et al., 1997b).

TBX1 (DiGeorge Syndrome)

DiGeorge syndrome, now synonymous with velo-cardio-facial syndrome, Shprintzen syndrome, and CATCH22, involves a range of variable clinical phenotypes including CHD, neonatal hypocalcemia, cellular immune deficiency, palatal defects, characteristic facies, and cognitive and psychiatric disorders. The cardiac defects are typically conotruncal or outflow tract malformations associated with defective neural crest migration, including interrupted aortic arch, truncus arteriosus, and tetralogy of Fallot, especially tetralogy with pulmonary atresia (McDonald-McGinn et al., 1999). This is a well established contiguous gene syndrome caused by microdeletion of a critical region located at chromosome 22q11.2 (a minority of DiGeorge patients have an alternate deletion at 10p14–13). The human *TBX1* gene,

encoding another T-box transcription factor and expressed in neural crest and the developing cardiac outflow tract (conotruncus), was mapped to the center of the DiGeorge critical region (Chieffo et al., 1997). Mice with targeted heterozygous *Tbx1* deletion developed cardiac outflow tract malformations and certain extra-cardiac phenotypes characteristic of human DiGeorge syndrome (Jerome and Papaioannou, 2001; Lindsay et al., 2001; Merscher et al., 2001). Furthermore, both the cardiac and extracardiac phenotypes could be rescued by the introduction of an extrachromosomal copy of *Tbx1*. These findings lend strong support to the hypothesis that human *TBX1* mutations may be sufficient to cause the cardiac defects that are characteristic DiGeorge syndrome and, indeed, similar lesions in patients who do not have the syndrome. However, rather surprisingly, isolated *TBX1* mutations or deletions have not been found in large numbers of non-deleted patients with conotruncal malformations (Conti et al., 2003; Gong et al., 2001). There is a sole report of *TBX1* frameshift and missense mutations in a small kindred and two unrelated individuals with DiGeorge-like phenotypes and no 22q11.2 deletion (Yagi et al., 2003). Thus, a role for *TBX1* mutation in non-syndromic conotruncal malformations seems very likely but awaits more definitive proof.

NKX2-5

The mammalian homeobox transcription factor Nkx2-5 was first identified as a homolog of the Drosophila *tinman* gene, essential for specification of heart muscle progenitors (Bodmer, 1993; Komuro and Izumo, 1993). Targeted disruption of *Nkx2-5* in the mouse was found to cause arrested cardiac development at the linear heart tube stage (Lyons et al., 1995). Subsequent studies showed that Nkx2-5 interacts with Tbx5, the Holt-Oram factor, to promote cardiomyocyte differentiation (Hiroi et al., 2001). Human *NKX2-5* was recognized as a likely candidate gene in the linked region on chromosome 5q35 following genome-wide scans in families with non-syndromic, multigenerational CHD (Schott et al., 1998). Direct sequencing identified a series of mutations that segregated with the CHD. Subsequently, re-sequencing in large cohorts of CHD patients has revealed an appreciable frequency of *NKX2-5* mutations or polymorphisms in association with sporadic, apparently non-familial CHD involving a broad range of cardiac defects (Benson et al., 1999; Elliott et al., 2003; Goldmuntz et al., 2001; McElhinney et al., 2003). Somatic mutations of *NKX2-5* have also been identified in the hearts of patients with cardiac malformations although causation has not been established (Reamon-Buettner et al., 2004).

GATA-4

GATA-4, one of a family of closely related zinc-finger transcription factors, is expressed in the developing heart. Targeted homozygous deletion of *Gata4* in the mouse abrogated normal formation of the primitive heart tube in the early embryo (Kuo et al., 1997; Molkentin et al., 1997), while haploinsufficiency and conditional myocardial deletion resulted in a range of cardiac malformations (Pu et al., 2004; Zeisberg et al., 2005). Human *GATA4* was identified as a candidate gene in an interval at 8p22–23 that was linked to CHD, principally atrial septal defects, in a genome-wide scan in a large multigenerational kindred (Garg et al., 2003). Direct sequencing identified mutations that segregated with non-syndromic CHD in this and a second independent kindred. GATA4 interacts with Tbx5, the Holt-Oram factor, and with Nkx2-5 to regulate cardiac gene expression, and disease-causing mutations abrogate these interactions (Durocher et al., 1997; Garg et al., 2003). Somatic mutations of *GATA4* have also been identified in the hearts of patients with cardiac malformations although again causation has not been established (Reamon-Buettner and Borlak, 2005).

ZFPM2/FOG2

FOG2 or ZFPM2, one of a family of zinc finger transcription factors, is expressed in the embryonic and adult heart, co-localizes with GATA-4, and interacts physically with GATA-4 to downregulate its activity (Svensson et al., 1999). Mice with targeted homozygous deletion of *Zfpm2/Fog2* develop cardiac lesions including tetralogy of Fallot, complete atrioventricular canal, and tricuspid atresia (Svensson et al., 2000; Tevosian et al., 2000). In a candidate gene approach, two missense mutations in *ZFPM2/FOG2* were identified in two of 47 patients with non-syndromic tetralogy of Fallot (Pizzuti et al., 2003). However, these mutations had at most a slight effect on protein function *in vitro* and, moreover, co-existing mutations in either *GATA-4* or *NKX2-5* were not excluded. Additional investigation will be required if *ZFPM2/FOG2* is to be implicated more definitively in CHD.

SALL4 (Okihiro Syndrome)

Okihiro syndrome is an autosomal dominant disorder involving multiple congenital anomalies including a characteristic eye movement disorder (Duane anomaly), forelimb defects, deafness, and CHD, typically atrial septal defects, ventricular septal defects, or tetralogy of Fallot. Linkage analysis in three affected pedigrees mapped the disease gene to a 21.6 cM region on chromosome 20 containing *SALL4*, a novel member of a zinc finger transcription factor family, in which nonsense and frameshift mutations were identified (Al-Baradie et al., 2002). Another group, working contemporaneously, also identified *SALL4* as the Okihiro disease locus using a candidate gene approach when they recognized certain non-cardiac phenotypic similarities between Okihiro syndrome and another syndrome caused by mutations in *SALL1* (Kohlhase et al., 2002). Recent work has shown that in the developing mouse, *Sall4* expression is regulated by *Tbx5*, the Holt-Oram gene, and the two interact in cardiac and forelimb patterning (Koshiba-Takeuchi et al., 2006).

ZIC3 (Heterotaxy)

Heterotaxy refers to ambiguous visceral situs in which the positions of normally asymmetric organs such as the liver and spleen are disordered with respect to the midline. Cardiac lesions associated with heterotaxy are often among the most severe

encountered, including complex malformations of the atria and ventricles and their venous and arterial connections. Many cases of heterotaxy appear to be sporadic but familial cases, notably with X-linked transmission, have been described. Linkage analysis in one such family mapped the heterotaxy phenotype to a locus at Xq24-q27.1 (Casey et al., 1993), and mutation in the gene *ZIC3*, a zinc finger transcription factor, was identified by positional cloning (Gebbia et al., 1997). Subsequently, additional *ZIC3* mutations have been identified by re-sequencing in other heterotaxy kindreds and isolated individuals with similar CHD phenotypes (Ware et al., 2004).

TFAP2B (Char Syndrome)

Char syndrome is an autosomal dominant disorder comprising facial dysmorphism, hand anomalies, and patent ductus arteriosus. Linkage analysis in two large multigenerational kindreds mapped the phenotype to a 3.1 cM region on chromosome 6 (Satoda et al., 1999). Candidate gene analysis then revealed missense mutations in *TFAP2B*, encoding a transcription factor expressed in neural crest, a contributor to the embryogenesis of the ductus arteriosus (Satoda et al., 2000). Examination of genotype–phenotype correlations among several *TFAP2B* mutations revealed that those in the DNA-binding domain of the protein produce the full Char phenotype while one in the trans-activation domain is associated with patent ductus arteriosus but only mild facial and no limb abnormalities (Zhao et al., 2001).

Cell Signaling Genes

Four CHD genes in Table 31.1 encode proteins involved in cell signal transduction, three of which are in the Notch pathways that regulate cell fate decisions during embryonic development.

JAG1 and *NOTCH2* (Alagille Syndrome)

Alagille syndrome is an autosomal dominant disorder comprising hepatic, cardiac, ocular, and vertebral abnormalities in addition to a characteristic facial appearance (Alagille et al., 1987). The cardiac defects are typically right-sided, with valvar pulmonary stenosis the single most common finding and tetralogy of Fallot the most common complex lesion. Linkage analysis and positional cloning in a series of affected families led to the definition of a locus on chromosome 20p12 that included *JAG1*, encoding a Notch ligand, in which a series of mutations were identified (Li et al., 1997a; Oda et al., 1997). Analysis of *JAG1* during mammalian embryogenesis showed expression in the developing heart and vessels, consistent with a critical role in patterning of the right heart and pulmonary vasculature (Loomes et al., 1999). Subsequently, in a candidate gene approach, a *JAG1* missense mutation was identified as the cause of non-Alagille autosomal dominant tetralogy of Fallot segregating in a large kindred, shown in Figure 31.1 (Eldadah et al., 2001). This finding further validated the hypothesis that genes associated with certain cardiac phenotypes in genetic syndromes may also be found to be mutated in non-syndromic patients with similar cardiac defects. Recently, in a candidate gene approach, mutations in the JAG1 receptor *NOTCH2* were found to segregate with cardiac and other phenotypes in two rare Alagille families without JAG1 mutations, strengthening the understanding that these phenotypes are attributable to defective Notch pathway signaling (McDaniell et al., 2006).

NOTCH1 (Bicuspid Aortic Valve)

Bicuspid (two-leaflet) aortic valve is the most prevalent congenital cardiac malformation, estimated to occur in as many as 2–3% of individuals. It is inconsequential in many children, but associated with stenosis of the valve in others. Further, bicuspid valves have a predilection to calcify in later life leading to progressive aortic stenosis and regurgitation. Although most cases appear to be sporadic, families with autosomal dominant bicuspid aortic valve have been identified. Whole genome linkage analysis in one such family and additional studies in other kindreds led to the identification of mutations in the *NOTCH1* transmembrane receptor that appear to be responsible both for the congenital valve deformity and later de-repression of osteoblast calcium deposition in these valves (Garg et al., 2005).

PTPN11 (Noonan Syndrome)

Noonan syndrome is an autosomal dominant disorder comprising dysmorphic facial features, skeletal malformations, short stature, and cardiac abnormalities of which pulmonary stenosis and hypertrophic cardiomyopathy are most characteristic. Genome-wide linkage analysis in a large kindred mapped the Noonan locus to the long arm of chromosome 12 (Jamieson et al., 1994). *PTPN11*, encoding the SHP-2 intracellular protein tyrosine phosphatase, was identified in this region and, in a candidate gene approach, was found to be mutated in approximately 50% of Noonan cases (Tartaglia et al., 2001); a locus or loci for the remaining cases remains to be mapped. Genotype-phenotype analyses showed that pulmonary stenosis was seen more frequently, and hypertrophic cardiomyopathy less frequently, in Noonan patients with *PTPN11* mutations compared to those without (Tartaglia et al., 2002).

Extracellular Matrix Protein Genes

Mutations in two genes encoding extracellular matrix proteins cause congenital syndromes involving arteriopathies of different forms.

FIBRILLIN-1 (Marfan Syndrome)

Marfan syndrome is an autosomal dominant disease of connective tissue principally involving the cardiovascular, skeletal, and ocular systems. Cardiovascular manifestations include mitral valve prolapse and regurgitation, presenting in infancy in the most severe cases, and progressive aneurismal dilation of the aortic root with the potential for catastrophic aortic dissection and rupture. Marfan syndrome was first mapped to chromosome 15 using traditional genetic linkage analysis (Dietz et al., 1991b; Kainulainen et al., 1990). At about the same time, immunohistochemical studies performed on skin and dermal fibroblasts from Marfan patients revealed that abnormal microfibrils were

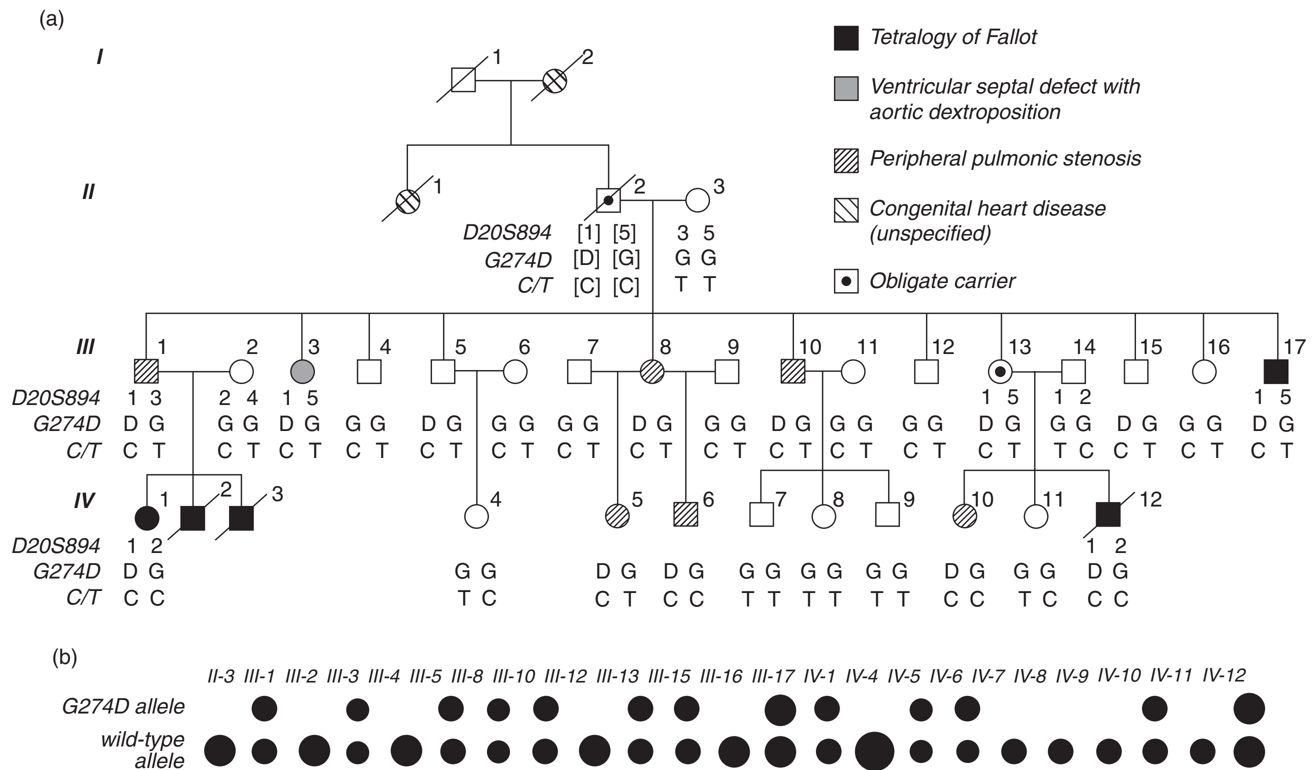


Figure 31.1 Alagille kindred. (a) Pedigree of four generations of a family with Alagille syndrome. Haplotypes include marker alleles of *D20S894* (1–5), glycine (G) or aspartic acid (D) at the mutation site (G274D), and cytosine (C) or thymidine (T) at nucleotide 765, a silent polymorphism (information in brackets was inferred). All individuals with a clinical phenotype inherited the D274 allele. (b) Genomic DNA from the indicated individuals was hybridized with radiolabeled probes specific for the mutant G274D and wild-type coding sequences. (Reprinted, with permission, from Eldadah et al., 2001.)

deficient in the protein fibrillin (Hollister et al., 1990). Then, taking a candidate gene approach, it was possible to localize *FIBRILLIN-1* to 15q21.1 (Magenis et al., 1991) and to demonstrate conclusively that mutations in this gene were tightly linked to the Marfan phenotype in several affected families (Dietz et al., 1991a; Lee et al., 1991). Recognizing this, fibrillin has long been assumed to be critical in the aortic wall and other connective tissues as a structural protein. However, recent work has revealed that fibrillin has a regulatory role in TGF-β signaling and dysregulation of this pathway may instead underlie Marfan pathogenesis (Habashi et al., 2006; Neptune et al., 2003).

ELASTIN (Williams Syndrome)

Williams or Williams-Beuren syndrome is a contiguous gene syndrome involving a 1.5–2 Mb microdeletion of chromosome 7q11.23. The phenotype comprises characteristic endocrine, cognitive, and facial features in association with areas of arterial narrowing, most typically supravalvar aortic stenosis (SVAS). SVAS is a constriction of the proximal aorta above the valve that increases the pressure work of the left ventricle, leading potentially to ventricular hypertrophy and congestive heart failure. On histopathological examination the architecture of the aortic wall elastic lamina is disrupted. The specific gene responsible for SVAS in Williams syndrome remained unknown until a non-syndromic form of autosomal dominant SVAS was recognized in a number of families. Linkage studies and candidate gene analysis in one such kindred identified *ELASTIN* as the disease gene in non-syndromic SVAS (Ewart et al., 1993). As this gene resides in the Williams critical region on 7q11.23, *ELASTIN* hemizygosity appears to underlie the characteristic arteriopathy of Williams syndrome.

Chromosomal Aneuploidies Associated with CHD

Down syndrome (trisomy 21) is the most prevalent genetic syndrome causing congenital heart malformations. It is most closely associated with endocardial cushion defects such as complete atrioventricular canal. The gene or genes on chromosome 21 that underlie the predilection for cardiac defects in Down syndrome are the subject of ongoing intense investigation (Reeves, 2006). Putative cardiac genes for trisomy 13 and trisomy 18, much less common than trisomy 21 but also involving heart defects, similarly remain to be elucidated. In Turner syndrome (45,XO), haploinsufficiency of one or more as yet unidentified genes on the X chromosome is inferred to cause coarctation of the aorta, the cardiovascular lesion most frequently associated with this condition.

GENOMIC STRATEGIES FOR CHD GENE DISCOVERY

It is evident from the foregoing overview that conventional genetic methods have been effective tools in the discovery of CHD genes and mutations. However, these methods are limited to the extent that they allow, in general, for the study of only one or a few genes at a time. The advent of genomic technologies has afforded opportunities in CHD, as in other fields, to analyze thousands of genes if not the entire genome simultaneously with the promise of accelerated discovery.

Large Scale cDNA Sequencing

The first systematic effort to sequence large numbers of cardiac cDNAs was pioneered in the early 1990s by Liew and coworkers (Hwang et al., 1994; Liew et al., 1994). A cDNA library was constructed from human fetal heart and more than 3000 clones identified, nearly half of which were novel at the time by comparison to existing public databases. Early experiments in array transcription profiling, in the form of dot blot filters, demonstrated differential fetal versus adult expression of some of these genes. These investigators expanded their analysis to encompass multiple cardiac libraries and found more than 43,000 expressed sequence tags using automated sequencing; an additional 41,000 were culled *in silico* from existing databases (Hwang et al., 1997). Further bioinformatics analyses allowed for the identification of potential tissue-, developmental stage-, and disease-specific expression of subsets of these genes.

Subtractive Hybridization and Differential Display

Understanding that developmental stage-specific genes are crucial to cardiac morphogenesis, additional genomic strategies have been brought to bear to identify them. Srivastava and coworkers developed a novel subtractive hybridization technique, termed subtractive and selective PCR amplification, to select genes expressed specifically in the earliest embryonic cardiac precursor cells (Gottlieb et al., 2002). Among these was *Bop*, a transcriptional repressor that they then showed to be necessary for normal development of the right ventricle, potentially through regulation of the HAND2 transcription factor, also critical for formation of the right ventricle (Srivastava et al., 1997). These and other investigators also used subtractive and differential display techniques to compare gene expression in *Hand2*$^{-/-}$ versus wild-type mice and elucidate cardiac molecular pathways downstream of HAND2 (Yamagishi et al., 1999). One such pathway gene mapped to human chromosome 10p13-p14, which is deleted in patients with a variant of DiGeorge syndrome that also includes cardiac malformations (Villanueva et al., 2002).

Microarray Transcriptional Profiling

More recently microarray technology has been used to study gene expression patterns related to CHD. Some of these experiments have focused on regional differences in cardiac gene expression in the early vertebrate embryo or in the adult heart (Afrakhte and Schultheiss, 2004; Zhao et al., 2002). Quertermous and coworkers undertook a comprehensive analysis of gene expression patterns in the four chambers and interventricular septum of the mouse heart using a 42,300 element mouse cDNA microarray representing more than 25,000 unique genes and expressed sequence tags (Tabibiazar et al., 2003). This strategy yielded large sets of chamber-specific genes, some in gene families well established to be involved in cardiac development and others more novel.

In the first genome-wide array analysis comparing normal and congenitally malformed human hearts, Sperling and coworkers hybridized human unigene cDNA arrays with right ventricular and/or right atrial RNA probes prepared from normal donors and from patients with simple ventricular septal defects, tetralogy of Fallot, and other defects leading to right ventricular hypertrophy (Kaynak et al., 2003). As represented in Figure 31.2, the levels and probability values for the differentially expressed genes constitute lesion- and chamber-specific genomic fingerprints. Similar microarray analyses of gene expression in tetralogy of Fallot and other lesions producing right ventricular outflow obstruction have followed (Konstantinov et al., 2004; Sharma et al., 2006). The identification of transcriptional patterns associated with particular congenital malformations is intriguing. The extent to which these reflect the primary genetic etiologies of these lesions versus secondary perturbations – due to cyanosis, pressure overload, volume overload, chamber interactions, for example – merits further investigation.

Microarray analysis is an increasingly important tool for the elucidation of molecular pathways that lie downstream of key transcriptional regulators of cardiac development and CHD, including Nkx2-5, Tbx5, and GATA6. In a study of gene expression in the myopathic ventricles of conditional *Nkx2-5*$^{-/-}$ knockout mice, genes associated with myocardial cell proliferation and trabeculation were markedly dysregulated (Pashmforoush et al., 2004). Among these was a gene previously established to lie downstream of *Nkx2-5*, as well as a series of others newly associated with this pathway. Other investigators have profiled gene expression in a mouse model of DiGeorge syndrome with *Tbx1* haploinsufficiency (Prescott et al., 2005) and in cardiac-derived cell lines with or without overexpression of *TBX5* or *GATA6*, revealing multiple putative downstream target genes in each pathway (Plageman and Yutzey, 2006; Alexandrovich et al., 2006).

Mutagenesis and Phenotypic Screens

Genomic strategies for CHD gene discovery are complemented by "phenomic" approaches in the mouse. The Jackson Laboratory has undertaken a phenotype-driven project to identify new murine mutations leading to the development of cardiovascular disease (Svenson et al., 2003). In collaboration with the Laboratory of Developmental Biology at the National Heart Lung and Blood Institute, non-invasive two-dimensional and Doppler echocardiography was used to screen 7546 mouse fetuses from

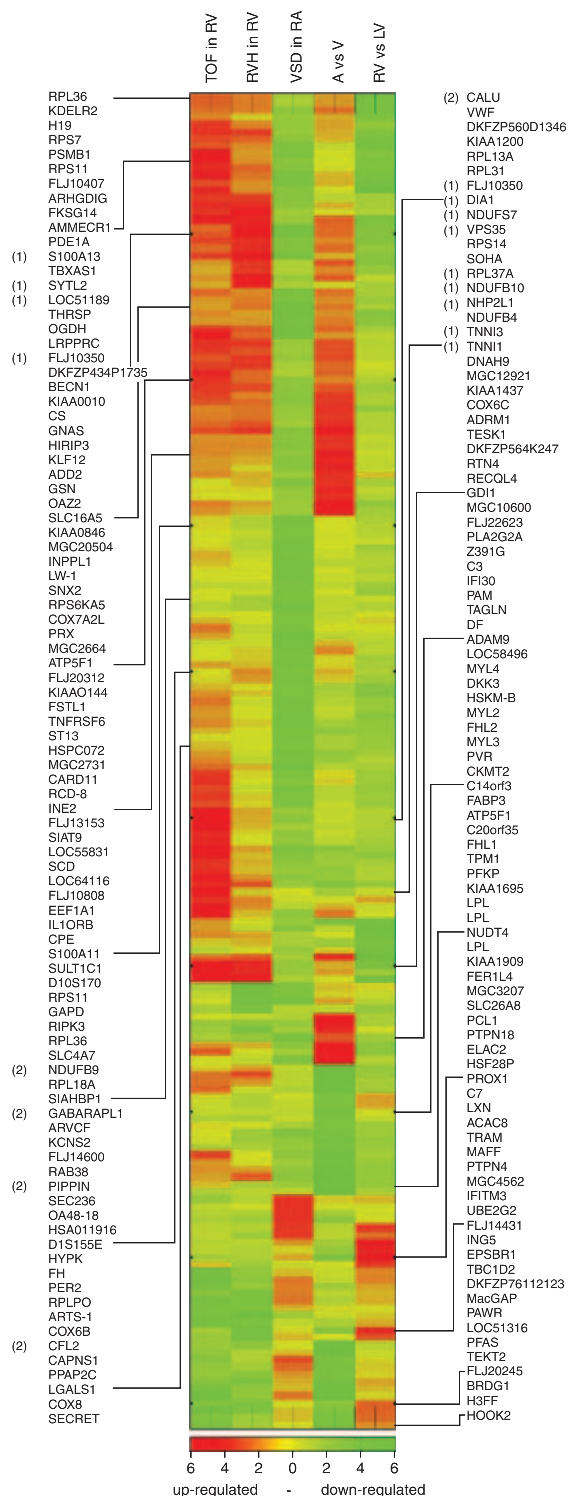


Figure 31.2 Transcription profiling. Graphic display of gene expression levels and probability values profiled across several cardiac phenotypes using microarray analysis. Gene names are listed. Each comparison is represented by a column and each gene by a row. The $-\log_{10}(P)$ for each gene in each comparison is color coded in yellow-to-red for upregulated and yellow-to-green for down-regulated genes, for example a value of 2 represents $P=0.01$ and is color coded in red if the gene is upregulated. (Reprinted, with permission, from Kaynak et al., 2003.)

262 chemically (*N*-ethyl-*N*-nitrosourea) mutagenized families (Yu et al., 2004). Remarkably, congenital heart defects were identified in 124 of these families, among them several with virtual phenocopies of human genetic syndromes including Holt–Oram and DiGeorge syndromes. Chromosomal localization of these mutations using polymorphic microsatellite markers revealed that they mapped to loci other than *Tbx5* and *Tbx1*, the genes previously linked respectively to these syndromes. Thus it is likely that novel CHD genes, some in established molecular pathways and others in entirely new pathways, will be discovered in these and other chemically mutagenized mouse models.

CYTOGENETIC AND MOLECULAR GENETIC TESTING

Cytogenetic and molecular techniques used in the discovery of genes associated with CHD also have practical utility when evaluating the underlying cause of CHD in a patient. Table 31.2 provides a comparison of available diagnostic genetic testing platforms and their applications in CHD.

Structural chromosome anomalies may explain a finding of CHD in conjunction with other congenital anomalies or developmental issues. A standard metaphase karyotype (450–550 bands) is useful for identifying extra or missing chromosomes, whereas high-resolution banding (550–800 bands) is effective at identifying more subtle structural abnormalities including deletions, duplications, translocations, and inversions.

Fluorescence *in situ* hybridization (FISH), another cytogenetic technique, has greater specificity and resolution for identification of microdeletions or duplications that could be missed in a karyotype analysis. FISH is routinely used to detect microdeletion syndromes that are associated with CHD including velocardiofacial syndrome (deletion of 22q11.2) and Williams–Beuren syndrome (deletion of 7q11.23).

Array-based comparative genomic hybridization (CGH) is increasingly performed as an adjunct to standard high-resolution karyotype analysis. Array-CGH can detect abnormalities as small as 80 kb whereas high-resolution chromosome analysis has a minimum resolution of 5 mb. Array-CGH typically includes subtelomeric evaluation to analyze the most distal ends of the chromosomes. Array-CGH cannot detect balanced rearrangements or abnormalities in the genome that are not covered by the selected DNA segments in the microarray. However, array-CGH may be a useful means of evaluating a patient with CHD for multiple genetic abnormalities at once.

Lastly, targeted mutation analysis of a known gene, whether by sequencing, array hybridization, or multiplex ligation-dependent probe amplification (MLPA), can be useful in confirming a clinical diagnosis in a patient whose features are highly suggestive of a particular genetic syndrome. Targeted mutation analysis is complicated by locus heterogeneity, where it is likely that other causal genes exist but have not been discovered. Thus,

TABLE 31.2 Cytogenetic and molecular diagnostic testing in CHD

Assay platform	Genetic abnormality	Resolution	Examples
Standard metaphase karyotype (450–550 bands)	Aneuploidies (monosomies, trisomies)	5–10 Mb	Trisomy 13, 18, 21
Karyotype with high-resolution banding (550–850 bands)	Chromosome structural anomalies: duplications, translocations, interstitial or terminal deletions	3–5 Mb	Trisomy 13, 18, 21, other chromosome structural anomaly syndromes
Fluorescence *in situ* hybridization (FISH)	Subtle chromosome structural anomalies: microdeletions, microduplications, subtle translocations	30–40 kb in clinical setting; as small as 1 kb in research setting	Deletion syndromes such as Velocardiofacial Syndrome (del22q11.2), Williams-Beuren Syndrome (del7q11.23), some instances of Noonan syndrome (*PTPN11*)
Array comparative genomic hybridization (array-CGH)	Identifies deletions and duplications in regions of the genome that are represented on the microarray	~80 kb clinical setting, smaller in research setting	Ability to evaluate multiple deletion/duplication syndromes simultaneously
Multiplex ligation-dependent probe amplification (MLPA)	Detects very small genetic aberrations, to the level of a single nucleotide change	Single base pair alterations	Small deletions below level of detection by FISH
Targeted DNA mutation detection (sequencing)	Single gene mutations, including point mutations, deletions, duplications within the gene	Single base pair alterations	Holt-Oram syndrome (*TBX5*), Alagille syndrome (*JAG1*), Char syndrome (*TFAP2B*), Noonan syndrome (*PTPN11*)

a negative test does not rule out the condition and provides limited recurrence risk information to the family.

MEDICAL EVALUATION AND COUNSELING RECOMMENDATIONS

A thorough medical evaluation, including family history, physical exam, and genetic testing, can aid in distinguishing isolated or sporadic CHD from syndromic CHD. This distinction can provide critical information for cardiac management as well as screening for extra-cardiac issues. Identification of an underlying genetic cause provides answers to the family and aids in recurrence risk counseling. Lastly, given the variability in phenotypic expression of many of the genetic syndromes associated with CHD, knowing the genetic cause and inheritance pattern helps identify at-risk relatives who may need cardiac evaluation and potential intervention. The following approach is recommended to the treating clinician, also outlined in Figure 31.3.

Family history: A three-generation family history can help to direct genetic testing. Documentation of details related to CHD in other family members is critical. It may be useful to request cardiac evaluations of other family members to rule out subtle cardiac defects. Additionally, it is important to document any history of other birth defects, learning disabilities, mental retardation, multiple miscarriages, and still births.

Physical exam and clinical imaging: During the physical exam, attention should be paid to subtle dysmorphic facial features, as well as the involvement of other organ systems. The exam should also focus on ear and eye abnormalities, skeletal issues such as limb reduction defects, gastrointestinal and urologic defects. Standard radiological studies may provide unique diagnostic clues; spine anomalies, aortic arch anomalies, and stomach situs can often be seen in a routine chest radiograph. For example, a right aortic arch, in combination with the diagnosis of tetralogy of Fallot and a family history of learning disabilities is suggestive of velocardiofacial syndrome.

Genetic testing: If the diagnosis of CHD was made prenatally, the patient may have had a standard karyotype performed via amniocentesis to rule out obvious chromosomal abnormalities. It is reasonable to order a high-resolution chromosome analysis if this has not been done, particularly if the infant or child has dysmorphic features or other congenital anomalies. If a chromosome abnormality is detected, it is recommended that the family consult with a genetic specialist for counseling and evaluation of other family members.

Genetic counseling: In the case of syndromic CHD, the clinician can provide referrals to appropriate specialists who can help evaluate and manage non-cardiac manifestations. A discussion about the genetic cause, inheritance pattern and available genetic testing for at-risk family members is imperative, as well as options for prenatal diagnosis in future pregnancies.

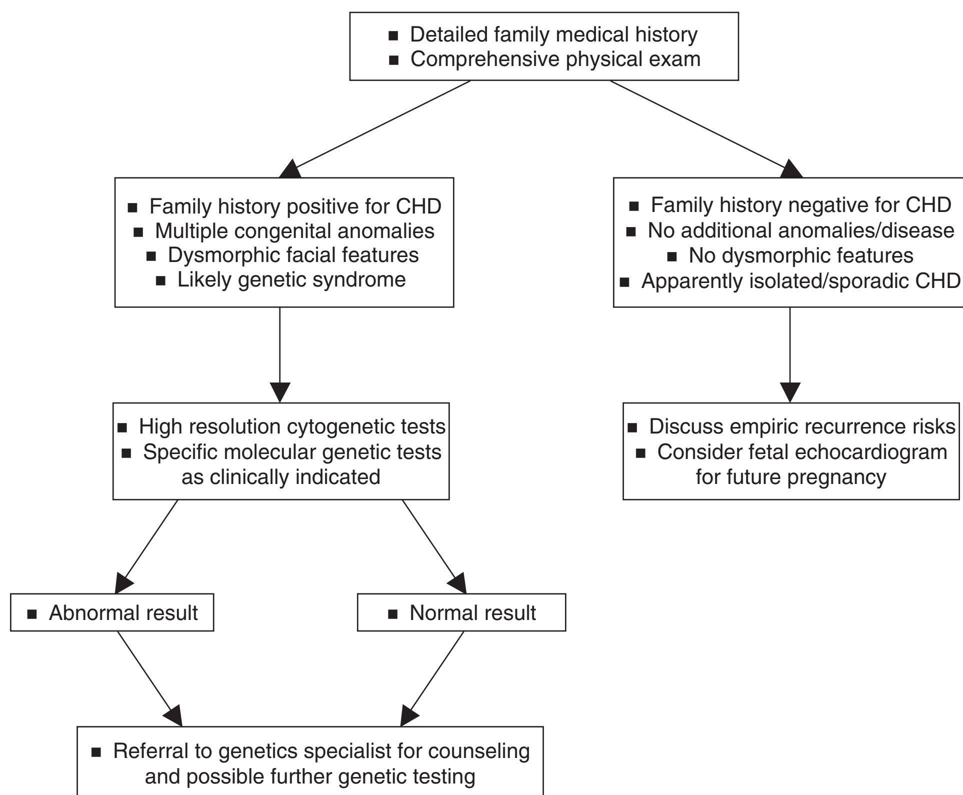


Figure 31.3 Strategy for genetic evaluation of patient with CHD.

A referral to a genetics specialist may be valuable in addressing the various questions and concerns of these families. If the family history is not indicative of familial CHD and the individual does not have features that are suggestive of a syndromic cause, the clinician is encouraged to discuss empiric evidence of an increased risk of CHD in subsequent children. The family can also be informed about fetal echocardiography to monitor future pregnancies.

2009 UPDATE

Genomic studies in humans and model organisms have continued to yield new knowledge of the genetic underpinnings of CHD, as well as the molecular pathways that ensue. A third T-box transcription factor, TBX20, known for some time to play a key role in murine cardiac development via interaction with NKX2-5 and GATA-4, has now been implicated in human cardiac malformations. Germline missense and nonsense *TBX20* mutations are associated with familial CHD, comprising a spectrum of septation, chamber, and valvular abnormalities (Kirk et al., 2007). *TBX20* mutations have subsequently been identified in sporadic cases of tetralogy of Fallot, atrial septal defect, and total anomalous pulmonary venous connection (Liu et al., 2008). *TBX20* transcripts are upregulated in cardiac biopsies of patients with tetralogy of Fallot hearts as compared with simple ventricular septal defects and normal controls (Hammer et al., 2008).

Further studies of GATA-4 in murine models and human CHD have also been revealing (Rajagopal et al., 2007). Mice with heterozygous *GATA-4* mutation developed a range of cardiac phenotypes including atrial and ventricular septal defects, endocardial cushion defects, and right ventricular hypoplasia. In human candidate gene analyses, *GATA-4* mutations were found associated with an overlapping spectrum of cardiac malformations including atrial septal defects, endocardial cushion defects, and right ventricular hypoplasia with double-inlet left ventricle. These mutations were absent in a large control cohort and resulted in nonconservative amino acid substitutions in key protein domains, strengthening the assertion that these are indeed causative CHD mutations.

Novel sequence variants in *CFC1*, encoding a cofactor for NODAL-related signaling in early embryogenesis and

previously associated with laterality defects, have recently been identified in heterotaxy patients with associated cardiac malformations including double-outlet right ventricle (Selamet Tierney et al., 2007). Transmission of some of these variant alleles from unaffected parents, and the presence of others in unaffected controls, indicates incomplete penetrance and suggests a role for other factors in the development of these phenotypes. The importance of complex genetic interactions has been highlighted by functional genomics investigation of the NODAL pathway comprising *FOXH1*, *SMAD2*, *GDF1*, and *TDGF1*, in addition to *CFC1* and *NODAL* itself (Roessler et al., 2008). Sequence variants in several of these component genes combine to reduce NODAL signaling strength. Further, loss-of-function mutations of *GDF1* alone have been associated with human CHD (Karkera et al., 2007).

The discovery of chromosomal rearrangements associated with CHD provides a key opportunity to identify potential new CHD genes at translocation, duplication, and deletion breakpoints (Behjati et al., 2008; Mefford et al., 2008). In addition to conventional cytogenetic analysis, the technique of array comparative genomic hybridization has revealed heretofore undetected submicroscopic chromosomal aberrations in patients with CHD and other birth defects (Erdogan et al., 2008, Lu et al., 2008; Richards et al., 2008).

The next phase of genomic research in CHD resides in large measure in the elucidation of complex genetic and epigenetic interactions. The effects of complex genetics are evident in mice with *GATA-4* mutation in which the cardiac phenotype is strongly influenced by the genetic background in different lines (Rajagopal et al., 2007). It is likely that such interactions are responsible for the incomplete penetrance and variable expressivity that characterize nearly all known CHD mutations, including *TBX20*, *GATA-4*, and *CFC-1* (Kirk et al., 2007; Rajagopal et al., 2007; Selamet Tierney et al., 2007).

Several informative reviews of CHD genetics have been published recently (Andelfinger, 2008; Bruneau, 2008; Pierpont et al., 2007).

2009 UPDATE REFERENCES

Andelfinger, G. (2008). Genetic factors in congenital heart malformation. *Clin Genet* 73, 516–527.

Behjati, F., Shafeghati, Y., Kahrizi, K., Firouzabadi, S.G., Najmabadi, H., Dixon, N. and Davies, A.F. (2008). Interstitial deletion of the short arm of chromosome 10 del(10)(p11.2p12.32) in a patient with congenital heart disease, minor dysmorphism, and mental retardation. *Am J Med Genet A* 146A, 3223–3226.

Bruneau, B. (2008). The developmental genetics of congenital heart disease. *Nature* 451, 943–948.

Erdogan, F., Larsen, L.A., Zhang, L., Tümer, Z., Tommerup, N., Chen, W., Jacobsen, J.R., Schubert, M., Jurkatis, J., Tzschach, A. et al. (2008). High frequency of submicroscopic genomic aberrations detected by tiling path array comparative genomic hybridization in patients with isolated congenital heart disease. *J Med Genet* 11, 704–709.

Hammer, S., Toenjes, M., Lange, M., Fischer, J.J., Dunkel, I., Mebus, S., Grimm, C.H., Hetzer, R., Berger, F. and Sperling, S. (2008). Characterization of TBX20 in human hearts and its regulation by TFAP2. *J Cell Biochem* 104, 1022–1033.

Karkera, J.D., Lee, J.S., Roessler, E., Banerjee-Basu, S., Ouspenskaia, M.V., Mez, J., Goldmuntz, E., Bowers, P., Towbin, J., Belmont, J.W. et al. (2007). Loss-of-function mutations in growth differentiation factor-1 (GDF1) are associated with congenital heart defects in humans. *Am J Hum Genet* 81, 987–994.

Kirk, E.P., Sunde, M., Costa, M.W., Rankin, S.A., Wolstein, O., Castro, M.L., Butler, T.L., Hyun, C., Guo, G., Otway, R. et al. (2007). Mutations in the cardiac T-box factor gene TBX20 are associated with diverse cardiac pathologies, including defects of septation and valvulogenesis and cardiomyopathy. *Am J Hum Genet* 81, 280–291.

Liu, C., Shen, A., Li, X., Jiao, W., Zhang, X. and Li, Z. (2008). T-box transcription factor TBX20 mutations in Chinese patients with congenital heart disease. *Eur J Med Genet* 51, 580–587.

Lu, X.Y., Phung, M.T., Shaw, C.A., Pham, K., Neil, S.E., Patel, A., Sahoo, T., Bacino, C.A., Stankiewicz, P., Kang, S.H. et al. (2008). Genomic imbalances in neonates with birth defects: high detection rates by using chromosomal microarray analysis. *Pediatrics* 122, 1310–1318.

Mefford, H.C., Sharp, A.J., Baker, C., Itsara, A., Jiang, Z., Buysse, K., Huang, S., Maloney, V.K., Crolla, J.A., Baralle, D. et al. (2008). Recurrent rearrangments of chromosome 1q21.1 and variable pediatric phenotypes. *N Engl J Med* 359, 1685–1699.

Pierpont, M.E., Basson, C.T., Benson, D.W., Gelb, B.D., Giglia, T.M., Goldmuntz, E., McGee, G., Sable, C.A., Srivastava, D. and Webb, C.L. (2007). Genetic basis for congenital heart defects: Current knowledge. *Circulation* 115, 3015–3038.

Rajagopal, S.K., Ma, Q., Obler, D., Shen, J., Manichaikul, A., Tomita-Mitchell, A., Boardman, K., Briggs, C., Garg, V., Srivastava, D., et al. (2007). Spectrum of heart disease associated with murine and human GATA4 mutation. *J Mol Cell Cardiol* 43, 677–685.

Richards, A.A., Santos, L.J., Nichols, H.A., Crider, B.P., Elder, F.F., Hauser, N.S., Zinn, A.R. and Garg, V Cryptic chromosomal abnormalities identified in children with congenital heart disease. *Pediatr Res* 64(4), 358–363.

Roessler, E., Ouspenskaia, M.V., Karkera, J.D., Vélez, J.I., Kantipong, A., Lacbawan, F., Bowers, P., Belmont, J.W., Towbin, J.A., Goldmuntz, E. et al. (2008). Reduced NODAL signaling strength via mutation of several pathway members including FOXH1 is linked to human heart defects and holoprosencephaly. *Am J Hum Genet* 83, 18–29.

Selamet Tierney, E.S., Marans, Z., Rutkin, M.B. and Chung, W.K. (2007). Variants of the CFC1 gene in patients with laterality defects associated with congenital heart disease. *Cardiol Young* 17, 268–274.

ACKNOWLEDGEMENTS

The authors thank Christine Seidman and Jon Seidman for helpful discussions and Kristin Tassinari for expert administrative assistance. This work was supported in part by NIH P50 HL074734 and by the Boston Children's Heart Foundation. The authors have no conflicts of interest related to this publication.

REFERENCES

Afrakhte, M. and Schultheiss, T.M. (2004). Construction and analysis of a subtracted library and microarray of cDNAs expressed specifically in chicken heart progenitor cells. *Dev Dyn* 230, 290–298.

Al-Baradie, R., Yamada, K., St Hilaire, C., Chan, W.M., Andrews, C., McIntosh, N., Nakano, M., Martonyi, E.J., Raymond, W.R., Okumura, S. et al. (2002). Duane radial ray syndrome (Okihiro syndrome) maps to 20q13 and results from mutations in SALL4, a new member of the SAL family. *Am J Hum Genet* 71, 1195–1199.

Alagille, D., Estrada, A., Hadchouel, M., Gautier, M., Odievre, M. and Dommergues, J.P. (1987). Syndromic paucity of interlobular bile ducts (Alagille syndrome or arteriohepatic dysplasia): review of 80 cases. *J Pediatr* 110, 195–200.

Alexandrovich, A., Arno, M., Patient, R.K., Shah, A.M., Pizzey, J.A. and Brewer, A.C. (2006). Wnt2 is a direct downstream target of GATA6 during early cardiogenesis. *Mech Dev* 123, 297–311.

Basson, C.T., Bachinsky, D.R., Lin, R.C., Levi, T., Elkins, J.A., Soults, J., Grayzel, D., Kroumpouzou, E., Traill, T.A., Leblanc-Straceski, J. et al. (1997). Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome. *Nat Genet* 15, 30–35.

Benson, D.W., Silberbach, G.M., Kavanaugh-McHugh, A., Cottrill, C., Zhang, Y., Riggs, S., Smalls, O., Johnson, M.C., Watson, M.S., Seidman, J.G. et al. (1999). Mutations in the cardiac transcription factor NKX2.5 affect diverse cardiac developmental pathways. *J Clin Invest* 104, 1567–1573.

Bodmer, R. (1993). The gene tinman is required for specification of the heart and visceral muscles in Drosophila. *Development* 118, 719–729.

Bruneau, B.G., Nemer, G., Schmitt, J.P., Charron, F., Robitaille, L., Caron, S., Conner, D.A., Gessler, M., Nemer, M., Seidman, C.E. et al. (2001). A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. *Cell* 106, 709–721.

Casey, B., Devoto, M., Jones, K.L. and Ballabio, A. (1993). Mapping a gene for familial situs abnormalities to human chromosome Xq24-q27.1. *Nat Genet* 5, 403–407.

Chieffo, C., Garvey, N., Gong, W., Roe, B., Zhang, G., Silver, L., Emanuel, B.S. and Budarf, M.L. (1997). Isolation and characterization of a gene from the DiGeorge chromosomal region homologous to the mouse Tbx1 gene. *Genomics* 43, 267–277.

Conti, E., Grifone, N., Sarkozy, A., Tandoi, C., Marino, B., Digilio, M.C., Mingarelli, R., Pizzuti, A. and Dallapiccola, B. (2003). DiGeorge subtypes of nonsyndromic conotruncal defects: evidence against a major role of TBX1 gene. *Eur J Hum Genet* 11, 349–351.

Dietz, H.C., Cutting, G.R., Pyeritz, R.E., Maslen, C.L., Sakai, L.Y., Corson, G.M., Puffenberger, E.G., Hamosh, A., Nanthakumar, E.J., Curristin, S.M. et al. (1991a). Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. *Nature* 352, 337–339.

Dietz, H.C., Pyeritz, R.E., Hall, B.D., Cadle, R.G., Hamosh, A., Schwartz, J., Meyers, D.A. and Francomano, C.A. (1991b). The Marfan syndrome locus: confirmation of assignment to chromosome 15 and identification of tightly linked markers at 15q15-q21.3. *Genomics* 9, 355–361.

Durocher, D., Charron, F., Warren, R., Schwartz, R.J. and Nemer, M. (1997). The cardiac transcription factors Nkx2-5 and GATA-4 are mutual cofactors. *EMBO J* 16, 5687–5696.

Eldadah, Z.A., Hamosh, A., Biery, N.J., Montgomery, R.A., Duke, M., Elkins, R. and Dietz, H.C. (2001). Familial Tetralogy of Fallot caused by mutation in the jagged1 gene. *Hum Mol Genet* 10, 163–169.

Elliott, D.A., Kirk, E.P., Yeoh, T., Chandar, S., McKenzie, F., Taylor, P., Grossfeld, P., Fatkin, D., Jones, O., Hayes, P. et al. (2003). Cardiac homeobox gene NKX2-5 mutations and congenital heart disease: associations with atrial septal defect and hypoplastic left heart syndrome. *J Am Coll Cardiol* 41, 2072–2076.

Ewart, A.K., Morris, C.A., Ensing, G.J., Loker, J., Moore, C., Leppert, M. and Keating, M. (1993). A human vascular disorder, supravalvular aortic stenosis, maps to chromosome 7. *Proc Natl Acad Sci USA* 90, 3226–3230.

Garg, V., Kathiriya, I.S., Barnes, R., Schluterman, M.K., King, I.N., Butler, C.A., Rothrock, C.R., Eapen, R.S., Hirayama-Yamada, K., Joo, K. et al. (2003). GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature* 424, 443–447.

Garg, V., Muth, A.N., Ransom, J.F., Schluterman, M.K., Barnes, R., King, I.N., Grossfeld, P.D. and Srivastava, D. (2005). Mutations in NOTCH1 cause aortic valve disease. *Nature* 437, 270–274.

Gebbia, M., Ferrero, G.B., Pilia, G., Bassi, M.T., Aylsworth, A., Penman-Splitt, M., Bird, L.M., Bamforth, J.S., Burn, J., Schlessinger, D. et al. (1997). X-linked situs abnormalities result from mutations in ZIC3. *Nat Genet* 17, 305–308.

Goldmuntz, E., Geiger, E. and Benson, D.W. (2001). NKX2.5 mutations in patients with tetralogy of fallot. *Circulation* 104, 2565–2568.

Gong, W., Gottlieb, S., Collins, J., Blescia, A., Dietz, H., Goldmuntz, E., McDonald-McGinn, D.M., Zackai, E.H., Emanuel, B.S., Driscoll, D.A. et al. (2001). Mutation analysis of TBX1 in non-deleted patients with features of DGS/VCFS or isolated cardiovascular defects. *J Med Genet* 38, E45.

Gottlieb, P.D., Pierce, S.A., Sims, R.J., Yamagishi, H., Weihe, E.K., Harriss, J.V., Maika, S.D., Kuziel, W.A., King, H.L., Olson, E.N. et al. (2002). Bop encodes a muscle-restricted protein containing MYND and SET domains and is essential for cardiac differentiation and morphogenesis. *Nat Genet* 31, 25–32.

Habashi, J.P., Judge, D.P., Holm, T.M., Cohn, R.D., Loeys, B.L., Cooper, T.K., Myers, L., Klein, E.C., Liu, G., Calvi, C. et al. (2006).

Losartan, an AT1 antagonist, prevents aortic aneurysm in a mouse model of Marfan syndrome. *Science* 312, 117–121.

Hiroi, Y., Kudoh, S., Monzen, K., Ikeda, Y., Yazaki, Y., Nagai, R. and Komuro, I. (2001). Tbx5 associates with Nkx2-5 and synergistically promotes cardiomyocyte differentiation. *Nat Genet* 28, 276–280.

Hollister, D.W., Godfrey, M., Sakai, L.Y. and Pyeritz, R.E. (1990). Immunohistologic abnormalities of the microfibrillar-fiber system in the Marfan syndrome. *N Engl J Med* 323, 152–159.

Hwang, D.M., Dempsey, A.A., Wang, R.X., Rezvani, M., Barrans, J. D., Dai, K.S., Wang, H.Y., Ma, H., Cukerman, E., Liu, Y.Q. et al. (1997). A genome-based resource for molecular cardiovascular medicine: toward a compendium of cardiovascular genes. *Circulation* 96, 4146–4203.

Hwang, D.M., Hwang, W.S. and Liew, C.C. (1994). Single pass sequencing of a unidirectional human fetal heart cDNA library to discover novel genes of the cardiovascular system. *J Mol Cell Cardiol* 26, 1329–1333.

Jamieson, C.R., van der Burgt, I., Brady, A.F., van Reen, M., Elsawi, M.M., Hol, F., Jeffery, S., Patton, M.A. and Mariman, E. (1994). Mapping a gene for Noonan syndrome to the long arm of chromosome 12. *Nat Genet* 8, 357–360.

Jerome, L.A. and Papaioannou, V.E. (2001). DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1. *Nat Genet* 27, 286–291.

Kainulainen, K., Pulkkinen, L., Savolainen, A., Kaitila, I. and Peltonen, L. (1990). Location on chromosome 15 of the gene defect causing Marfan syndrome. *N Engl J Med* 323, 935–939.

Kaynak, B., von Heydebreck, A., Mebus, S., Seelow, D., Hennig, S., Vogel, J., Sperling, H.P., Pregla, R., Alexi-Meskishvili, V., Hetzer, R. et al. (2003). Genome-wide array analysis of normal and malformed human hearts. *Circulation* 107, 2467–2474.

Kohlhase, J., Heinrich, M., Schubert, L., Liebers, M., Kispert, A., Laccone, F., Turnpenny, P., Winter, R.M. and Reardon, W. (2002). Okihiro syndrome is caused by SALL4 mutations. *Hum Mol Genet* 11, 2979–2987.

Komuro, I. and Izumo, S. (1993). Csx: a murine homeobox-containing gene specifically expressed in the developing heart. *Proc Natl Acad Sci USA* 90, 8145–8149.

Konstantinov, I.E., Coles, J.G., Boscarino, C., Takahashi, M., Goncalves, J., Ritter, J. and Van Arsdell, G.S. (2004). Gene expression profiles in children undergoing cardiac surgery for right heart obstructive lesions. *J Thorac Cardiovasc Surg* 127, 746–754.

Koshiba-Takeuchi, K., Takeuchi, J.K., Arruda, E.P., Kathiriya, I.S., Mo, R., Hui, C.C., Srivastava, D. and Bruneau, B.G. (2006). Cooperative and antagonistic interactions between Sall4 and Tbx5 pattern the mouse limb and heart. *Nat Genet* 38, 175–183.

Kuo, C.T., Morrisey, E.E., Anandappa, R., Sigrist, K., Lu, M.M., Parmacek, M.S., Soudais, C. and Leiden, J.M. (1997). GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. *Genes Dev* 11, 1048–1060.

Lee, B., Godfrey, M., Vitale, E., Hori, H., Mattei, M.G., Sarfarazi, M., Tsipouras, P., Ramirez, F. and Hollister, D.W. (1991). Linkage of Marfan syndrome and a phenotypically related disorder to two different fibrillin genes. *Nature* 352, 330–334.

Li, L., Krantz, I.D., Deng, Y., Genin, A., Banta, A.B., Collins, C.C., Qi, M., Trask, B.J., Kuo, W.L., Cochran, J. et al. (1997a). Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1. *Nat Genet* 16, 243–251.

Li, Q.Y., Newbury-Ecob, R.A., Terrett, J.A., Wilson, D.I., Curtis, A.R., Yi, C.H., Gebuhr, T., Bullen, P.J., Robson, S.C., Strachan, T. et al. (1997b). Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. *Nat Genet* 15, 21–29.

Liew, C.C., Hwang, D.M., Fung, Y.W., Laurenssen, C., Cukerman, E., Tsui, S. and Lee, C.Y. (1994). A catalogue of genes in the cardiovascular system as identified by expressed sequence tags. *Proc Natl Acad Sci USA* 91, 10645–10649.

Lindsay, E.A., Vitelli, F., Su, H., Morishima, M., Huynh, T., Pramparo, T., Jurecic, V., Ogunrinu, G., Sutherland, H.F., Scambler, P.J. et al. (2001). Tbx1 haploinsufficieny in the DiGeorge syndrome region causes aortic arch defects in mice. *Nature* 410, 97–101.

Loomes, K.M., Underkoffler, L.A., Morabito, J., Gottlieb, S., Piccoli, D.A., Spinner, N.B., Baldwin, H.S. and Oakey, R.J. (1999). The expression of Jagged1 in the developing mammalian heart correlates with cardiovascular disease in Alagille syndrome. *Hum Mol Genet* 8, 2443–2449.

Lyons, I., Parsons, L.M., Hartley, L., Li, R., Andrews, J.E., Robb, L. and Harvey, R.P. (1995). Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. *Genes Dev* 9, 1654–1666.

Magenis, R.E., Maslen, C.L., Smith, L., Allen, L. and Sakai, L.Y. (1991). Localization of the fibrillin (FBN) gene to chromosome 15, band q21.1. *Genomics* 11, 346–351.

McDaniell, R., Warthen, D.M., Sanchez-Lara, P.A., Pai, A., Krantz, I.D., Piccoli, D.A. and Spinner, N.B. (2006). NOTCH2 mutations cause Alagille syndrome, a heterogeneous disorder of the notch signaling pathway. *Am J Hum Genet* 79, 169–173.

McDonald-McGinn, D.M., Kirschner, R., Goldmuntz, E., Sullivan, K., Eicher, P., Gerdes, M., Moss, E., Solot, C., Wang, P., Jacobs, I. et al. (1999). The Philadelphia story: the 22q11.2 deletion: report on 250 patients. *Genet Couns* 10, 11–24.

McElhinney, D.B., Geiger, E., Blinder, J., Benson, D.W. and Goldmuntz, E. (2003). NKX2.5 mutations in patients with congenital heart disease. *J Am Coll Cardiol* 42, 1650–1655.

Merscher, S., Funke, B., Epstein, J.A., Heyer, J., Puech, A., Lu, M.M., Xavier, R.J., Demay, M.B., Russell, R.G., Factor, S. et al. (2001). TBX1 is responsible for cardiovascular defects in velo-cardio-facial/DiGeorge syndrome. *Cell* 104, 619–629.

Molkentin, J.D., Lin, Q., Duncan, S.A. and Olson, E.N. (1997). Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. *Genes Dev* 11, 1061–1072.

Neptune, E.R., Frischmeyer, P.A., Arking, D.E., Myers, L., Bunton, T.E., Gayraud, B., Ramirez, F., Sakai, L.Y. and Dietz, H.C. (2003). Dysregulation of TGF-beta activation contributes to pathogenesis in Marfan syndrome. *Nat Genet* 33, 407–411.

Oda, T., Elkahloun, A.G., Pike, B.L., Okajima, K., Krantz, I.D., Genin, A., Piccoli, D.A., Meltzer, P.S., Spinner, N.B., Collins, F.S. et al. (1997). Mutations in the human Jagged1 gene are responsible for Alagille syndrome. *Nat Genet* 16, 235–242.

Pashmforoush, M., Lu, J.T., Chen, H., Amand, T.S., Kondo, R., Pradervand, S., Evans, S.M., Clark, B., Feramisco, J.R., Giles, W. et al. (2004). Nkx2-5 pathways and congenital heart disease; loss of ventricular myocyte lineage specification leads to progressive cardiomyopathy and complete heart block. *Cell* 117, 373–386.

Pizzuti, A., Sarkozy, A., Newton, A.L., Conti, E., Flex, E., Digilio, M.C., Amati, F., Gianni, D., Tandoi, C., Marino, B. et al. (2003). Mutations of ZFPM2/FOG2 gene in sporadic cases of tetralogy of Fallot. *Hum Mutat* 22, 372–377.

Plageman, T.F., Jr and Yutzey, K.E. (2006). Microarray analysis of Tbx5-induced genes expressed in the developing heart. *Dev Dyn* 235, 2868–2880.

Prescott, K., Ivins, S., Hubank, M., Lindsay, E., Baldini, A. and Scambler, P. (2005). Microarray analysis of the Df1 mouse model of the 22q11 deletion syndrome. *Hum Genet* 116, 486–496.

Pu, W.T., Ishiwata, T., Juraszek, A.L., Ma, Q. and Izumo, S. (2004). GATA4 is a dosage-sensitive regulator of cardiac morphogenesis. *Dev Biol* 275, 235–244.

Reamon-Buettner, S.M. and Borlak, J. (2005). GATA4 zinc finger mutations as a molecular rationale for septation defects of the human heart. *J Med Genet* 42, e32.

Reamon-Buettner, S.M., Hecker, H., Spanel-Borowski, K., Craatz, S., Kuenzel, E. and Borlak, J. (2004). Novel NKX2-5 mutations in diseased heart tissues of patients with cardiac malformations. *Am J Pathol* 164, 2117–2125.

Reeves, R.H. (2006). Down syndrome mouse models are looking up. *Trends Mol Med* 12, 237–240.

Satoda, M., Pierpont, M.E., Diaz, G.A., Bornemeier, R.A. and Gelb, B.D. (1999). Char syndrome, an inherited disorder with patent ductus arteriosus, maps to chromosome 6p12-p21. *Circulation* 99, 3036–3042.

Satoda, M., Zhao, F., Diaz, G.A., Burn, J., Goodship, J., Davidson, H.R., Pierpont, M.E. and Gelb, B.D. (2000). Mutations in TFAP2B cause Char syndrome, a familial form of patent ductus arteriosus. *Nat Genet* 25, 42–46.

Schott, J.J., Benson, D.W., Basson, C.T., Pease, W., Silberbach, G.M., Moak, J.P., Maron, B.J., Seidman, C.E. and Seidman, J.G. (1998). Congenital heart disease caused by mutations in the transcription factor NKX2-5. *Science* 281, 108–111.

Sharma, H.S., Peters, T.H., Moorhouse, M.J., van der Spek, P.J. and Bogers, A.J. (2006). DNA microarray analysis for human congenital heart disease. *Cell Biochem Biophys* 44, 1–9.

Srivastava, D., Thomas, T., Lin, Q., Kirby, M.L., Brown, D. and Olson, E.N. (1997). Regulation of cardiac mesodermal and neural crest development by the bHLH transcription factor, dHAND. *Nat Genet* 16, 154–160.

Svenson, K.L., Bogue, M.A. and Peters, L.L. (2003). Invited review: Identifying new mouse models of cardiovascular disease: a review of high-throughput screens of mutagenized and inbred strains. *J Appl Physiol* 94, 1650–1659. discussion 1673.

Svensson, E.C., Huggins, G.S., Lin, H., Clendenin, C., Jiang, F., Tufts, R., Dardik, F.B. and Leiden, J.M. (2000). A syndrome of tricuspid atresia in mice with a targeted mutation of the gene encoding Fog-2. *Nat Genet* 25, 353–356.

Svensson, E.C., Tufts, R.L., Polk, C.E. and Leiden, J.M. (1999). Molecular cloning of FOG-2: a modulator of transcription factor GATA-4 in cardiomyocytes. *Proc Natl Acad Sci USA* 96, 956–961.

Tabibiazar, R., Wagner, R.A., Liao, A. and Quertermous, T. (2003). Transcriptional profiling of the heart reveals chamber-specific gene expression patterns. *Circ Res* 93, 1193–1201.

Tartaglia, M., Kalidas, K., Shaw, A., Song, X., Musat, D.L., van der Burgt, I., Brunner, H.G., Bertola, D.R., Crosby, A., Ion, A. et al. (2002). PTPN11 mutations in Noonan syndrome: molecular spectrum, genotype-phenotype correlation, and phenotypic heterogeneity. *Am J Hum Genet* 70, 1555–1563.

Tartaglia, M., Mehler, E.L., Goldberg, R., Zampino, G., Brunner, H.G., Kremer, H., van der Burgt, I., Crosby, A.H., Ion, A., Jeffery, S. et al. (2001). Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome. *Nat Genet* 29, 465–468.

Tevosian, S.G., Deconinck, A.E., Tanaka, M., Schinke, M., Litovsky, S.H., Izumo, S., Fujiwara, Y. and Orkin, S.H. (2000). FOG-2, a cofactor for GATA transcription factors, is essential for heart morphogenesis and development of coronary vessels from epicardium. *Cell* 101, 729–739.

Villanueva, M.P., Aiyer, A.R., Muller, S., Pletcher, M.T., Liu, X., Emanuel, B., Srivastava, D. and Reeves, R.H. (2002). Genetic and comparative mapping of genes dysregulated in mouse hearts lacking the Hand2 transcription factor gene. *Genomics* 80, 593–600.

Ware, S.M., Peng, J., Zhu, L., Fernbach, S., Colicos, S., Casey, B., Towbin, J. and Belmont, J.W. (2004). Identification and functional analysis of ZIC3 mutations in heterotaxy and related congenital heart defects. *Am J Hum Genet* 74, 93–105.

Yagi, H., Furutani, Y., Hamada, H., Sasaki, T., Asakawa, S., Minoshima, S., Ichida, F., Joo, K., Kimura, M., Imamura, S. et al. (2003). Role of TBX1 in human del22q11.2 syndrome. *Lancet* 362, 1366–1373.

Yamagishi, H., Garg, V., Matsuoka, R., Thomas, T. and Srivastava, D. (1999). A molecular pathway revealing a genetic basis for human cardiac and craniofacial defects. *Science* 283, 1158–1161.

Yu, Q., Shen, Y., Chatterjee, B., Siegfried, B.H., Leatherbury, L., Rosenthal, J., Lucas, J.F., Wessels, A., Spurney, C.F., Wu, Y.J. et al. (2004). ENU induced mutations causing congenital cardiovascular anomalies. *Development* 131, 6211–6223.

Zeisberg, E.M., Ma, Q., Juraszek, A.L., Moses, K., Schwartz, R.J., Izumo, S. and Pu, W.T. (2005). Morphogenesis of the right ventricle requires myocardial expression of Gata4. *J Clin Invest* 115, 1522–1531.

Zhao, F., Weismann, C.G., Satoda, M., Pierpont, M.E., Sweeney, E., Thompson, E.M. and Gelb, B.D. (2001). Novel TFAP2B mutations that cause Char syndrome provide a genotype-phenotype correlation. *Am J Hum Genet* 69, 695–703.

Zhao, X.S., Gallardo, T.D., Lin, L., Schageman, J.J. and Shohet, R.V. (2002). Transcriptional mapping and genomic analysis of the cardiac atria and ventricles. *Physiol Genomics* 12, 53–60.

RECOMMENDED RESOURCES

Reviews and Texts

1. Clark, K.L., Yutzey, K.E., and Benson, D.W. (2006). Transcription factors and congenital heart defects. *Annu Rev Physiol* 68, 97–121. Comprehensive review of CHD genetics with a focus on the biochemistry of the key transcription factors and related developmental mechanisms.
2. Mandel, E.M., Callis, T.E., Wang, D.-Z., and Conlon, F.L. (2005). Transcriptional mechanisms of congenital heart disease. *Drug Discov Today* 2, 33–38. Overview of cardiac transcription factors implicated in CHD with a discussion of potential therapeutic strategies based on this knowledge.
3. Harper, P.S. (2004). *Practical Genetic Counseling, 6th edition*. Oxford University Press, Oxford, 376 pp. Comprehensive overview of genetic testing and counseling applicable to CHD and other heritable diseases.

Websites

1. www.nhlbi.nih.gov/resources/pga – Program for Genomic Applications, National Heart Lung and Blood Institute, National Institutes of Health, main website with links to the several component research sites. The principal goal of this program is to develop and disseminate information, tools, and resources to link genes to biological function and diseases, including CHD, using genomic applications.
2. http://dir.nhlbi.nih.gov/labs/ldb/chd/ldbmouse.asp – Laboratory of Developmental Biology, National Heart Lung and Blood Institute, National Institutes of Health. Program of mouse chemical mutagenesis and *in utero* ultrasound screening to recover mutation causing congenital heart anomalies.
3. www.ncbi.nlm.nih.gov/omim/ – Online Mendelian Inheritance in Man database of human genes and genetic disorders, based at Johns Hopkins University and sponsored by the National Center for Biotechnology Information.
4. www.cardiogenetics.org – Congenital Cardiovascular Genetics Program, Children's Hospital Boston. Links to research and clinical resources related to CHD genetics.
5. www.childrenshospital.org/cfapps/research/data_admin/Site2234/mainpageS2234P1.html – Tetralogy of Fallot Genetics Registry, Children's Hospital Boston. Genetics resources for patients and families with tetralogy of Fallot and other CHD.

Section 5

Clinical (Oncology)

32. Genomics in the Management of Lymphomas

33. Genomics in Leukemias

34. Genomics of Lung Cancer

35. Genomics in the Diagnosis and Management of Breast Cancer

36. Colorectal Cancer

37. Genomic Evaluation and Management of Prostate Cancer

38. Genomic Assessment of Ovarian Cancer

39. Genomic Evaluation of Pancreatic Neoplasms

40. Genomic Evaluation of Head and Neck Cancer

41. Genomic Evaluation of Brain Tumors and Gliomas

42. Targeted Therapies for Cancer

This page intentionally left blank

CHAPTER

Genomics in the Management of Lymphomas

Lisa Rimsza

INTRODUCTION

"Lymphoma" is not a single disease but a general term encompassing nearly 40 different types of lymphoid malignancies broadly divided into Hodgkin and non-Hodgkin categories, the latter including precursor B or T-cell neoplasms and mature B or T-cell neoplasms, as well as rare natural killer (NK) cell neoplasms. Each type of lymphoma has unique histogenesis. In other words, each type of lymphoma has been derived from a particular lymphocyte subset defined by stage of differentiation, immune compartment, function, activation status, and/or exposure to antigens. The diversity of the immune system is reflected by the many types of lymphoma ranging from indolent to aggressive, highly curable to almost invariably lethal diseases. Any age, race, or sex of patients can be affected, as can any location in the body. Lymphoid neoplasms including lymphoma (non-Hodgkin and Hodgkin lymphoma), chronic lymphocytic leukemia and multiple myeloma accounted for 93,420 new cases and 38,000 deaths in the United States in 2005. Over one-half of new cases are non-Hodgkin lymphoma (Morton et al., 2006). The incidence of lymphoma has increased 50% from 1970 to 1990, although the rate of increase appears to be slowing. Increased lymphoma incidence is partly attributable to the AIDS epidemic and increased numbers of immunocompromised patients and partly unexplained. The cause of the increased incidence in non-immunocompromised patients is unknown but thought to be due to unidentified factors in the environment such as viral or other infections, or exposure to toxins. Table 32.1 contains a list of Hodgkin and non-Hodgkin lymphoma types according to the World Health Organization (Jaffe et al., 2001).

Predisposition

Certain groups of patients are more likely to get lymphoma than others. In general, lymphoma is a disease of adults. As mentioned above, immunocompromised patients, either due to congenital immune deficiency, HIV infection, or iatrogenic immunosuppression (for the purpose of preventing organ or tissue rejection or for treatment of autoimmune and other types of disease) are at increased risk. Presumably, in immunosuppressed states, immunosurveillance by T-cells is compromised, allowing growth of a malignant lymphoid clone. Other risk factors include genetic diseases that promote genomic instability, exposure to ionizing radiation or chemotherapy, and exposure to certain infectious agents. Epstein-Barr virus (EBV) is particularly associated with Burkitt lymphoma (BL), Hodgkin lymphoma, and lymphomas in immunocompromised patients. Human T-cell leukemia virus-1 (HTLV1) is highly associated with adult T-cell leukemia/lymphoma. Kaposi's Sarcoma Human Virus-8 (KSHV8) is associated with primary effusion lymphoma (PEL). Chronic bacterial infections such as *Helicobacter pylori* in the stomach or *Borrelia burgdorferi* in the skin can create situations of

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

TABLE 32.1 Lymphoma disease categories, characteristic chromosomal aberrations, and involved genes

Lymphoma	Characteristic chromosome aberrations	Involved genes
B-cell		
Lymphoplasmacytic/Waldenstrom's macroglobulinemia	t(9;14)(p13;q32)	*PAX5, IgH*
Splenic Marginal Zone	Loss of 7q21-32	
Extranodal marginal zone (MALT)	t(11;18)(q21;q21) t(14;18)(q32;q21) t(1;14)(p22;q21) 5(1;2)(p22;p12)	*AP12, MALT1* *MALT1, IgH* *BCL-10, IgH BCL-10, Igκ*
Nodal marginal zone		
Follicular	t(14;18)(q32;q21)	*BCL-2, IgH*
Mantle cell	t(11;14)(q13;q32)	*CCND1* (cyclin D1), *IgH*
Diffuse large B cell	t(14;18)(q32;q21) t(3;14)(q27;q32) t(3;22)(q27;q11) t(2;3)(p12;q27) 2p13–15 amplification	*BCL-2, IgH* *BCL-6, IgH* *BCL-6, Igλ BCL-6, Igκ* *REL* amplification
Mediastinal (thymic) large B cell	Gains at 9p	*JAK* amplification
Intravascular large B cell		
Primary effusion		
Burkitt	t(8;14)(q24;q32) t(8:22) (q24;q11) t(2;8) (p11;q24)	*C-MYC, IgH* *C-MYC, Igλ* *C-MYC, Igκ*
Lymphomatoid granulomatosis		
T/NK cell		
Adult T-cell leukemia/lymphoma		
Extranodal NK/T-cell, nasal type		
Enteropathy-type T-cell		
Hepatosplenic T-cell	Isochromosome 7q	
Subcutaneous panniculitis-like T-cell		
Blastic NK-cell		
Mycosis fungoides/Sezary syndrome		
Primary cutaneous CD30-positive T-cell lymphoproliferative disorders		
Angioimmunoblastic T-cell		
Peripheral T-cell, unspecified		
Anaplastic large cell	t(2;5)(p23;q35) other translocations involving ALK	*ALK, NPM* *ALK*, others
Hodgkin		
Nodular lymphocyte predominant		
Nodular sclerosis		
Mixed cellularity		
Lymphocyte rich		
Lymphocyte depleted		
Immunodeficiency associated lymphoproliferative disorders		
Primary immune disorder-related		
Human immunodeficiency virus-related		
Post-transplant		
Methotrexate-associated		

chronic inflammation which in turn lead to lymphoid proliferation and an increased risk of lymphoma (Aster, 2005).

Diagnosis

The diagnosis of lymphoma is a highly specialized branch of medicine practiced by pathologists, often with subspecialty training (hematopathologists). The diagnosis of lymphoma, and even lymphoma type, can be suspected based on clinical history, laboratory data, or imaging studies, but the definitive diagnosis can only be made from a tissue biopsy. The histologic appearance of the biopsy alone is frequently insufficient for current classification of lymphoma. Additional studies to determine the lineage and subtype of the malignant lymphoid cells as well as any characteristic genetic changes are often necessary. These studies may include immunophenotyping by flow cytometry or immunohistochemistry, cytogenetics, molecular studies such as Southern blot, PCR, RT-PCR, or FISH to assess lymphoid clonality or to discover characteristic chromosomal aberrations. Molecular diagnostic techniques are almost routine in the medical diagnosis of hematopoietic neoplasms. Table 32.1 includes a description of the most common known genetic changes in lymphomas identified by karyotyping or FISH, summarized from a recent review article (Spagnolo et al., 2004). Because of the extreme complexity of diagnosis in lymphomas and the present need to integrate information from numerous different testing modalities, a search for a single unifying technique for diagnosis is particularly appealing. Gene expression profiling (GEP) is therefore being used extensively in studies aimed at establishing a precise clinical diagnosis of particular lymphoma subtypes based on characteristic GEP patterns or "signatures."

Prognosis

In a patient with a recent diagnosis of lymphoma, prognosis is the next question to be considered. It is imperative to know what type of clinical course the patient might be facing in order to plan therapy. Therapy for lymphoma can range from careful observation without treatment, to local radiation, immunomodulation, chemotherapy, and even hematopoietic stem cell transplant. The International Prognostic Index (IPI) is commonly used in patients with diffuse large B-cell lymphoma (DLBCL) to determine the likely clinical outcome of the patient (Shipp et al., 1993). This IPI score includes age, serum lactate dehydrogenase (LDH) level, performance status, stage of disease, and number of extra-nodal sites. Patients with high IPI scores usually have a worse outcome and may be selected for more aggressive or alternative therapies. However, even within groups of patients with similar IPI scores, the outcome is variable (from cure to death). Thus, there are clearly additional factors about each patient's tumor which influence the course of the disease. GEP has been recently used to identify gene expression patterns that have additional prognostic information for an individual patient independent of the clinical IPI score (Rosenwald et al., 2002; Shipp et al., 2002).

Novel and Emerging Therapeutics

After the questions of diagnosis and prognosis, a search for therapeutic targets has been the next information sought in lymphoma GEP studies. As stated previously, even within groups of patients with the same diagnosis and with similar IPI scores, there is still variability in patient outcome. Many of the known disease subtypes have a dismal prognosis. Therefore, a search for new therapeutic targets is a pressing concern. Recent studies have identified the NF-κB pathway (Lam et al., 2005), PKCβ, MAPK (Elenitoba-Johnson et al., 2003), or immune modulation (Dave et al., 2004; Rimsza et al., 2004) as likely targets.

GEP in Lymphoma Research

Lymphoma and leukemia are particularly amenable to GEP analysis due to the relative homogeneity of tissue compared to solid tumors and to the lack of strong cohesion between the cells which makes homogenization of tissue samples relatively easy. Because of these factors, the early use of cytogenetics helped to define the genetic basis of the different diseases. The cytogenetics of these diseases in turn led to development of PCR, RT-PCR, and FISH tests to identify the characteristic genomic changes with more sensitivity and with shorter turnaround time to test completion. Consequently, acquisition of fresh tissue for flow cytometry and/or snap freezing became standard practice for hematologic malignancies in many medical centers, which stands in contrast to most of the non-hematologic solid tumors for which tissue biopsy specimens are immediately placed in fixatives to preserve tissue appearance under the microscope after paraffin embedding and staining. Therefore, when GEP was developed as a new technique, research groups interested in lymphoma had access to high quality frozen tissue. Many groups also had experience performing molecular techniques such as flow sorting, RNA/DNA extraction and qualitative or quantitative PCR and RT-PCR for diagnostic purposes. This eased the way for lymphoma researchers to apply GEP in their studies. In 1998, after completion of the Human Genome Project with a draft of the entire human DNA sequence, Dr. Robert Klausner, then director of the National Institutes of Health (NIH), announced the Director's Challenge Program to establish the molecular diagnosis of cancer. The Leukemia and Lymphoma Molecular Profiling Project (LLMPP) got underway immediately with the development of the "LymphoChip." This chip was a spotted oligonucleotide microarray containing numerous elements known to represent genes expressed by B or T lymphoid cells, genes involved in the immune response, or genes expressed by lymphoma and leukemia cell lines. These genes included cytokines, chemokines, chemokine receptors, adhesion molecules, cell surface differentiation molecules, signal transduction molecules, transcription factors, cell cycle regulators, apoptosis genes, tumor suppressor genes, oncogenes, viral genes, and genes induced/repressed during the immune response (Alizadeh et al., 1999). Under the NIH Director's Challenge Program, this group examined a series of snap frozen lymph node biopsies from different types of lymphomas and benign tissues. The results demonstrated the striking differences in GEP of the several lymphomas from each other and from benign tissues (Alizadeh et al., 2000). This was the first demonstration of the potential of GEP in the investigation of lymphoma. Other papers by the LLMPP

and other research groups, which will be detailed under particular disease subtypes in the following chapters, soon followed. Besides the Lymphochip, other groups have used different array platforms including high density oligonucleotide arrays with focused gene sets (sub-genomic) such as HY6800 or U95 chips and later on total genome chips such as U133A + B (all from the Affymetrix corporation, Santa Clara, CA). Subsequent studies have also used tumor samples (sorted, unsorted, or microdissected) or cell lines representing different types of lymphomas with a focus on diagnosis, identification of genes involved in lymphomagenesis, discovery of new subtypes, prognostically relevant genes, and therapeutic targets. The different diseases, techniques, analysis strategies, specimen sources and questions asked have led to an explosive amount of information over the last 5 years. The major advances in the most common diseases or findings illustrating a general principle will be discussed in the sections below.

DIFFUSE LARGE B-CELL LYMPHOMA

DLBCL is the most common of the aggressive lymphomas, accounting for nearly 40% of lymphomas overall. With multi-agent chemotherapy, the overall survival is approximately 50% at 5 years including some patients who are completely cured and some who die of initial or relapsed disease. The clinical IPI score for DLBCL, described above, helps to stratify patients into risk categories based on clinical features. However, even within categories, patient outcome is variable. There are three common chromosomal alterations in DLBCL. The first involves translocations of the *BCL-6* gene at the 3q27 locus with the immunoglobulin heavy chain (*IgH*) or light chain (*Igκ* or *Igλ*) genes (see Table 32.1). Since *BCL-6* is a critical gene in germinal center formation, alterations in *BCL-6* fundamentally influence the developmental program of the B-cells. Secondly, the t(14;18)(q32;q21) translocation places the anti-apoptotic gene *BCL-2* under the regulation of IgH enhancers possibly leading to increased anti-apoptosis signals. Thirdly, amplification of chromosomal region 2p13–15 leads to amplification of the *REL* gene which is a member of the NF-κB family. DLBCL can occur at any anatomic location and is characterized histologically by sheets of large B-lymphocytes which may have a variety of cytologic appearances. Several subtypes have been described based on histologic appearance (centroblastic, immunoblastic, anaplastic) or sites of presentation (immune-privileged sites, intravascular) (Jaffe et al., 2001). Because of the variable patient outcome, it has long been suspected that other biologically relevant subtypes may exist.

Diagnosis and Histogenesis

In the first GEP paper published on lymphoma, the investigators used the Lymphochip to analyze 96 samples of several subtypes of lymphoid malignancies including DLBCL, benign lymphoid cell samples, and DLBCL cell lines. They discovered that in otherwise histologically indistinguishable cases, there were at least two different subtypes of DLBCL termed the germinal center B-cell (GCB) and activated B-cell (ABC) subtypes with better and worse survival respectively. This was the first documentation that DLBCL tumors with the same appearance could have unique gene expression signatures which yielded prognostically useful information. Other researchers used Affymetrix HU6800 oligonucleotide arrays to differentiate DLBCL from follicular lymphoma (FL). DLBCL had increased expression of LDH and transferrin (known prognostic markers in lymphoma) as well as increased cellular proliferation genes, invasion and metastasis genes (cathep-sins B and D), myc-targeted genes, hematopoietic cell kinases (induced with CD44 signaling) and apoptosis inhibitors. The GCB subtype was later found to correlate with the presence of the t(14;18)(q32;q21) translocation (23–35% of the cases) which made intuitive sense because the GCB subtype of DLBCL is of follicular cell origin and may share the translocation as an early pathogenic event similar to FL in which the t(14;18) is the characteristic genetic abnormality. These findings lent credence to the importance of the GCB and ABC distinction since these subtypes could be related to a previously well known prognostically important genetic translocation (Huang et al., 2002; Iqbal et al., 2004). A short panel of only 3 immunohistochemical stains was then developed from the GEP signatures, which was capable of dividing DLBCL into GCB and non-GCB subtypes yielding similar prognostic information to the GEP analysis (Chang et al., 2004; Hans et al., 2004). This latter study demonstrated that the strength and complexity of GEP analysis is reducible into a format which can be put into immediate clinical use for patient diagnosis.

Prognosis

The LLMPP group then expanded on the initial set of 40 DLBCL cases with a follow-up study of 240 cases of untreated, untransformed DLBCL with an emphasis on gene expression patterns related to prognosis. They discovered four key gene expression signatures that were highly related to patient survival: "germinal center," "major histocompatibility (MHC) class II," "lymph node," and "proliferation." Two of the four signatures, "MHC Class II" and "lymph node," contain antigen presenting molecules, T-cell-associated genes, inflammatory genes, and stromal genes, which directly reflect the critical role of tumor immunosurveillance and host response in relation to patient outcome. Representative genes from these four signatures with the addition of the gene *BMP6* were used to create a 17-gene outcome predictor score, which provide additional prognostic value independent of the clinical IPI score. A follow-up study on the same cases used comparative genomic hybridization (CGH) to assess chromosomal losses, gains, or amplifications and found that this added prognostic information to the GEP results. In particular, chromosomal gains involving 3p11–12 correlated with a worse outcome and were related to decreased MHC Class II signature while loss at 6q21 also correlated with worse outcome but with increased proliferation signature (Bea et al., 2005). Shipp

and colleagues, however, found no prognostic significance based on the cell-of-origin (GCB versus ABC) in their data using the HU6800 chip or in their re-analysis of the Lymphochip data. However, they did find that the Lymphochip data confirmed the prognostic significance of the PKC-β2A isoform. Since these two datasets appeared to have different conclusions, the LLMPP investigators downloaded and re-analyzed the Shipp dataset. Using a different analysis technique, they were able to reconcile somewhat the two different datasets and conclusions (Wright et al., 2003). This exercise emphasizes the importance of making GEP data available to the research community for interactive re-analysis to validate scientific conclusions. It also underscores that different conclusions can be reached as a result of different technology (gene arrays, samples) and analysis strategies. The Shipp research group then used the Affymetrix total genome chip (U133A + B) on 175 DLBCL cases and identified three distinct GEP-defined subgroups termed "oxidative-phosphorylation (Ox-Phos)," "B-cell receptor/proliferation" and "host response." The "B-cell receptor/proliferation" cluster is characterized by increased expression of genes related to not only signaling and proliferation but also replication, DNA repair, and transcription factors. The "host response" cluster showed high expression of T-cell, dendritic cell, macrophage, and antigen presentation genes which may reflect the prominent inflammatory background seen in some types of DLBCL (T-cell/histiocyte rich large B-cell lymphoma). The "Ox-Phos" cluster included genes related to oxidative-phosphorylation, mitochondrial function, and the electron transport chain (Monti et al., 2005). Comparison of the consensus clusters identified in this study with the GCB and ABC cell-of-origin signatures described by Rosenwald et al. indicated that these two classification schema identify different aspects of tumor biology. Some overlap of gene expression patterns was identified in that there was enrichment of genes from the "lymph node" signature in the "host response" consensus cluster as well as increased "proliferation" signature genes in the "B-cell receptor/signaling" consensus cluster. Furthermore, the "B-cell receptor/signaling" consensus cluster genes were enriched in the ABC subtype while the "Ox-Phos" consensus cluster genes were enriched in the GCB subtype.

The importance of the loss of the "MHC Class II" gene expression signature as reported by Rosenwald et al. was further documented by a reanalysis of the LLMPP dataset. This analysis revealed that for every incremental decrease in MHC Class II expression, there was an incremental increase in the hazard ratio of death and that loss of MHC Class II expression was associated with a decreased percentage of tumor infiltrating lymphocytes. The same investigators later used GEP data to create a positional expressional profiling map which used gene expression data organized by chromosomal position to create a "virtual" CGH with implications about the genetic structure of the MHC Class II region. These "virtual" results correlated well with actual laboratory CGH results demonstrating that complete mRNA data can be used assess the possibility of DNA alterations (Rimsza et al., 2006).

In order to create a more practical platform for outcome prediction in DLBCL, Lossos and co-workers analyzed 66 cases using quantitative real-time RT-PCR to assess mRNA levels of 36 genes identified by univariate analysis in either the Shipp or LLMPP GEP data or in previous studies reported in the literature using a variety of investigational methods. These investigators identified six genes (*LM02*, *BCL-6*, *FN1*, *CCND2*, *SCYA3*, and *BCL-2*) as most important to overall survival. Using these six genes, they created a multivariate model with prognostic predictive ability independent of the clinical IPI score (Lossos et al., 2004). This work again demonstrated that GEP data can be reduced to a small number of gene analyses that might be quickly applicable in the clinical laboratory setting to determine patient prognosis.

Additional Subtypes

Other investigators have used GEP to investigate phenotypically different subtypes of DLBCL in order to determine whether these represent different diseases. One group focused on DLBCL which either does or does not express the usually T-cell associated antigen CD5 (Kobayashi et al., 2003). A second group of investigators compared site-specific DLBCL either involving or outside of the central nervous system (Fujii et al., 2005). Both of these groups discovered GEP patterns suggesting that there were additional DLBCL subtypes yet to be identified. Another group analyzed cases of DLBCL with primary involvement of the skin. There has been a long debate over whether those cases of cutaneous DLBCL which involve the lower leg were a different disease from those involving other areas of the skin. These investigators discovered that those tumors involving the leg were frequently of the ABC subtype while those tumors involving other areas of the skin were generally related to the GCB subtype (Hoefnagel et al., 2005). Their results correlated well with the known worse outcome in DLBCL of the leg compared to other sites. Why the ABC subtype of cutaneous DLBCL would involve the lower leg more frequently than other skin locations is currently a matter for speculation.

Therapeutic Targets

Using the GEP data on the ABC subtype of DLBCL, the LLMPP group identified activation of the NFκB pathway as a potential therapeutic target not typically seen in the GCB subtype of lymphoma. They then treated ABC and GCB DLBCL cell lines with IκB kinase inhibitors and demonstrated toxicity only in the ABC cell lines. This work mined the GEP data, identified an abnormal pathway, and demonstrated the *in vitro* effects of targeting that pathway, illustrating the strength of GEP for identification of new therapeutic targets.

PRIMARY MEDIASTINAL LARGE B-CELL LYMPHOMA

Primary Mediastinal Large B-Cell Lymphoma (PMBCL) has been a controversial entity that was originally thought to be

an otherwise typical DLBCL that happened to involve the mediastinum, while others described this as a unique subtype of DLBCL derived from thymic B-cells in the mediastinum. Clinical and pathologic evidence to support the latter view included a propensity of occurrence in young women, a characteristic morphologic pattern including large lymphoid cells with polylobated nuclei and abundant clear cytoplasm in a background of sclerosis, and a unique immunophenotype including lack of surface immunoglobulin and expression of FIG1 (Copie-Bergman et al., 2003). Characteristic genomic changes included chromosomal gains at 9p and amplification at 2p13–15, which includes the *REL* and *BCL-11A* loci (Jaffe et al., 2001). However the relationship between DLBCL and PMBCL remained unclear.

Diagnosis and Lymphomagenesis

The LLMPP groups used GEP with the Lymphochip to establish a molecular signature of PMBCL. Using 35 PMBCL cases defined by a >6 cm mediastinal mass at clinical presentation, they found a GEP signature which was different from either the ABC or GCB subtypes of DLBCL. Characteristic genes that were over expressed included two genes previously described as over-expressed in PMBCL, MAL, and FIG1, as well as genes under-expressed as compared to DLBCL. 21/35 cases were true PMBCL while the others appeared to be ABC or GCB subtypes of DLBCL involving the mediastinum. Surprisingly, the GEP analysis also discovered similarities to Hodgkin lymphoma (HL) cell lines and microdissected malignant HL cells from patient lymph nodes. In particular, loss of expression of genes associated with the B-cell receptor signaling pathway were noticed in both lymphomas implying a similar lymphomagenic mechanism. HL is a different lymphoma in histologic appearance and classification, but some overlapping clinicopathologic features with PMBCL are apparent. After the GEP discovery, it was reinforced that both lymphomas often had involvement of the mediastinum, predominance in young women for cases with mediastinal involvement, frequent lack of surface immunoglobulin, genomic gains at 9p and 2p, and expression of CD30 (Rosenwald et al., 2003a). Simultaneously, the Shipp research group, using the Affymetrix U133A + B platform described similar findings: that PMBCL was a separate entity from DLBCL that had overlapping features with HL. These similarities included decreased expression of molecules involved in B-cell receptor signaling and activation of the NF-κB pathway (Savage et al., 2003). Excitingly, these landmark studies reached the same conclusions using different microarray platforms, samples, and analysis methods, indicating the strength of the conclusions. The discovery that PMBCL and HL may share a similar pathogenic mechanism shed light on the origins of the controversial cases of so-called "mediastinal gray zone" lymphomas. These rare lymphomas have features of both diseases and have been difficult to diagnose as either Hodgkin or non-Hodgkin lymphoma. Perhaps the "gray zone lymphomas" represent the missing link between PMBCL and HL (Calvo et al., 2004).

Therapeutic Targets

Further investigation of the NF-κB pathway as a potential therapeutic target in PMBCL was confirmed by demonstration of REL protein localization to the nucleus, high levels of NF-κB binding activity, and NF-κB inhibition (achieved by transduction with a super-repressor form of IκBα) of proliferation in a PMBCL cell line. The investigators further defined an "NF-κB activation" signature by re-analysis of the GEP data. This signature included increased expression of genes promoting cell survival and anti-apoptotic signaling via TNFα which was significantly different from the NF-κB signature exhibited by the ABC type DLBCL. They also noted partial overlap with the "host response" consensus cluster in DLBCL which may reflect the rich stromal background characteristic of PMBCL (Feuerhake et al., 2005). NF-κB activation is frequently but not invariably associated with gene amplification, suggesting the possibility of multiple pathogenic mechanisms. Taken together, these data clearly identify NF-κB as a potential therapeutic target in PMBCL. Therapies lowering NF-κB activity, such as the proteasome inhibitor, bortezomib, which decreases degradation of the NF-κB inhibitor IκBα, are under development for clinical use in lymphomas.

HODGKIN LYMPHOMA

HL has traditionally been kept separate from the non-Hodgkin lymphomas because of very characteristic clinical and pathologic features. HL always spreads in a contiguous fashion from one lymph node group to the adjacent lymph nodes on the same side of the diaphragm (above or below), next through the lymphatic system to lymph nodes on the opposite side of the diaphragm, then to the spleen, liver, and bone marrow. Involvement outside of the hematolymphoid system is therefore rare. Because of this predictable pattern of spread, radiotherapy can be curative in the early stage disease. The most common subtype of HL is the Nodular Sclerosis variant, which occurs classically in young women and involves lymph nodes above the diaphragm. This disease has a generally good prognosis. The Mixed Cellularity variant of HL is more often seen in male and HIV+ patients, has less mediastinal involvement, a higher stage at diagnosis, and a generally less favorable outcome (Jaffe et al., 2001). Histologically, HL is also very unusual in that the malignant cells, termed Reed-Sternberg cells, and their variants are uncommon in the tumor tissue. It is estimated that the malignant cells make up fewer than 1% of all cells present with the remaining cells being inflammatory cells: lymphocytes, eosinophils, plasma cells, and stromal elements. To further complicate the diagnosis, Reed-Sternberg cells generally do not express the characteristic immunophenotype of any normal hematopoietic cells. In fact, they lack CD45 (leukocyte common antigen), specific T-cell antigens, and have weak or no expression of specific B-cell antigens. For a long time, the lineage of Reed-Sternberg cells was therefore a matter for speculation. Investigators used laser microdissection to isolate primary Reed-Sternberg cells

directly from frozen tissue sections. PCR analysis of these cells demonstrated that they were B-cells with rearrangements of the immunoglobulin heavy and light chain genes (Kanzler et al., 1996; Marafioti et al., 2000). GEP studies of HL have often relied on HL cell lines derived from Reed-Sternberg cells since the process of microdissection to obtain enough mRNA for complete analysis is difficult and time consuming. However, some investigators have first used cell lines, which allows analysis of homogenous samples, then confirmed key findings using primary tissue samples (Kuppers et al., 2003). Cell sorting has also been used to isolate primary Reed-Sternberg cells from tissues (Cossman, 2001).

Lymphomagenesis

Building on the knowledge that immunoglobulin heavy and light chain genes could be successfully rearranged in HL but the proteins were not expressed, investigators used Affymetrix U95 microarrays on HL cell lines and discovered a loss of the B-lineage specific expression program which was confirmed by immunohistochemistry. These results agreed with those described by other research groups studying PMBCL previously described in this chapter. Interestingly, continued expression of the Pax-5 gene, a transcription factor essential for B-cell commitment in early development and maintenance of the mature B-cell phenotype, was noted, although many genes known to be transcriptionally activated by Pax-5 (CD19, CD21, CD22, CD79a, BLNK) were down-regulated. No mutations in the Pax-5 gene were found to account for this phenomenon, indicating that there must be another mechanism of loss of the B-cell phenotype (Schwering et al., 2003b). In the same year, the same group compared HL cell lines to GCB DLBCL cell lines using U95 Affymetrix microarrays and confirmed some of the findings with SAGE and RT-PCR on microdissected Reed-Sternberg cells from patient lymph nodes. They found that Reed-Sternberg cells had an altered expression of hundreds of genes including potential oncogenes (rhoC, L-myc, and PTP4A) and transcription factors (ATF-5, ATBF1 and p21SNFT) and most closely resembled the ABC type of DLBCL. (Schwering et al., 2003a). Comparison of HL to NHL (non-Hodgkin lymphoma) cell lines identified expression of the activating transcription factor, ATF3, exclusively in HL. RNA interference assays selectively diminished growth of the HL cell lines compared to the NHL lines indicating the importance of ATF3 in HL growth (Janz et al., 2005). GEP using chemokine, chemokine receptor, and cytokine DNA chips on whole lymph nodes of HL identified a mixed T helper1/T helper2 response by the background T-cells and found differing chemokine expression profiles mainly related to presence or absence of EBV infection rather than histologic subtype (Ohshima et al., 2003).

Prognosis and Subtypes

Another group analyzed whole lymph nodes with GEP using their own spotted microarrays and found 3 subtypes of HL; these corresponded to the known Mixed Cellularity and Nodular Sclerosis variants with the latter divided into good and poor outcome groups. The poor outcome genes related to angiogenesis, proliferation, cell adhesion, and growth factor receptors. Good outcome genes were involved with increased apoptosis, cell signaling, cytokines, cytokine receptors and transduction molecules (Devilard et al., 2002).

FOLLICULAR LYMPHOMA

FL is the most common of the low-grade lymphomas, accounting for nearly 40% of all NHL in the United States. This disease occurs mainly in adults and is thought to originate from GCB-cells. Patients present with one or multiple sites of lymph node involvement. The tumor cells express markers of germinal center differentiation and recapitulate normal germinal center architecture. The genetic hallmark of the disease is the t(14;18)(q32;q21) that brings the anti-apoptosis gene, *BCL-2*, to the vicinity of the immunoglobulin heavy chain gene enhancers resulting in the over expression of BCL-2 protein and prolonged cell survival. The disease follows a variable waxing and waning course over several years (Jaffe et al., 2001). Therapy has been generally reserved for those patients with discomfort or other symptoms related to the bulk or location of their disease. However, new monoclonal antibody therapies directed against tumor cell surface markers may be changing the treatment paradigm (Fisher et al., 2005). About 25–35% of patients' tumors will undergo "transformation" to DLBCL which will then require multi-agent chemotherapy (Jaffe et al., 2001). Efforts to predict the clinical course of individual patients have relied on clinical features similar to the IPI score developed for DLBCL (Solal-Celigny et al., 2004). Histologic grading based on counting the number of large "centroblastic" cells per high power microscopic field has also been used, but with variable reproducibility (Jaffe et al., 2001; The Non-Hodgkin's Lymphoma Classification Project, 1997). However, speculation has remained high regarding whether there are other features of FL useful in guiding clinical treatment decision making.

Diagnosis

One of the first papers on GEP in FL compared normal GCB-cells to malignant B-cells, both enriched from involved lymph nodes by negative selection. This procedure is different from most previously discussed procedures because it compared cell types rather than complete lymph nodes. Using the Atlas cDNA Expression array (Clontech, Mountain View, CA) containing 588 cDNAs and confirming their results with quantitative RT-PCR, it was reported that FL, compared to benign B-cells, upregulated genes related to cell cycle regulation, transcription factors including Pax-5, cell–cell interactions, tumor necrosis factor, interleukin-2 receptor, and interleukin-4 receptor while decreasing expression of genes related to adhesion including MRP8 and MRP14. BCL-2 expression was increased as expected from the characteristic chromosomal translocation in FL involving BCL-2. Some of these genes mapped to regions previously reported

to be altered in FL such as 6p21.1^{CIP1} and 6p21.3^{TNF} (Husson et al., 2002). Other investigators analyzed DLBCL cell lines with t(14;18)(q32;21) compared to mantle cell lymphoma (MCL) cell lines harboring the t(11;14)(q13;q32) involving the immunoglobulin heavy chain gene and CCND1 which encodes the cyclin D1 protein. Using a cDNA spotted microarray with 4364 genes printed at their own facility, they compared the t(14;18)-containing cell lines to purified B-cells from benign tonsils and described increased expression of genes related to cellular proliferation, survival, and metabolism in the cell lines. Numerous under-expressed genes identified in the t(14;18) containing cell lines included negative regulators of cell activation and growth. Not surprisingly, t(11;14) and t(14;18) containing cell lines had distinctly different GEP profiles, with t(11;14) cell lines expressing increased cyclin D1 and associated downstream genes as expected in MCL (Robetorye et al., 2002).

Transformation

In order to investigate the molecular basis for the transformation of low grade FL into DLBCL, investigators compared 7 paired tumor samples from the same patients taken pre- and post-transformation. They found increased expression of 36 genes related to proliferation and metabolism and decreased expression of 66 genes including T-cell genes, dendritic and stromal cell genes, and transcription factors previously known to be involved with transformation, along with 25 novel genes (de Vos et al., 2003a). Other researchers evaluated 12 matched pairs proven to be clonally related to each other using an in-house spotted microarray and identified increased expression of p38-mitogen-activated protein kinase (MAPK) and growth factor cytokine receptors in the transformed samples. Results were confirmed with quantitative RT-PCR and an independent, blinded set of FL and DLBCL cases yielded similar results supporting the validity of the observations. The same paper described that pharmacologic manipulation of phosphorylated MAPK in a t(14;18)/MAPK over-expressing FL cell line inhibited growth in vitro and *in vivo* in a NOD/SCID mouse model. These results suggested MAPK as a therapeutic target in FL. A third group of investigators studied transformed FL compared to untransformed cases using array CGH to assess genomic alterations and compared these results to their previously reported GEP results on the same cases. Interestingly, they described 2 different GEP signatures in transformed cases that were related to over- or under-expression of the oncogene, *c-myc*. Heterogeneous acquired chromosomal aberrations were also observed in the transformed FL. For example, gains at 18q21.3 were observed, but not related to BCL-2 levels, implicating other genes in that region as important in transformation. Gains or amplifications at the 2p16 locus implicated the REL and BCL11A genes. This work supported the idea that there may be more than one genetic mechanism for transformation to a higher grade neoplasm (Lossos et al., 2002; Martinez-Climent et al., 2003).

Prognosis

Due to the variable clinical outcome of patients with survival ranging from 1 to 20 years, there has always been a medical interest in identifying prognostic factors in FL. The LLMPP group analyzed a training set of 95 samples, using the Affymetrix whole genome oligonucleotide arrays (U133A + B) to define a prognostic GEP signature. Two different signatures related to patient outcome, independent of clinical risk factors, were described, termed immune response 1 (IR1) and immune response 2 (IR2). The IR1 gene signature is enriched in transcripts related to T-cell mediated immune responses including monocytes and cytotoxic T-cell genes while the IR2 signature is enriched in transcripts expressed by histiocytes and dendritic cells. Both signatures can be combined to create a survival predictor score. When patients were split into 4 quartiles based on this score, median survival varied from 3.9 to 13.6 years. Most interestingly, the tumor microenvironment, particularly the non-malignant immune cells, not the tumor cells, predicted patient survival. Differences in IR1 and IR2 between individual patients may reflect genetic variation in immune response and regulation between different individuals. Or, the differences in IR1 and IR2 between different tumors may reflect variation in the molecular biology of the malignant FL cells. There are 2 hypotheses about why IR1 and IR2 are associated with good and bad prognosis, respectively. In the "immune response" hypothesis, IR1 reflects an effective antitumor response by the adaptive immune system. In contrast, IR2 reflects genes expressed in the innate immune system implying that FL cases with increased IR2 signature have learned to evade the adaptive immune system. A second hypothesis termed "immune cell dependence" suggests that the tumor-associated immune cells supply trophic survival signals to the malignant cells. The malignant cells may therefore have difficulty leaving the lymph node. FL cases with increased IR2 signature may have lost dependence on the immune cells and be able to grow without such requirements, resulting in a more aggressive tumor and poorer prognosis. Further investigation will be necessary to dissect these interactions (Staudt and Dave, 2005). An independent study compared microdissected follicles from 10 FL and 10 lymph nodes with benign reactive follicular hyperplasia using a cytokine and chemokine cDNA microarray. They identified interleukin receptor 3-alpha as increased in FL follicles among several other increased and decreased genes. Since this receptor is derived from activated follicular dendritic cells, these findings also imply the importance of the tumor microenvironment in FL growth. Other investigators used paired samples from patients with low grade (Grade 1 or 2) FL compared to aggressive disease (Grade 3b or DLBCL) based on current histologic criteria (Jaffe et al., 2001). They constructed an 81-gene model to represent these categories. In aggressive disease, genes involved in cell cycle control, DNA synthesis, and metabolism are upregulated. Genes increased in indolent disease included T-cell and macrophage-derived genes, which could either implicate host immune and microenvironmental factors in good prognosis FL or just be a reflection of higher abundance of stromal elements in low grade versus high grade lymphomas. This model accurately classified 93% of the FL samples in an independent validation set of 58 cases. The authors noted that development of diagnostic assays to make the low versus high

grade distinction using custom-made mini-chips or multiplex RT-PCR may be clinically feasible (Glas et al., 2005).

MANTLE CELL LYMPHOMA

MCL is a NHL of B-lymphocytes thought to originate from the mantle zone surrounding the germinal center. Patients are middle aged to elderly, predominantly male, and have frequent involvement of the gastrointestinal tract. These lymphomas have very poor outcomes. Patient survival is variable ranging from less than 1 to over 10 years, with a median of approximately 3 years. Most patients cannot be cured (Jaffe et al., 2001). Therefore in MCL, similar to DLBCL and FL, there is a real clinical need to identify patients with a poor prognosis who might be candidates for alternate therapy as well as patients with a good prognosis who may require less aggressive treatment. MCL has a characteristic translocation t(11;14)(q12;q32), which brings the gene encoding the cell cycle protein, cyclin D1, to the vicinity of the immunoglobulin heavy chain enhancers leading to overexpression of cyclin D1 (Jaffe et al., 2001). This deregulation of cell cycle control presumably confers a proliferative advantage to these cells. Diagnostically, the presence of the t(11;14)(q13;q32) translocation, high expression levels of cyclin D1, and characteristic immunophenotype have made MCL a fairly uniform diagnostic category. Whether or not MCL can occur without the characteristic translocation and whether the sometimes variable nuclear cytologic appearance (blastoid variant (BV) represents transformation or different disease subtypes are other diagnostic questions.

Diagnosis

The initial two GEP papers on MCL focused on the differences between MCL and benign lymphoid populations. The first paper evaluated whole MCL lymph nodes versus whole benign hyperplastic nodes. Differential expression of apoptotic genes were identified and confirmed with quantitative RT-PCR in addition to the expected over-expression of cyclin D1 and associated genes (Hofmann et al., 2001). The second paper compared MCL to pre-germinal center (antigen-naïve) and post-germinal center (antigen-experienced) B-cells sorted from normal tonsils. They described normal levels of expression of the CCR7 receptor but decreased expression of other receptors that are expressed during transition of B-cells from primary follicles to GCB-cells. They challenged the previous belief that MCL originates from antigenically naïve, pre-GCB-cells suggesting a slightly later functional stage. MCL was characterized by decreased trafficking and differentiation genes, deregulated growth factors and growth factor receptors and oncogenes (*Cyclin D1*, *BCL-2*, and *MERTK*), increased metastasis and angiogenesis genes, defective apoptosis, and differential expression of neurotransmitter receptors. Two GEP patterns were identified indicating that there might be two different subtypes of MCL (Ek et al., 2002). Two other groups compared MCL to flow cytometry-sorted normal mantle zone lymphocytes. Altered apoptosis and signaling pathways were identified compared to normal populations. Gene expression signatures for pathologically defined subtypes included signatures for cases with somatic mutations of the immunoglobulin heavy chain gene, a proliferation signature, and a blastoid cytology signature (Martinez et al., 2003; Rizzatti et al., 2005). De Vos and colleagues investigated the BV of MCL and identified altered progress through the G1/S and G2/M checkpoints along with cyclin D1-associated genes indicating that the BV cases may represent MCL with additional alterations in the gene expression pattern related to cell cycle (de Vos et al., 2003b). A comparison of MCL with 2 other types of lymphoma, small lymphocytic lymphoma and marginal zone lymphoma, which sometimes have overlapping clinicopathologic variables, demonstrated that MCL had an upregulation of cell cycle control genes, as anticipated from the known association with t(11;14)(q13;q32), as well as multidrug resistance genes. This latter finding may have implications for treatment options (Thieblemont et al., 2004).

Prognosis

The LLMPP group analyzed 101 cases of MCL to identify a diagnostic and prognostic signature unique to this type of lymphoma. Genes associated with outcome were analyzed to develop a survival predictor model. A large proportion of the survival predictor genes (24/48, 58%) formed a signature related to proliferation. However, these proliferation genes were different than the proliferation signature genes that constitute an important survival predictor in DLBCL. In MCL, the proliferation genes were associated with cell cycle progression and DNA synthesis, but not with c-myc as seen in the DLBCL proliferation signature. Furthermore, the higher the expression of cyclin D1 transcripts, the worse the patient outcome. Independent of the elevated cyclin D1, deletion of the *INK4a/ARF* locus encoding the $p16^{INK4a}$ and $p14^{ARF}$ tumor suppressor genes related to cell cycle progression was more frequent in MCL with high expression of the proliferation signature. Thus, a quantitative model based on the proliferation signature was able to predict patient outcome (Rosenwald et al., 2003b). Interestingly, approximately 5% of cases submitted to the study but not included in the above analysis, had otherwise typical clinical pathologic features but did not overexpress the cyclin D1 mRNA or protein. However, they had other diagnostic, morphologic, and immunophenotype features of MCL as well as a characteristic MCL GEP. These cases expressed high levels of cyclin D2 or D3, perhaps as an alternative molecular mechanism of pathogenesis (Fu et al., 2005).

BURKITT LYMPHOMA

BL is a type of B-cell NHL which is thought to originate from germinal center centroblasts. Classically, the malignant cells are medium-sized, round, and uniform with frequent mitotic figures and frequent apoptosis. The apoptotic cells and necrotic debris are often engulfed by macrophages resulting in the characteristic "starry sky" effect. By immunophenotyping, the cells strongly

express B-cell antigens, show light chain restriction, and have markers of germinal center differentiation including CD10 and BCL-6 as well as high proliferation (Ki-67 antigen expression approaching 100%). BCL-2, found in many DLBCL, is typically negative. Genetically, BL is defined by a translocation between the *c-myc* oncogene and either the immunoglobulin heavy chain gene or one of the light chain genes. This translocation results in over expression of c-myc and dysregulated cell cycle. BL is a relatively common lymphoma in children, but much rarer in adults. BL is associated with EBV infection, particularly in the endemic form, which is found in children in Africa. Sporadic BL is less often associated with EBV and more common in adults. BL also occurs in immunocompromised patients (Jaffe et al., 2001). Because of the extremely high proliferation rate, BL is exquisitely sensitive to chemotherapy using cell-cycle specific drugs. Intensive chemotherapy regimens can be curative, but are also difficult to tolerate, particularly for adults. Thus, the clinical distinction between BL and DLBCL, which is more common and receives a less intensive chemotherapy regimen, is critical. Currently, diagnosis is based on a combination of morphologic and immunophenotypic features along with evidence of the characteristic c-myc translocation. Atypical Burkitt cases or "Burkitt-like" cases have also been recognized that have most features of BL but with cytologic features outside what is typically encountered (Jaffe et al., 2001). Cases with otherwise typical DLBCL morphology can also harbor the c-myc translocation. Cases with overlapping features are not infrequent, particularly in adult patients, a situation that can lead to a difficult diagnostic dilemma with significant clinical consequences.

Diagnosis

Because of the real-world difficulty in accurately diagnosing BL, two research groups independently pursued definition of a characteristic GEP pattern for classic BL as compared to DLBCL. Their results were published simultaneously in the same journal (Dave et al., 2006; Hummel et al., 2006). The first paper was published by the LLMPP group using Affymetrix's U133A + B chips. Their analysis of 303 cases of aggressive B-cell lymphomas clearly identified a characteristic gene expression signature associated with classic BL cases as diagnosed by a panel of expert hematopathologists with morphology, immunophenotyping, and FISH for the c-myc translocation. In addition, they found discrepant cases with either the clinicopathologic diagnosis of BL and GEP of DLBCL or the reverse, a clinicopathologic diagnosis of DLBCL or high grade DLBCL, but with the GEP signature of BL. Of interest, they found that cases with the BL GEP signature most likely benefited from more intensive chemotherapy regimens, regardless of the clinicopathologic classification. The genes that were highly expressed in BL included a subset of GCB-cell associated genes and targets of c-myc, while there was diminished expression of NF-kappa-B targets and MHC class I genes. The second study used Affymetrix U133A chips to study 220 aggressive B-cell lymphomas to develop a GEP signature for the disease. FISH and CGH studies were also performed. They defined a GEP signature for BL which was predominantly associated with Ig-c-myc rearrangement and low chromosomal complexity score. Other cases without the characteristic signature (mainly DLBCL) were usually without c-myc rearrangment. Intermediate cases with a mixture of features were also identified. Both of these papers examined large numbers of cases and arrived at very similar conclusions: BL is a molecularly definable group of lymphoma that can be characterized by GEP and translocation status, that the classic pathologic diagnostic approach can be enhanced with the addition of GEP information, and that intermediate cases with features of both DLBCL and BL do occur. This latter group of cases will require further study to define exact categories and implications for therapy. It was anticipated by both groups that the GEP definition of BL in some form will be taken forward and applied to clinical research and practice.

2009 UPDATE

Additional gene expression studies on peripheral T-cell lymphomas (PTCL) have been reported recently, aimed at defining the underlying mechanisms of disease (de Leval et al., 2007; Lamant et al., 2007; Piccaluga et al., 2007a, b). While the number of cases in these studies is still quite limited, there are some interesting findings, including the possible derivation of angioimmunoblastic T-cell lymphoma from T-follicular helper cells (de Leval et al., 2007; Piccaluga et al., 2007a).

There has been speculation that the originally described prognosticator for DLBCL may not hold with the addition of rituximab to the treatment regimen. A recent study provided evidence that the cell of origin classification is still predictive of survival and, in addition, the stromal signature is also highly significant (Lenz et al., 2008b). Other global investigations such as microRNA profiling (Lu et al., 2005; Zhang et al., 2006), CGH including high resolution array CGH (Bea et al., 2005; Kimm et al., 2007; Lenz et al., 2008; Salaverria et al., 2007), methylation studies (Pike et al., 2008), and mutation analysis (McLendon, 2008; Shendure and Ji, 2008; Wood et al., 2007) are being pursued, and all this information can be integrated to give a better understanding of the pathogenesis and evolution of hematologic malignancies. Interesting examples include the activation of the NF-kB pathway by recently discovered mutations affecting *CARD11* and *TNFAIP3* (Lenz et al., 2008a; Pasqualucci et al., 2008), which partially account for the frequent upregulation of the NF-kB pathway in the ABC-type of DLBCL.

It is uncertain how this information will be utilized in the clinical setting and with which platform. A diagnostic

array may be constructed for hematologic malignancies. Alternatively, useful information from GEP can be condensed and represented by a much smaller number of transcripts and adapted for quantitative RT-PCR (Malumbres et al., 2008) or immunohistochemistry (Hans et al., 2004) that can be applied to the study of paraffin embedded materials. It is also not clear if a single diagnostic platform can be designed that will incorporate important findings from multiple sources (e.g., GEP, aCGH, methylation, mutation) or whether several assays will have to be used to capture all relevant information. However, it is likely that some form of diagnostic assay for hematologic malignancies based on these global analyses will emerge in the next few years to provide additional relevant molecular information for clinical use.

2009 UPDATE REFERENCES

Bea, S., Zettl, A. et al. (2005). Diffuse large B-cell lymphoma subgroups have distinct genetic profiles that influence tumor biology and improve gene-expression-based survival prediction. *Blood* 106(9), 3183–3190.

de Leval, L., Rickman, D.S. et al. (2007). The gene expression profile of nodal peripheral T-cell lymphoma demonstrates a molecular link between angioimmunoblastic T-cell lymphoma (AITL) and follicular helper T (TFH) cells. *Blood* 109(11), 4952–4963.

Hans, C.P., Weisenburger, D.D. et al. (2004). Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood* 103(1), 275–282.

Kimm, L.R., deLeeuw, R.J. et al. (2007). Frequent occurrence of deletions in primary mediastinal B-cell lymphoma. *Genes Chromosomes Cancer* 46(12), 1090–1097.

Lamant, L., de Reynies, A. et al. (2007). Gene-expression profiling of systemic anaplastic large-cell lymphoma reveals differences based on ALK status and two distinct morphologic ALK+ subtypes. *Blood* 109(5), 2156–2164.

Lenz, G., Davis, R.E. et al. (2008a). Oncogenic CARD11 mutations in human diffuse large B cell lymphoma. *Science* 319(5870), 1676–1679.

Lenz, G., Wright, G. et al. (2008b). Stromal gene signatures in large-B-cell lymphomas. *N Engl J Med* 359(22), 2313–2323.

Lu, J., Getz, G. et al. (2005). MicroRNA expression profiles classify human cancers. *Nature* 435(7043), 834–838.

Malumbres, R., Chen, J. et al. (2008). Paraffin-based 6-gene model predicts outcome in diffuse large B-cell lymphoma patients treated with R-CHOP. *Blood* 111(12), 5509–5514.

McLendon, R., Friedman, A., Bigner, D., Van Meir, E.G., Brat, D.J., Mastrogianakis, G.M., Olson, J.J., Mikkelsen, T., Lehman, N., Aldape, K. et al. (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455(7216), 1061–1068.

Pasqualucci, L., Compagno, M., Lim, W.K., Grunn, A., Nandula, S.V., Scandurra, M., et al. (2008). Mutations in multiple genes cause deregulation of the NFkB pathway in diffuse large B-C lymphoma. *ASH Abstract* 801.

Piccaluga, P.P., Agostinelli, C. et al. (2007a). Gene expression analysis of angioimmunoblastic lymphoma indicates derivation from T follicular helper cells and vascular endothelial growth factor deregulation. *Cancer Res* 67(22), 10703–10710.

Piccaluga, P.P., Agostinelli, C. et al. (2007b). Gene expression analysis of peripheral T cell lymphoma, unspecified, reveals distinct profiles and new potential therapeutic targets. *J Clin Invest* 117(3), 823–834.

Pike, B.L., Greiner, T.C. et al. (2008). DNA methylation profiles in diffuse large B-cell lymphoma and their relationship to gene expression status. *Leukemia* 22(5), 1035–1043.

Salaverria, I., Zettl, A. et al. (2007). Specific secondary genetic alterations in mantle cell lymphoma provide prognostic information independent of the gene expression-based proliferation signature. *J Clin Oncol* 25(10), 1216–1222.

Shendure, J. and Ji, H. (2008). Next-generation DNA sequencing. *Nat Biotechnol* 26(10), 1135–1145.

Wood, L.D., Parsons, D.W. et al. (2007). The genomic landscapes of human breast and colorectal cancers. *Science* 318(5853), 1108–1113.

Zhang, L., Huang, J. et al. (2006). microRNAs exhibit high frequency genomic alterations in human cancer. *Proc Natl Acad Sci U S A* 103(24), 9136–9141.

REFERENCES

Alizadeh, A., Eisen, M., Davis, R.E., Ma, C., Sabet, H., Tran, T., Powell, J.I., Yang, L., Marti, G.E., Moore, D.T. et al. (1999). The lymphochip: A specialized cDNA microarray for the genomic-scale analysis of gene expression in normal and malignant lymphocytes. *Cold Spring Harb Symp Quant Biol* 64, 71–78.

Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X. et al. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403, 503–511.

Aster, J.C. (2005). Diseases of white blood cells, lymph nodes, spleen, and thymus. In *Robbins and Contran Pathologic Basis of Disease* (V. Kumar, A.K. Abbas and N. Fausto, eds), Elsevier Saunders, Philadelphia, pp. 661–709.

Bea, S., Zettl, A., Wright, G., Salaverria, I., Jehn, P., Moreno, V., Burek, C., Ott, G., Puig, X., Yang, L. et al. (2005). Diffuse large B-cell lymphoma subgroups have distinct genetic profiles that influence tumor biology and improve gene-expression-based survival prediction. *Blood* 106, 3183–3190.

Calvo, K.R., Traverse-Glehen, A., Pittaluga, S. and Jaffe, E.S. (2004). Molecular profiling provides evidence of primary mediastinal large B-cell lymphoma as a distinct entity related to classic Hodgkin lymphoma: Implications for mediastinal gray zone lymphomas as

an intermediate form of B-cell lymphoma. *Adv Anat Pathol* 11, 227–238.

Chang, C.C., McClintock, S., Cleveland, R.P., Trzpuc, T., Vesole, D.H., Logan, B., Kajdacsy-Balla, A. and Perkins, S.L. (2004). Immunohistochemical expression patterns of germinal center and activation B-cell markers correlate with prognosis in diffuse large B-cell lymphoma. *Am J Surg Pathol* 28, 464–470.

Copie-Bergman, C., Boulland, M.L., Dehoulle, C., Moller, P., Farcet, J.P., Dyer, M.J., Haioun, C., Romeo, P.H., Gaulard, P. and Leroy, K. (2003). Interleukin 4-induced gene 1 is activated in primary mediastinal large B-cell lymphoma. *Blood* 101, 2756–2761.

Cossman, J. (2001). Gene expression analysis of single neoplastic cells and the pathogenesis of Hodgkin's lymphoma. *J Histochem Cytochem* 49, 799–800.

Dave, S.S., Wright, G., Tan, B., Rosenwald, A., Gascoyne, R.D., Chan, W.C., Fisher, R.I., Braziel, R.M., Rimsza, L.M., Grogan, T.M. et al. (2004). Prediction of survival in follicular lymphoma based on molecular features of tumor-infiltrating immune cells. *N Engl J Med* 351, 2159–2169.

Dave, S.S., Fu, K., Wright, G.W., Lam, L.T., Kluin, P., Boerma, E.J., Greiner, T. C., Weisenburger, D.D., Rosenwald, A., Ott, G., et al. (2006). Molecular diagnosis of Burkitt's lymphoma. *N Engl J Med* 354, 2431–2442.

de Vos, S., Hofmann, W.K., Grogan, T.M., Krug, U., Schrage, M., Miller, T.P., Braun, J.G., Wachsman, W., Koeffler, H.P. and Said, J.W. (2003a). Gene expression profile of serial samples of transformed B-cell lymphomas. *Lab Invest* 83, 271–285.

de Vos, S., Krug, U., Hofmann, W.K., Pinkus, G.S., Swerdlow, S.H., Wachsman, W., Grogan, T.M., Said, J.W. and Koeffler, H.P. (2003b). Cell cycle alterations in the blastoid variant of mantle cell lymphoma (MCL-BV) as detected by gene expression profiling of mantle cell lymphoma (MCL) and MCL-BV. *Diagn Mol Pathol* 12, 35–43.

Devilard, E., Bertucci, F., Trempat, P., Bouabdallah, R., Loriod, B., Giaconia, A., Brousset, P., Granjeaud, S., Nguyen, C., Birnbaum, D. et al. (2002). Gene expression profiling defines molecular subtypes of classical Hodgkin's disease. *Oncogene* 21, 3095–3102.

Ek, S., Hogerkorp, C.M., Dictor, M., Ehinger, M. and Borrebaeck, C.A. (2002). Mantle cell lymphomas express a distinct genetic signature affecting lymphocyte trafficking and growth regulation as compared with subpopulations of normal human B cells. *Cancer Res* 62, 4398–4405.

Elenitoba-Johnson, K.S., Jenson, S.D., Abbott, R.T., Palais, R. A., Bohling, S.D., Lin, Z., Tripp, S., Shami, P.J., Wang, L.Y., Coupland, R.W. et al. (2003). Involvement of multiple signaling pathways in follicular lymphoma transformation: p38-mitogen-activated protein kinase as a target for therapy. *Proc Natl Acad Sci USA* 100, 7259–7264.

Feuerhake, F., Kutok, J.L., Monti, S., Chen, W., LaCasce, A.S., Cattoretti, G., Kurtin, P., Pinkus, G.S., De Leval, L., Harris, N.L. et al. (2005). NFkappaB activity, function, and target-gene signatures in primary mediastinal large B-cell lymphoma and diffuse large B-cell lymphoma subtypes. *Blood* 106, 1392–1399.

Fisher, R.I., LeBlanc, M., Press, O.W., Maloney, D.G., Unger, J.M. and Miller, T.P. (2005). New treatment options have changed the survival of patients with follicular lymphoma. *J Clin Oncol* 23, 8447–8452.

Fu, K., Weisenburger, D.D., Greiner, T.C., Dave, S., Wright, G., Rosenwald, A., Chiorazzi, M., Iqbal, J., Gesk, S., Siebert, R. et al. (2005). Cyclin D1-negative mantle cell lymphoma: A clinicopathologic study based on gene expression profiling. *Blood* 106, 4315–4321.

Fujii, A., Oshima, K., Hamasaki, M., Utsunomiya, H., Okazaki, M., Kagami, Y., Seto, M. and Kikuchi, M. (2005). Differential expression of cytokines, chemokines and their receptors in follicular lymphoma and reactive follicular hyperplasia: Assessment by complementary DNA microarray. *Oncol Rep* 13, 819–824.

Glas, A.M., Kersten, M.J., Delahaye, L.J., Witteveen, A.T., Kibbelaar, R.E., Velds, A., Wessels, L.F., Joosten, P., Kerkhoven, R.M., Bernards, R. et al. (2005). Gene expression profiling in follicular lymphoma to assess clinical aggressiveness and to guide the choice of treatment. *Blood* 105, 301–307.

Hans, C.P., Weisenburger, D.D., Greiner, T.C., Gascoyne, R.D., Delabie, J., Ott, G., Muller-Hermelink, H.K., Campo, E., Braziel, R.M., Jaffe, E.S. et al. (2004). Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood* 103, 275–282.

Hoefnagel, J.J., Dijkman, R., Basso, K., Jansen, P.M., Hallermann, C., Willemze, R., Tensen, C.P. and Vermeer, M.H. (2005). Distinct types of primary cutaneous large B-cell lymphoma identified by gene expression profiling. *Blood* 105, 3671–3678.

Hofmann, W.K., de Vos, S., Tsukasaki, K., Wachsman, W., Pinkus, G.S., Said, J.W. and Koeffler, H.P. (2001). Altered apoptosis pathways in mantle cell lymphoma detected by oligonucleotide microarray. *Blood* 98, 787–794.

Huang, J.Z., Sanger, W.G., Greiner, T.C., Staudt, L.M., Weisenburger, D.D., Pickering, D.L., Lynch, J.C., Armitage, J.O., Warnke, R.A., Alizadeh, A.A. et al. (2002). The t(14;18) defines a unique subset of diffuse large B-cell lymphoma with a germinal center B-cell gene expression profile. *Blood* 99, 2285–2290.

Hummel, M., Bentink, S., Berger, H., Klapper, W., Wessendorf, S., Barth, T.F.E., Bernd, H.W., Cogliatti, S.B., Dierlamm, J., Feller, A.C., et al. (2006). A biologic definition of Burkitt's lymphoma from transcriptional and genomic profiling. *N Engl J Med* 354, 2419–2430.

Husson, H., Carideo, E.G., Neuberg, D., Schultze, J., Munoz, O., Marks, P.W., Donovan, J.W., Chillemi, A.C., O'Connell, P. and Freedman, A.S. (2002). Gene expression profiling of follicular lymphoma and normal germinal center B cells using cDNA arrays. *Blood* 99, 282–289.

Iqbal, J., Sanger, W.G., Horsman, D.E., Rosenwald, A., Pickering, D.L., Dave, B., Dave, S., Xiao, L., Cao, K., Zhu, Q. et al. (2004). BCL2 translocation defines a unique tumor subset within the germinal center B-cell-like diffuse large B-cell lymphoma. *Am J Pathol* 165, 159–166.

Jaffe, E.S., Harris, N.L., Stein, H. and Vardiman, J.W. (2001). *WHO Classification of Tumours, Pathology and Genetics, Tumours of Haematopoietic and Lymphoid Tissues*. IARC Press, Lyon, France.

Janz, M., Hummel, M., Truss, M., Wollert-Wulf, B., Mathas, S., Johrens, K., Hagemeier, C., Bommert, K., Stein, H., Dorken, B. et al. (2006). Classical Hodgkin lymphoma is characterized by high constitutive expression of activating transcription factor 3 (ATF3) which promotes viability of Hodgkin/Reed-Sternberg cells. *Blood* 107, 2536–2539.

Kanzler, H., Kuppers, R., Hansmann, M.L. and Rajewsky, K. (1996). Hodgkin and Reed-Sternberg cells in Hodgkin's disease represent the outgrowth of a dominant tumor clone derived from (crippled) germinal center B cells. *J Exp Med* 184, 1495–1505.

Kobayashi, T., Yamaguchi, M., Kim, S., Morikawa, J., Ogawa, S., Ueno, S., Suh, E., Dougherty, E., Shmulevich, I., Shiku, H. et al. (2003). Microarray reveals differences in both tumors and vascular specific gene expression in *de novo* CD5+ and CD5− diffuse large B-cell lymphomas. *Cancer Res* 63, 60–66.

Kuppers, R., Klein, U., Schwering, I., Distler, V., Brauninger, A., Cattoretti, G., Tu, Y., Stolovitzky, G.A., Califano, A., Hansmann, M.L. et al. (2003). Identification of Hodgkin and Reed-Sternberg cell-specific genes by gene expression profiling. *J Clin Invest* 111, 529–537.

Lam, L.T., Davis, R.E., Pierce, J., Hepperle, M., Xu, Y., Hottelet, M., Nong, Y., Wen, D., Adams, J., Dang, L. et al. (2005). Small molecule inhibitors of IkappaB kinase are selectively toxic for subgroups of diffuse large B-cell lymphoma defined by gene expression profiling. *Clin Cancer Res* 11, 28–40.

Lossos, I.S., Alizadeh, A.A., Diehn, M., Warnke, R., Thorstenson, Y., Oefner, P.J., Brown, P.O., Botstein, D. and Levy, R. (2002). Transformation of follicular lymphoma to diffuse large-cell lymphoma: Alternative patterns with increased or decreased expression of c-myc and its regulated genes. *Proc Natl Acad Sci USA* 99, 8886–8891.

Lossos, I.S., Czerwinski, D.K., Alizadeh, A.A., Wechser, M.A., Tibshirani, R., Botstein, D. and Levy, R. (2004). Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. *N Engl J Med* 350, 1828–1837.

Marafioti, T., Hummel, M., Foss, H.D., Laumen, H., Korbjuhn, P., Anagnostopoulos, I., Lammert, H., Demel, G., Theil, J., Wirth, T. et al. (2000). Hodgkin and Reed-Sternberg cells represent an expansion of a single clone originating from a germinal center B-cell with functional immunoglobulin gene rearrangements but defective immunoglobulin transcription. *Blood* 95, 1443–1450.

Martinez, N., Camacho, F.I., Algara, P., Rodriguez, A., Dopazo, A., Ruiz-Ballesteros, E., Martin, P., Martinez-Climent, J.A., Garcia-Conde, J., Menarguez, J. et al. (2003). The molecular signature of mantle cell lymphoma reveals multiple signals favoring cell survival. *Cancer Res* 63, 8226–8232.

Martinez-Climent, J.A., Alizadeh, A.A., Segraves, R., Blesa, D., Rubio-Moscardo, F., Albertson, D.G., Garcia-Conde, J., Dyer, M.J., Levy, R., Pinkel, D. et al. (2003). Transformation of follicular lymphoma to diffuse large cell lymphoma is associated with a heterogeneous set of DNA copy number and gene expression alterations. *Blood* 101, 3109–3117.

Monti, S., Savage, K.J., Kutok, J.L., Feuerhake, F., Kurtin, P., Mihm, M., Wu, B., Pasqualucci, L., Neuberg, D., Aguiar, R.C. et al. (2005). Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. *Blood* 105, 1851–1861.

Morton, L.M., Wang, S.S., Devesa, S.S., Hartge, P., Weisenburger, D.D. and Linet, M.S. (2006). Lymphoma incidence patterns by WHO subtype in the United States, 1992–2001. *Blood* 107, 265–276.

Ohshima, K., Karube, K., Hamasaki, M., Tutiya, T., Yamaguchi, T., Suefuji, H., Suzuki, K., Suzumiya, J., Ohga, S. and Kikuchi, M. (2003). Differential chemokine, chemokine receptor and cytokine expression in Epstein-Barr virus-associated lymphoproliferative diseases. *Leuk Lymphoma* 44, 1367–1378.

Rimsza, L.M., Roberts, R.A., Miller, T.P., Unger, J.M., LeBlanc, M., Braziel, R.M., Weisenberger, D.D., Chan, W.C., Muller-Hermelink, H.K., Jaffe, E.S. et al. (2004). Loss of MHC class II gene and protein expression in diffuse large B-cell lymphoma is related to decreased tumor immunosurveillance and poor patient survival regardless of other prognostic factors: A follow-up study from the Leukemia and Lymphoma Molecular Profiling Project. *Blood* 103, 4251–4258.

Rimsza, L.M., Roberts, R.A., Campo, E., Grogan, T.M., Bea, S., Salaverria, I., Zettl, A., Rosenwald, A., Ott, G., Muller-Hermelink, K. et al. (2006). Loss of major histocompatibility class II expression in non-immune privileged site diffuse large B cell lymphoma is highly coordinated and not due to chromosomal deletions. *Blood* 107, 1101–1107.

Rizzatti, E.G., Falcao, R.P., Panepucci, R.A., Proto-Siqueira, R., Anselmo-Lima, W.T., Okamoto, O.K. and Zago, M.A. (2005). Gene expression profiling of mantle cell lymphoma cells reveals aberrant expression of genes from the PI3K-AKT, WNT and TGFbeta signalling pathways. *Br J Haematol* 130, 516–526.

Robetorye, R.S., Bohling, S.D., Morgan, J.W., Fillmore, G.C., Lim, M.S. and Elenitoba-Johnson, K.S. (2002). Microarray analysis of B-cell lymphoma cell lines with the t(14;18). *J Mol Diagn* 4, 123–136.

Rosenwald, A., Wright, G., Chan, W.C., Connors, J.M., Campo, E., Fisher, R.I., Gascoyne, R.D., Muller-Hermelink, H.K., Smeland, E.B., Giltnane, J.M. et al. (2002). The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* 346, 1937–1947.

Rosenwald, A., Wright, G., Leroy, K., Yu, X., Gaulard, P., Gascoyne, R.D., Chan, W.C., Zhao, T., Haioun, C., Greiner, T.C. et al. (2003a). Molecular diagnosis of primary mediastinal B cell lymphoma identifies a clinically favorable subgroup of diffuse large B cell lymphoma related to Hodgkin lymphoma. *J Exp Med* 198, 851–862.

Rosenwald, A., Wright, G., Wiestner, A., Chan, W.C., Connors, J.M., Campo, E., Gascoyne, R.D., Grogan, T.M., Muller-Hermelink, H.K., Smeland, E.B. et al. (2003b). The proliferation gene expression signature is a quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma. *Cancer Cell* 3, 185–197.

Savage, K.J., Monti, S., Kutok, J.L., Cattoretti, G., Neuberg, D., De Leval, L., Kurtin, P., Dal Cin, P., Ladd, C., Feuerhake, F. et al. (2003). The molecular signature of mediastinal large B-cell lymphoma differs from that of other diffuse large B-cell lymphomas and shares features with classical Hodgkin lymphoma. *Blood* 102, 3871–3879.

Schwering, I., Brauninger, A., Distler, V., Jesdinsky, J., Diehl, V., Hansmann, M.L., Rajewsky, K. and Kuppers, R. (2003a). Profiling of Hodgkin's lymphoma cell line L1236 and germinal center B cells: Identification of Hodgkin's lymphoma-specific genes. *Mol Med* 9, 85–95.

Schwering, I., Brauninger, A., Klein, U., Jungnickel, B., Tinguely, M., Diehl, V., Hansmann, M.L., Dalla-Favera, R., Rajewsky, K. and Kuppers, R. (2003b). Loss of the B-lineage-specific gene expression program in Hodgkin and Reed-Sternberg cells of Hodgkin lymphoma. *Blood* 101, 1505–1512.

Shipp, M.A., Harrington, D.P., Anderson, J.R., Armitage, J.O., Bonadonna, G., Brittinger, G., Cabanillas, F., Canellos, G.P., Coiffier, B., Connors, J.M. et al. (1993). A predictive model for aggressive non-Hodgkins-lymphoma. *N Engl J Med* 329, 987–994.

Shipp, M.A., Ross, K.N., Tamayo, P., Weng, A.P., Kutok, J.L., Aguiar, R.C., Gaasenbeek, M., Angelo, M., Reich, M., Pinkus, G.S. et al. (2002). Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat Med* 8, 68–74.

Solal-Celigny, P., Roy, P., Colombat, P., White, J., Armitage, J.O., Arranz-Saez, R., Au, W.Y., Bellei, M., Brice, P., Caballero, D. et al. (2004). Follicular lymphoma international prognostic index. *Blood* 104, 1258–1265.

Spagnolo, D.V., Ellis, D.W., Juneja, S., Leong, A.S., Miliauskas, J., Norris, D.L. and Turner, J. (2004). The role of molecular studies in lymphoma diagnosis: A review. *Pathology* 36, 19–44.

Staudt, L.M. and Dave, S. (2005). The biology of human lymphoid malignancies revealed by gene expression profiling. *Adv Immunol* 87, 163–208.

The Non-Hodgkin's Lymphoma Classification Project (1997). A clinical evaluation of the International Lymphoma Study Group classification of non-Hodgkin's lymphoma. *Blood* 89, 3909–3918.

Thieblemont, C., Nasser, V., Felman, P., Leroy, K., Gazzo, S., Callet-Bauchu, E., Loriod, B., Granjeaud, S., Gaulard, P., Haioun, C. et al. (2004). Small lymphocytic lymphoma, marginal zone B-cell lymphoma, and mantle cell lymphoma exhibit distinct gene-expression profiles allowing molecular diagnosis. *Blood* 103, 2727–2737.

Wright, G., Tan, B., Rosenwald, A., Hurt, E.H., Wiestner, A. and Staudt, L.M. (2003). A gene expression based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proc Natl Acad Sci USA* 100, 9991–9996.

CHAPTER

Genomics in Leukemias

Lars Bullinger, Hartmut Dohner and Jonathan R. Pollack

INTRODUCTION

Recently, our understanding of the hematopoietic system has expanded dramatically, revealing leukemias to exhibit an extraordinary biologic and clinical heterogeneity. While the broad classification of leukemias is still based on the cell of origin (e.g. myeloid or lymphoid) and the rapidity of the clinical course (e.g., acute or chronic), over the past several years efforts have been made to characterize additional biologically and clinically relevant leukemia entities. At the end of the 1990's, the World Health Organization (WHO) summarized consensus practices to formulate a revised classification of tumors of hematopoietic and lymphoid tissues (summarized in Table 33.1) (Harris et al., 1999). This classification, for example, divides acute myeloid leukemias (AML) into four large subclasses (AML with recurrent cytogenetic abnormalities, with multilineage dysplasia, therapy related, and not otherwise categorized), which are further subdivided into distinct AML subtypes. However, for many leukemia classes the genetic or pathogenic events are still unknown, and within well-defined leukemia subgroups, like AML cases with a t(8;21) or an inv(16), considerable clinical heterogeneity is observed (Schlenk et al., 2004). Thus, there is a need for a refined classification based on an improved understanding of the molecular mechanisms leading to leukemia.

Leukemia Cytogenetics and Molecular Genetics

Leukemias have a long tradition of being "first", with chronic myelogenous leukemia (CML) being the first malignancy in which a recurring chromosomal abnormality, the Philadelphia chromosome, was found to result from a translocation of genetic material from one chromosome to another (Rowley, 1973). The fusion gene resulting from this translocation t(9;22), *BCR-ABL*, was then shown to be responsible for the myeloproliferation observed in CML (Konopka et al., 1985; Shtivelman et al., 1985). Representing still the gold standard for the investigation of aberrations in leukemias, chromosome banding analyses have since led to the discovery of well over 100 chromosome translocations and the identification of a number of recurring chromosomal gains and losses in leukemic cells, thereby transforming our understanding of the genetic mechanisms involved in leukemogenesis.

At the molecular level, many of these chromosomal translocations result in the deregulated expression of oncogenes like e.g. *MYC* in acute lymphoblastic leukemia (ALL). Alternatively, translocations can result in the creation of chimeric fusion proteins, many of which alter transcriptional programs to block cell differentiation. For example, in AML, t(8;21) and inv(16) create the *AML1-ETO* and *CBFB-MYH11* fusions, respectively, both of which deregulate the activity of the transcription factor complex Core Binding Factor (CBF), altering the expression of genes and disrupting cell differentiation (Frohling et al., 2005; Licht and Sternberg, 2005).

Nevertheless, while many such aberrations like e.g. t(8;21) or inv(16) can block the differentiation of myeloid cells, they are not by themselves sufficient to cause a myeloid leukemia. On the other hand, it has been shown that constitutively activated signaling molecules, such as FLT3 or RAS family members, can induce the complementary myeloproliferative phenotype. Today,

Copyright © 2010, Elsevier Inc.
All rights reserved.

Table 33.1 WHO Classification of Tumors of Hematopoietic and Lymphoid Tissues – Leukemias

	"Acute" clinical course	"Chronic" clinical course
Myeloid Origin	**AML with recurrent genetic abnormalities**	**Myeloproliferative diseases**
	AML with t(8;21)(q22;q22), (*AML1/ETO*)	Chronic myelomonocytic leukemia
	AML with inv(16)(p13q22), (*CBFβ/MYH11*)	Atypical chronic myeloid leukemia
	AML with t(15;17)(q22;q12), (*PML/RARα*)	Juvenile myelomonocytic leukemia etc.
	AML with 11q23 (*MLL*) abnormalities	
	AML with multilineage dysplasia	**Chronic myeloproliferative diseases**
	AML, therapy related	CML with t(9;22)(q34;q11), (*BCR/ABL*)
	AML not otherwise categorized	Chronic neutrophilic leukemia
		Chronic eosinophilic leukemia, etc.
Lymphoid Origin	**Precursor B-cell neoplasms**	**Mature B-cell neoplasms**
	Precursor B-ALL	CLL
	Burkitt leukemia	B-cell prolymphocytic leukemia
		Hairy cell leukemia etc.
	Precursor T-cell neoplasms	**Mature T-cell neoplasms**
	Precursor T-ALL	T-cell prolymphocytic leukemia
		T-cell granular lymphocytic leukemia, etc.

Table 33.2 Cytogenetic and Molecular Aberrations in AML

Risk category[a]	Cytogenetic findings [Genes involved]	Associated molecular abnormalities
Favorable	t(8;21)(q22;q23) [*ETO, AML1/RUNX1*]	*KIT* mutations (Exon 8 and 17, codon 816)
		CEBPA downregulation
	inv(16) (q13q22)/t(16;16) [*MYH11, CBFB*]	*KIT* mutations (Exon 8 and 17, codon 816)
		NRAS and *KRAS* mutations
		CEBPA downregulation
	del(9q)	*CEBPA* mutation
	t(15;17) (q22;q11∼q21) [*PML, RARA*][b]	*FLT3* ITD and *FLT3* activating loop mutation
Intermediate	Normal karyotype	*MLL* PTD
		FLT3 ITD and *FLT3* activating loop mutation
		CEBPA mutation
		NPM1 mutation
		BAALC over expression
	Trisomy 21	*RUNX1* mutation
	Trisomy 11	*MLL* PTD
Unfavorable	t(6;9)(p23;q34) [*DEK, CAN*]	*FLT3* ITD
	complex karyotype (3 or more aberrations)	*TP53* mutation

[a] Risk categories for overall survival according to the CALGB (Cancer and Leukemia Group B).

[b] Not included in the CALGB study, but favorable risk category in other studies.

there are further lines of evidence implicating a multistep leukemogenesis, and based on advances in molecular genetics many pathogenetically relevant mutations have been identified both in myeloid (Table 33.2) (Frohling et al., 2005; Licht and Sternberg, 2005) and lymphoid (Armstrong and Look, 2005; Pui and Evans, 2006) leukemias.

Leukemia Treatment

Regarding leukemias' tradition of being "first", acute promyelocytic leukemia (APL) was one of the first malignancies to be successfully treated with a molecularly targeted therapy, all-trans retinoic acid (ATRA). ATRA specifically targets the transforming potential of the fusion gene product, PML-RARA, that results from a t(15;17), which is commonly detected in APL. Likewise, CML was the first disorder in which a small molecule inhibitor had been designed to specifically target the disease-causing underlying molecular defect, the BCR-ABL fusion protein. Since then, additional novel drugs have been shown to be effective in leukemias, including numerous tyrosine kinase inhibitors (e.g. FLT3 inhibitors like PKC412), farnesyltransferase inhibitors (e.g. tipifarnib), demethylating agents (e.g. decitabine), histone deacetylase inhibitors (e.g. valproic acid), and monoclonal

TABLE 33.3 Prognostic Factors in Leukemias

	Clinical/Laboratory parameters	Cytogenetics	Molecular factors
AML	increasing age (↓) Secondary AML/t-AML (↓) Abnormal organ function (↓)[a] Poor performance status (↓)[a]	t(15;17), inv(16), t(8;21) (↑) normal karyotype, t(9;11), others (=) complex karyotype, inv(3) (↓) −5/del(5q), −7, others (↓)	MRD1 expression (↓) *NPM1*, *CEBPA* mutations (↑) *FLT3*, *MLL* mutations (↓) ↑ *BAALC*, *ERG* expression (↓)
CML	↑ Sokal and Hasford prognostic scores (↓) Hematologic response (↑) MRD persistence (↓) ↑ C50 imatinib (↓)	del(9q) (↓)	*OCT1* expression (↑)
ALL	High WBC (↓) Late achievement of CR (↓) Increasing age (↓) MRD persistence (↓) *In-vitro* resistance (↓) Immunophenotype: Pro-B-ALL	t(4;11)/ALL1-AF4 (↓) t(9;22)/BCR-ABL (↓) t(1;19)/PBX-E2A (↓) complex aberrant karyotype (↓)	MDR1 function (↓)
CLL	Binet/Rai stage Response to therapy (↓) ↑ lymphocyte doubling time ↑ β2-microglobulin (↓) ↑ Thymidine-kinase(↓)	del (13q14) (↑) del(17p13), del(11q22q23) (↓)	Unmutated *IgVH* status[b] (↓) V3.21 expression (↓) ↓ CD38 expression(↓) ↑ ZAP70 expression (↓)

(↓) Unfavorable;, (↑) favorable; (=) intermediate prognostic marker;
[a]Applies to all leukemias.

antibodies (e.g. the anti-CD33 antibody gemtuzumab ozogamicin) (Tallman, 2005). While these agents are currently being investigated within clinical treatment trials, the combination therapies of cytotoxic drugs with or without stem-cell transplantation are still the gold standard and have increased the cure rate of leukemia, especially in childhood acute lymphoblastic leukemia (ALL) with an overall cure rate of over 80% (Pui and Evans, 2006). However, cure rates for adults remain much lower, thereby signaling the need for treatment approaches incorporating the above mentioned novel targeted therapies. This should be based on an improved risk stratification that reflects the biological and clinical heterogeneity of leukemias in order to guide efficient patient management.

Prognostic Factors in Leukemias

In AML, patients are assigned to risk-groups based on the underlying leukemia karyotype (Byrd et al., 2002; Grimwade et al., 1998; Slovak et al., 2000), which represents one of the most powerful prognostic factors in this disease (Table 33.3). However, the identification of novel molecular markers has recently permitted to further dissect existing prognostic leukemia subclasses, like for example the large group of AML patients presenting with normal karyotype disease (Frohling et al., 2005; Licht and Sternberg, 2005). Internal tandem duplications (ITD) of the *FLT3* gene, partial tandem duplications (PTD) of the *MLL* gene, as well as mutations of *CEBPA* and *NPM1* are of prognostic relevance in this AML subgroup, as is the expression level of *BAALC*. In chronic lymphoid leukemias (CLL), cytogenetic aberrations also provide prognostic information in addition to the immunoglobulin variable heavy chain gene (V_H) mutational status (Seiler et al., 2006). Nevertheless, despite this recent progress there is still no commonly accepted risk stratification for many leukemia subtypes as the mechanisms of leukemogenesis are not yet fully understood.

GENOMICS IN LEUKEMIAS: INSIGHTS INTO LEUKEMIA BIOLOGY

Genomics-Based Class Prediction in Leukemias

Over the past several years, genomics methods have offered a range of experimental approaches to capture the molecular variation underlying the biological and clinical heterogeneity of

leukemias. Since genome-wide gene expression profiling (GEP) based on DNA microarrays represents one of the most powerful genomics tools, we will mainly focus on the impact of this technology on leukemia management, especially as it was the leukemias in which the utility and promise of GEP was first demonstrated. By analyzing AML and ALL samples, Golub *et al.*, were able to distinguish AML and ALL without previous knowledge of these leukemia classes, and, by developing a supervised class predictor, new leukemia cases could be accurately assigned to one of these two leukemia classes (Golub et al., 1999).

Since then, DNA microarray technology has already contributed significantly to the field of leukemia research (Ebert and Golub, 2004; Staudt, 2003), including the identification of a general gene dosage effect for the expression of genes located in areas of chromosomal gains and losses (Haslinger et al., 2004; Rucker et al., 2006). Based on supervised analytical approaches, in childhood ALL distinct expression profiles were identified to correlate with prognostically important leukemia subtypes like e.g. T-ALL, *E2A-PBX1*, *BCR-ABL*, *TEL-AML1*, *MLL* rearrangement, and ALL with a hyperdiploid karyotype (>50 chromosomes) (Yeoh et al., 2002), and these findings have since been confirmed in other studies of both childhood and adult ALL (Armstrong et al., 2002; Ross et al., 2003). In AML, a gene expression-based discrimination of the cytogenetically defined subgroups inv(16), t(8;21), t(15;17) and t(11q23)/*MLL* could also be demonstrated (Bullinger et al., 2004; Ross et al., 2004; Valk et al., 2004) (Figure 33.1). Notably, in acute leukemia the same classifiers have been shown to apply to both pediatric and adult leukemia cases exhibiting identical genetic aberrations (Kohlmann et al., 2004; Ross et al., 2004), and it has been demonstrated that the respective signatures are quite robust (Kohlmann et al., 2005; Mitchell et al., 2004). Thus, gene expression profiling may provide a useful alternative approach to classify known leukemia subgroups with high accuracy (Haferlach et al., 2005).

Likewise, gene expression patterns associated with molecular genetic aberrations like the V_H mutational status in CLL have been demonstrated and led to the identification ZAP-70 (Rosenwald et al., 2001), which is now used as surrogate marker for the V_H mutational status. In AML gene expression patterns associated with e.g. *FLT3* ITD (Bullinger et al., 2004; Valk et al., 2004), *CEBPA* (Valk et al., 2004), and *NPM1* mutations (Alcalay et al., 2005) have been described. However, in contrast to translocations involving the *MLL* gene (Armstrong et al., 2002), no characteristic pattern has been identified for cases with *MLL* PTD (Bullinger et al., 2004; Ross et al., 2004), thereby suggesting that cases with *MLL* PTD might be more heterogeneous at a molecular level and that not all genetic alterations in leukemia result in definably altered gene expression patterns.

Genomics-Based Class Discovery in Leukemias

By analyzing 285 AML with unsupervised analytical approaches Valk et al., identified sixteen AML subgroups that included novel prognostically-relevant subtypes of leukemia (Valk et al., 2004). While cases with favorable cytogenetics generally exhibited "homogeneous clustering", novel clusters were often characterized by specific molecular alterations, like for example MLL abnormalities or increased *EVI1* expression. On the other hand, "homogenously grouped" classes, like cases with an inv(16) or a t(8;21), were also characterized by molecular variation depending on the probe sets included in the analysis (Valk et al., 2004). Using a different unsupervised clustering approach this molecular heterogeneity within t(8;21) and inv(16) cases was also observed in a separate study (Bullinger et al., 2004), thereby suggesting that distinct patterns of gene expression within the t(8;21) and inv(16) subgroups reflect alternative cooperating events leading to transformation.

Interestingly, in this study, cases with normal karyotype also segregated mainly into two distinct groups (Figure 33.2), each of which included a small number of cases from other cytogenetic classes (Bullinger et al., 2004). While *FLT3* aberrations were more highly represented in one subgroup, FAB (French American British) M4/M5 morphologic subtypes were more prevalent in the other subgroup, and Kaplan-Meier analysis identified a statistically significant difference in overall survival between the two subclasses (Bullinger et al., 2004). In agreement, Valk and colleagues also identified normal karyotype-predominated clusters associated with *FLT3* ITD, as well as a cluster including mainly specimens from AML patients whose blasts displayed FAB M4 or M5 morphology (Valk et al., 2004).

GENOMICS IN LEUKEMIAS: EVALUATION OF DRUG EFFECTS

Molecular Signatures of Anti-Leukemic Drugs

Leukemia was again one of the first diseases in which DNA microarray technology was applied for monitoring drug effects by analyzing the effect of all-trans retinoic acid (ATRA) treatment in APL-derived cell lines (Tamayo et al., 1999). These analyses showed that ATRA-regulated genes included members of the tumor necrosis factor (TNF) pathway, suggesting that this pathway might intersect with ATRA signaling and play a role in regulating cell survival in response to ATRA. GEP has also been successfully used to evaluate the sensitivity of CML to imatinib mesylate (Tipping et al., 2003), which targets the ABL kinase activity of the BCR-ABL fusion. Using a cell line model, differentially-expressed genes correlated with imatinib mesylate resistance could be identified, suggesting that alternative pathways maintain viability and promote growth independently of BCR-ABL.

Similarly, Rosenwald et al., have recently investigated the molecular consequences of fludarabine treatment of CLL patient samples using GEP (Rosenwald et al., 2004). Both *in vitro* and *in vivo* exposure to fludarabine, a purine analog currently used in standard CLL treatment regimens, resulted in a consistent "response signature" characterized by p53 target genes and genes involved in DNA repair, thus, providing a molecular explanation for the drug resistance and aggressive clinical course often seen in p53-mutated CLL patients. This analysis further suggested

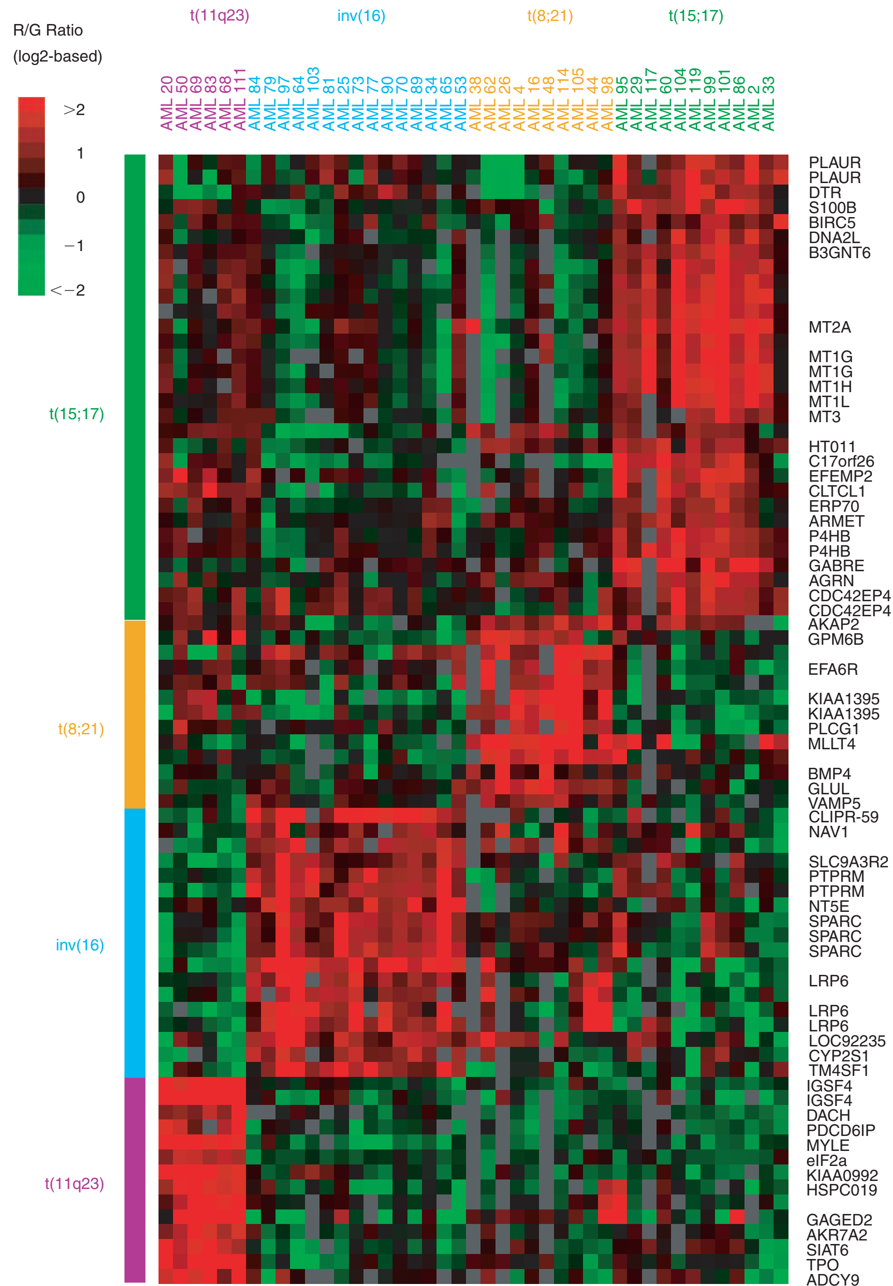


Figure 33.1 Gene expression signatures of leukemia cytogenetic classes. Shown is a “heat map” representation of selected genes identified by supervised analysis whose expression is significantly correlated with specific AML cytogenetic aberrations, like t(11q23), inv(16), t(8;21), and t(15;17). Gene-expression levels are depicted in a pseudo color scale as indicated (red indicates higher expression levels). The analysis shown was performed using publicly available microarray data that were based on a supervised analysis (Bullinger et al., 2004). Gene-expression signatures such as these can be used to classify the cytogenetic group of new specimens with high accuracy. AML 64 was initially characterized as “normal karyotype” by cytogenetic banding analysis, but for which RT-PCR subsequently identified the diagnostic *CBFB-MYH11* fusion transcript characteristic of inv(16) cases.

the importance of only treating patients that warrant therapy, as fludarabine treatment might select for p53 mutant CLL cells (Rosenwald et al., 2004). Another recent study sought to provide a better molecular understanding of resistance to L-asparaginase, an important component of most treatment regimens for ALL. *In vitro* exposure to L-asparaginase in cell lines and pediatric ALL samples followed by gene expression profiling revealed changes that reflect a consistent coordinated response to asparagine starvation, which is independent of asparagine synthetase base-line expression levels (Fine et al., 2005). Thus, targeting

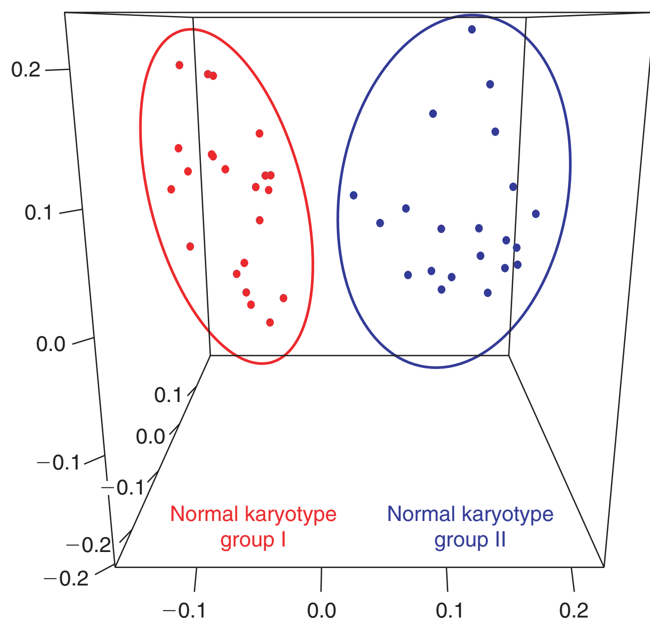


Figure 33.2 Discovery of new molecular subtypes of leukemia. Shown are the results of an unsupervised principal components analysis (PCA) displaying a projection of the first three principal components of variable gene expression for AML specimens with normal karyotype. Based on distinct patterns of genes expression, PCA identifies two novel subgroups of normal karyotype AML cases (indicated by red and blue dots, respectively). Group II cases exhibit more frequent myelomonocytic differentiation, while group I cases more often harbor *FLT3* mutations and are associated with shorter overall survival. Data are from Bullinger et al. (2004).

particular "amino acid starvation response" genes might provide a novel way to overcome L-asparaginase resistance of ALL cells.

Drug Response Prediction in Leukemia

To determine whether the cellular responses provoked by chemotherapeutic agents can be predicted, Cheok and colleagues profiled gene expression in childhood ALL cells before and after *in vivo* treatment with methotrexate and mercaptopurine, given either alone or in combination (Cheok et al., 2003). A gene expression pattern consisting of 124 genes, which included genes involved in apoptosis, mismatch repair, cell cycle control and stress response, accurately discriminated among the assigned treatments. Notably, this signature showed differences in cellular response to drug combinations versus single agents and indicated that common pathways of genomic response to the same treatment are shared by different ALL subtypes. Testing for *in vitro* sensitivity to prednisolone, vincristine, asparaginase, and daunorubicin, the identification of differentially expressed genes in drug-sensitive and drug-resistant ALL leukemia cells allowed the creation of a combined gene-expression score of resistance, which was shown to be of prognostic relevance in childhood ALL (Holleman et al., 2004).

Discovery of Novel Anti-Leukemic Drugs

Today, genomics approaches including DNA microarray technology play an important part in the drug target discovery process, not only by providing a valuable tool for the optimization and clinical validation of novel compounds. While at the beginning of the process GEP allows the identification and prioritization of potential therapeutic targets, subsequent expression profiling assists in drug discovery and toxicology. Then, various bioinformatics approaches are used to deduce from expression profiles the mechanism of action of new drugs as well as off-target effects. However, it is important to keep in mind the limitations of evaluating drug responses through measurements of mRNA abundance alone. Nevertheless, GEP provides a powerful tool in pharmacogenomics that promises an improved prediction of prognosis and drug response (Walgren et al., 2005).

In a recent example of the utility of GEP in drug discovery, a GEP-based high-throughput screening approach was applied to screen for chemical compounds with differentiation-inducing activity in leukemia (Stegmaier et al., 2004). Following definition and validation of a microarray-based differentiation signature, a high throughput screening method was designed to detect the respective gene pattern. Using this approach, treatment of a leukemia cell line with 1,739 different compounds revealed eight chemicals that reliably induced the differentiation signature (Stegmaier et al., 2004). Interestingly, one of these compounds inhibited epidermal growth factor receptor (EGFR) kinase activity, and the authors could show in a subsequent study that the Food and Drug Administration (FDA)–approved EGFR inhibitor gefitinib similarly promoted the differentiation of AML cell lines and primary patient–derived AML blasts *in vitro* (Stegmaier et al., 2005). Notably, the analyzed AML cells did not express EGFR, indicating an EGFR-independent mechanism of gefitinib induced differentiation in AML, thereby suggesting the presence of additional yet unknown key targets.

GENOMICS IN LEUKEMIAS: CLINICAL OUTCOME PREDICTION

Identification of Novel Surrogate Markers in Leukemia

Many groups have now demonstrated that GEP allows the identification of specific signatures correlated with and the subsequent prediction of known "favorable-risk" cytogenetic aberrations in both AML like e.g. cases with inv(16), t(8;21), or t(15;17) (Bullinger et al., 2004; Haferlach et al., 2005; Valk et al., 2004), and in ALL like e.g. cases with t(12;21) or hyperdiploidy (more than 50 chromosomes per leukemia cell) (Yeoh et al., 2002). Similarly, "unfavorable-risk" cytogenetics subgroups can be predicted, as well as prognostically relevant molecular genetic aberrations (Bullinger and Valk, 2005). Such analyses led to the identification of novel surrogate markers among which ZAP-70 represents the first to be clinically implemented. Coding for a

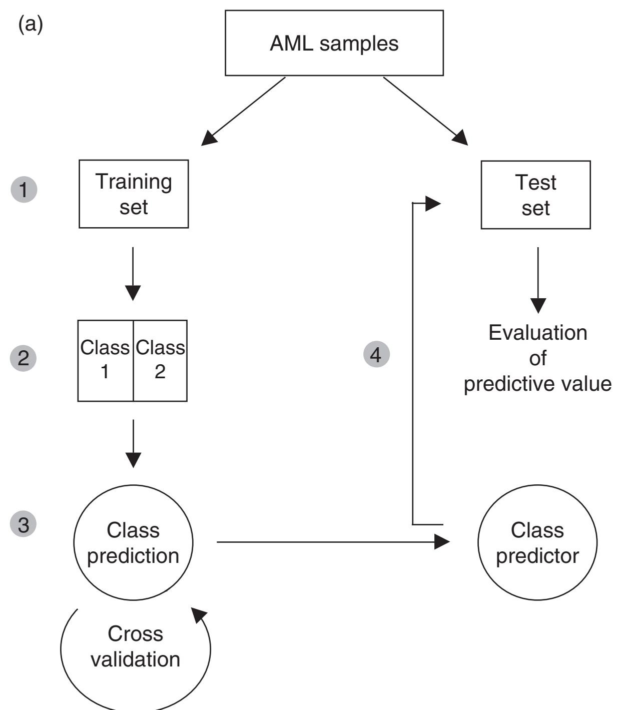
tyrosine kinase essential for T cell signaling, *ZAP-70* exhibited five-fold higher expression levels in patients with unmutated V_H genes compared to those in which the B cell clone had undergone the rearrangement of the immunoglobulin heavy-chain variable region (Rosenwald et al., 2001). The clinical usefulness of this novel marker, which can be readily measured at the protein level by flow cytometry, has recently been validated (Orchard et al., 2004).

Furthermore, GEP-based prediction of leukemias sensitive or resistant to chemotherapeutic agents can provide significant information regarding treatment outcome in leukemias. The fact that initial findings could be confirmed in independent sets of patients in ALL (Holleman et al., 2004; Lugthart et al., 2005) and AML (Heuser et al., 2005) suggests the feasibility of an improved treatment management in leukemias based on GEP. Furthermore, expression profiling might be used to predict leukemia cases with minimal residual disease (MRD) following treatment at high accuracy. Based on the hypothesis that this treatment resistance is reflected by an intrinsic feature of leukemia cells at the time of diagnosis, a prognostic signature predicting the MRD load following therapy could be defined (Cario et al., 2005).

Identification of Novel Prognostic Markers and Signatures in Leukemias

Using supervised approaches, many groups have tried to generate signatures correlated with good and poor outcome. However, in contrast to signatures correlated with resistance to chemotherapy, in our experience supervised signatures predictive of good and poor outcome have generally not been validated in independent data sets. This might be secondary to the limitation of strictly supervised approaches in outcome prediction as survival and survival time are impacted by many things other then the tumor cells themselves, and thus are likely to be very noisy surrogates for the underlying prognostically relevant tumor subclasses. In order to discover gene-expression signatures with prognostic value in addition to signatures correlated with cytogenetics, novel semi-supervised strategies combining the strengths of supervised and unsupervised approaches might be helpful in leukemia research (Bair and Tibshirani, 2004).

Semi-supervised approaches use the subset of genes correlating with survival time for supervised clustering (or supervised principal components analysis) of specimens to reveal the underlying prognostically relevant tumor subtypes, and then to build a predictor for these subtypes (Figure 33.3). Applying this approach to a cohort of AML patients, a 133 gene signature could be defined and validated as a significant independent outcome predictor, both across all cytogenetic classes and within the large subset of clinically-important AML cases with normal karyotype (Bullinger et al., 2004). Recently, this signature has been validated by applying it to an independent set of AML cases with normal karyotype (Marcucci et al., 2006). While these findings are definitely encouraging, further validation of results in larger cohorts and in independent studies are required before clinical implementation becomes feasible in leukemias.



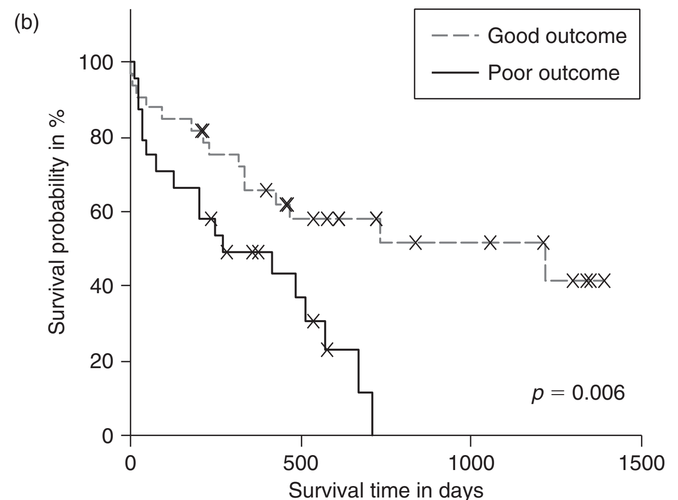


Figure 33.3 Semi-supervised approach for leukemia outcome prediction. (a) Schematic overview of a supervised clustering strategy. (1) AML specimens are randomized into separate training and test sets. (2) In the training set, genes whose expression correlates with survival are used to cluster samples into favorable and unfavorable outcome classes. (3) An optimal gene expression predictor is constructed for these outcome classes. (4) The outcome class predictor is then validated by predicting outcomes in the independent test set. (b) Evaluation of outcome predictor. Kaplan-Meier survival analysis of the independent test set validates the gene expression classifier (here comprising 133 genes) as a significant predictor of overall survival. Data are from Bullinger et al. (2004).

2009 UPDATE

During the last decade, microarray-based genomic approaches have contributed significantly to leukemia research, as demonstrated by microarray-based GEP findings in AML (Baldus and Bullinger, 2008; Wouters et al., 2009). Since its invention, GEP has contributed an important new facet to the exploration of hematologic malignancies; recent studies have confirmed its impact in leukemia classification (Verhaak et al., 2009) and have provided evidence that GEP might improve classification with regard to clinical importance (Bullinger et al., 2008). Furthermore, in AML, GEP-based class discovery approaches have identified novel core binding factor leukemia subgroups (Bullinger et al., 2007) and CEBPA mutated/silenced subgroups (Wouters et al., 2007). Finally, advances in prediction of treatment response to the farnesyltransferase inhibitor tipifarnib (Raponi et al., 2008) or the immunomodulatory drug lenalidomide (Ebert et al., 2008), as well as advances in outcome prediction (Metzeler et al., 2008), point to a potential usefulness of GEP-based molecular leukemia classification in the future. Increasingly, microarray assays will be applied to adequately diagnose leukemias, predict their course, and guide individualized treatment approaches.

Recently, novel molecular subgroups have been identified both in cytogenetically normal leukemia cases (Schlenk et al., 2008) and in leukemia subtypes characterized by distinct cytogenetic aberrations (Dohner and Dohner, 2008). However, first array CGH-based studies dissecting the genomic heterogeneity of leukemia by parallel analysis of global gene expression changes and copy number alterations (CNAs) provided not only useful information regarding novel leukemia relevant candidate (Rucker et al., 2006) but also suggested that genomic aberrations associated with leukemogenesis are likely to be far more complex than initially anticipated. Novel genome-wide single nucleotide polymorphism microarray-based analyses (SNP array) have not only revealed large-scale cryptic regions of acquired homozygosity in form of segmental uniparental disomy (Fitzgibbon et al., 2005; Gupta et al., 2008) but also submicroscopic CNAs and numerous germline copy number variations (CNVs) (Maciejewski and Mufti, 2008). For example, in ALL, a genome-wide analysis of leukemic cells using high-resolution SNP arrays and genomic DNA sequencing revealed deletion, amplification, point mutation and structural rearrangement in genes involved in B-cell development, and differentiation in B-progenitor ALL cases (Mullighan et al., 2007). Besides the *PAX5* gene, which was the most frequent target of somatic mutation being altered in almost one-third of cases (Mullighan et al., 2007), *IKZF1* (encoding the transcription factor Ikaros) was deleted in over 80% of *BCR/ABL1* positive ALL cases (Mullighan et al., 2008). *IKZF1* deletions resulted either in the expression of a dominant-negative Ikaros isoform or the complete loss of Ikaros expression, thereby suggesting that genetic lesions resulting in the loss of Ikaros function are an important event in the development of *BCR/ABL1* ALL (Mullighan et al., 2008). Thus, these data demonstrate the power of high-resolution, genome-wide SNP array approaches to identify new molecular lesions in leukemia.

In addition to genome-wide profiling for genomic aberrations, recent progress in profiling the expression of microRNAs (miRNAs), small nonprotein coding RNAs that play a crucial role in the transcriptional and posttranscriptional regulation of genes (Calin and Croce 2006), shed further light on the molecular heterogeneity of leukemias (Jongen-Lavrencic et al., 2008; Lu et al., 2005). Interestingly, distinct miRNA signatures might also be associated with outcome in both chronic (Calin et al., 2005) as well as acute leukemia (Marcucci et al., 2008). Next, microarrays also offer the possibility to screen for global epigenetic changes in leukemias (Esteller, 2007), and recent studies have already successfully demonstrated the power of this technology in investigating acute leukemia biology (Gebhard et al., 2006; Taylor et al., 2007). Furthermore, novel advances such as high-throughput sequencing technology that enable the timely sequence analysis of an entire genome might reveal novel leukemia-relevant gene aberrations, as suggested by the recent publication of an entire AML genome (Ley et al., 2008).

In the future, integrative analyses of high-dimensional genomics datasets, like SNP array and sequencing data, global transcriptome and epigenome information, as well as proteomics data (Kornblau et al., 2009), will allow the extraction of additional biological insights into leukemia pathogenesis. However, a prerequisite for a successful detection of leukemia relevant aberration will rely on additional functional analyses that possess the power to distinguish between driver and passenger aberrations in leukemia cells (Frohling et al., 2007).

2009 UPDATE REFERENCES

Baldus, C.D. and Bullinger, L. (2008). Gene expression with prognostic implications in cytogenetically normal acute myeloid leukemia. *Semin Oncol* 35(4), 356-364.

Bullinger, L., Dohner, K. et al. (2008). An FLT3 gene-expression signature predicts clinical outcome in normal karyotype AML. *Blood* 111(9), 4490-4495.

Bullinger, L., Rucker, F.G. et al. (2007). Gene-expression profiling identifies distinct subclasses of core binding factor acute myeloid leukemia. *Blood* 110(4), 1291-1300.

Calin, G.A. and Croce, C.M. (2006). MicroRNA signatures in human cancers. *Nat Rev Cancer* 6(11), 857-866.

Calin, G.A., Ferracin, M. et al. (2005). A microRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N Engl J Med* 353(17), 1793-1801.

Dohner, K. and Dohner, H. (2008). Molecular characterization of acute myeloid leukemia. *Haematologica* 93(7), 976-982.

Esteller, M. (2007). Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet* 8(4), 286-298.

Fitzgibbon, J., Smith, L.L. et al. (2005). Association between acquired uniparental disomy and homozygous gene mutation in acute myeloid leukemias. *Cancer Res* 65(20), 9152-9154.

Frohling, S., Scholl, C. et al. (2007). Identification of driver and passenger mutations of FLT3 by high-throughput DNA sequence analysis and functional assessment of candidate alleles. *Cancer Cell* 12(6), 501-513.

Gebhard, C., Schwarzfischer, L. et al. (2006). Genome-wide profiling of CpG methylation identifies novel targets of aberrant hypermethylation in myeloid leukemia. *Cancer Res* 66(12), 6118-6128.

Gupta, M., Raghavan, M. et al. (2008). Novel regions of acquired uniparental disomy discovered in acute myeloid leukemia. *Genes Chromosomes Cancer* 47(9), 729-739.

Jongen-Lavrencic, M., Sun, S.M. et al. (2008). MicroRNA expression profiling in relation to the genetic heterogeneity of acute myeloid leukemia. *Blood* 111(10), 5078-5085.

Kornblau, S.M., Tibes, R. et al. (2009). Functional proteomic profiling of AML predicts response and survival. *Blood*113(1), 154-164.

Ley, T.J., Mardis, E.R. et al. (2008). DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 456(7218), 66-72.

Lu, J., Getz, G. et al. (2005). MicroRNA expression profiles classify human cancers. *Nature* 435(7043), 834-838.

Maciejewski, J.P. and Mufti, G.J. (2008). Whole genome scanning as a cytogenetic tool in hematologic malignancies. *Blood* 112(4), 965-974.

Marcucci, G., Radmacher, M.D. et al. (2008). MicroRNA expression in cytogenetically normal acute myeloid leukemia. *N Engl J Med* 358(18), 1919-1928.

Metzeler, K.H., Hummel, M. et al. (2008). An 86-probe-set gene-expression signature predicts survival in cytogenetically normal acute myeloid leukemia. *Blood* 112(10), 4193-4201.

Mullighan, C.G., Goorha, S. et al. (2007). Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* 446(7137), 758-764.

Mullighan, C.G., Miller, C.B. et al. (2008). BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. *Nature* 453(7191), 110-114.

Raponi, M., Lancet, J.E. et al. (2008). A 2-gene classifier for predicting response to the farnesyltransferase inhibitor tipifarnib in acute myeloid leukemia. *Blood* 111(5), 2589-2596.

Rucker, F.G., Bullinger, L. et al. (2006). Disclosure of candidate genes in acute myeloid leukemia with complex karyotypes using microarray-based molecular characterization. *J Clin Oncol* 24(24), 3887-3894.

Schlenk, R.F., Dohner, K. et al. (2008). Mutations and treatment outcome in cytogenetically normal acute myeloid leukemia. *N Engl J Med* 358(18), 1909-1918.

Taylor, K.H., Pena-Hernandez, K.E. et al. (2007). Large-scale CpG methylation analysis identifies novel candidate genes and reveals methylation hotspots in acute lymphoblastic leukemia. *Cancer Res* 67(6), 2617-2625.

Verhaak, R.G., Wouters, B.J. et al. (2009). Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. *Haematologica* 94(1), 131-134.

Wouters, B.J., Jorda, M.A. et al. (2007). Distinct gene expression profiles of acute myeloid/T-lymphoid leukemia with silenced CEBPA and mutations in NOTCH1. *Blood* 110(10), 3706-3714.

Wouters, B.J., Lowenberg, B. et al. (2009). A decade of genome-wide gene expression profiling in acute myeloid leukemia: Flashback and prospects. *Blood* 113(2), 291-298.

ACKNOWLEDGMENTS

The authors have no conflicts of interest and have nothing to disclose.

REFERENCES

Alcalay, M., Tiacci, E., Bergomas, R., Bigerna, B., Venturini, E., Minardi, S.P., Meani, N., Diverio, D., Bernard, L., Tizzoni, L., et al.. (2005). Acute myeloid leukemia bearing cytoplasmic nucleophosmin (NPMc+ AML) shows a distinct gene expression profile characterized by up-regulation of genes involved in stem-cell maintenance. *Blood* 106, 899–902.

Armstrong, S.A. and Look, A.T. (2005). Molecular genetics of acute lymphoblastic leukemia. *J Clin Oncol* 23, 6306–6315.

Armstrong, S.A., Staunton, J.E., Silverman, L.B., Pieters, R., den Boer, M.L., Minden, M.D., Sallan, S.E., Lander, E.S., Golub, T.R. and Korsmeyer, S.J. (2002). MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat Genet* 30, 41–47.

Bair, E. and Tibshirani, R. (2004). Semi-supervised methods to predict patient survival from gene expression data. *PLoS Biol* 2, E108.

Bullinger, L. and Valk, P.J. (2005). Gene expression profiling in acute myeloid leukemia. *J Clin Oncol* 23, 6296–6305.

Bullinger, L., Dohner, K., Bair, E., Frohling, S., Schlenk, R.F., Tibshirani, R., Dohner, H. and Pollack, J.R. (2004). Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *N Engl J Med* 350, 1605–1616.

Byrd, J.C., Mrozek, K., Dodge, R.K., Carroll, A.J., Edwards, C.G., Arthur, D.C., Pettenati, M.J., Patil, S.R., Rao, K.W., Watson, M.S. et al. (2002). Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative incidence of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: Results from Cancer and Leukemia Group B (CALGB 8461). *Blood* 100, 4325–4336.

Cario, G., Stanulla, M., Fine, B.M., Teuffel, O., Neuhoff, N.V., Schrauder, A., Flohr, T., Schafer, B.W., Bartram, C.R., Welte, K. et al. (2005). Distinct gene expression profiles determine molecular treatment

response in childhood acute lymphoblastic leukemia. *Blood* 105, 821–826.

Cheok, M.H., Yang, W., Pui, C.H., Downing, J.R., Cheng, C., Naeve, C.W., Relling, M.V. and Evans, W.E. (2003). Treatment-specific changes in gene expression discriminate *in vivo* drug response in human leukemia cells. *Nat Genet* 34, 85–90.

Ebert, B.L. and Golub, T.R. (2004). Genomic approaches to hematologic malignancies. *Blood* 104, 923–932.

Ebert, B.L., Galili, N., Tamayo, P., Bosco, J., Mak, R., Pretz, J., Tanguturi, S., Ladd-Acosta, C., Stone, R., Golub, T.R. et al. (2008). An erythroid differentiation signature predicts response to lenalidomide in myelodysplastic syndrome. PLoS Med 5(2), e35.

Fine, B.M., Kaspers, G.J., Ho, M., Loonen, A.H. and Boxer, L.M. (2005). A genome-wide view of the *in vitro* response to l-asparaginase in acute lymphoblastic leukemia. *Cancer Res* 65, 291–299.

Frohling, S., Scholl, C., Gilliland, D.G. and Levine, R.L. (2005). Genetics of myeloid malignancies: Pathogenetic and clinical implications. *J Clin Oncol* 23, 6285–6295.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A. et al. (1999). Molecular classification of cancer: class discovery and Class prediction by gene expression monitoring. *Science* 286, 531–537.

Grimwade, D., Walker, H., Oliver, F., Wheatley, K., Harrison, C., Harrison, G., Rees, J., Hann, I., Stevens, R., Burnett, A. et al. (1998). The importance of diagnostic cytogenetics on outcome in AML: Analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties. *Blood* 92, 2322–2333.

Haferlach, T., Kohlmann, A., Schnittger, S., Dugas, M., Hiddemann, W., Kern, W. and Schoch, C. (2005). Global approach to the diagnosis of leukemia using gene expression profiling. *Blood* 106, 1189–1198.

Harris, N.L., Jaffe, E.S., Diebold, J., Flandrin, G., Muller-Hermelink, H.K., Vardiman, J., Lister, T.A. and Bloomfield, C.D. (1999). World Health Organization classification of neoplastic diseases of the hematopoietic and lymphoid tissues: Report of the Clinical Advisory Committee meeting-Airlie House, Virginia, November 1997. *J Clin Oncol* 17, 3835–3849.

Haslinger, C., Schweifer, N., Stilgenbauer, S., Dohner, H., Lichter, P., Kraut, N., Stratowa, C. and Abseher, R. (2004). Microarray gene expression profiling of B-cell chronic lymphocytic leukemia subgroups defined by genomic aberrations and VH mutation status. *J Clin Oncol* 22, 3937–3949.

Heuser, M., Wingen, L.U., Steinemann, D., Cario, G., von Neuhoff, N., Tauscher, M., Bullinger, L., Krauter, J., Heil, G., Dohner, H. et al. (2005). Gene-expression profiles and their association with drug resistance in adult acute myeloid leukemia. *Haematologica* 90, 1484–1492.

Holleman, A., Cheok, M.H., den Boer, M.L., Yang, W., Veerman, A.J., Kazemier, K.M., Pei, D., Cheng, C., Pui, C.H., Relling, M.V. et al. (2004). Gene-expression patterns in drug-resistant acute lymphoblastic leukemia cells and response to treatment. *N Engl J Med* 351, 533–542.

Kohlmann, A., Schoch, C., Schnittger, S., Dugas, M., Hiddemann, W., Kern, W. and Haferlach, T. (2004). Pediatric acute lymphoblastic leukemia (ALL) gene expression signatures classify an independent cohort of adult ALL patients. *Leukemia* 18, 63–71.

Kohlmann, A., Schoch, C., Dugas, M., Rauhut, S., Weninger, F., Schnittger, S., Kern, W. and Haferlach, T. (2005). Pattern robustness of diagnostic gene expression signatures in leukemia. *Genes Chromosomes Cancer* 42, 299–307.

Konopka, J.B., Watanabe, S.M., Singer, J.W., Collins, S.J. and Witte, O.N. (1985). Cell lines and clinical isolates derived from Ph1-positive chronic myelogenous leukemia patients express c-abl proteins with a common structural alteration. *Proc Natl Acad Sci USA* 82, 1810–1814.

Krober, A., Bloehdorn, J., Hafner, S., Buhler, A., Seiler, T., Kienle, D., Winkler, D., Bangerter, M., Schlenk, R.F., Benner, A. et al. (2006). Additional genetic high-risk features such as 11q deletion, 17p deletion, and V3–21 usage characterize discordance of ZAP-70 and VH mutation status in chronic lymphocytic leukemia. *J Clin Oncol* 24, 969–975.

Licht, J.D. and Sternberg, D.W. (2005). The molecular pathology of acute myeloid leukemia. *Hematology (Am Soc Hematol Educ Program)*, 137–142.

Lugthart, S., Cheok, M.H., den Boer, M.L., Yang, W., Holleman, A., Cheng, C., Pui, C.H., Relling, M.V., Janka-Schaub, G.E., Pieters, R. et al. (2005). Identification of genes associated with chemotherapy crossresistance and treatment response in childhood acute lymphoblastic leukemia. *Cancer Cell* 7, 375–386.

Marcucci, G., Radmacher, M.D., Ruppert, A.S., Mrozek, K., Kolitz, J.E., Whitman, S.P., Edwards, C.G., Vardiman, J.W., Caligiuri, M.A., Carroll, A.J. et al. (2006). Independent Validation of Prognostic Relevance of a Previously Reported Gene-Expression Signature in Acute Myeloid Leukemia (AML) with Normal Cytogenetics (NC): A Cancer and Leukemia Group B (CALGB) Study. *Blood, ASH Annual Meeting Abstracts* 106, 755.

Mitchell, S.A., Brown, K.M., Henry, M.M., Mintz, M., Catchpoole, D., LaFleur, B. and Stephan, D.A. (2004). Inter-platform comparability of microarrays in acute lymphoblastic leukemia. *BMC Genomics* 5, 71.

Orchard, J.A., Ibbotson, R.E., Davis, Z., Wiestner, A., Rosenwald, A., Thomas, P.W., Hamblin, T.J., Staudt, L.M. and Oscier, D.G. (2004). ZAP-70 expression and prognosis in chronic lymphocytic leukaemia. *Lancet* 363, 105–111.

Pui, C.H. and Evans, W.E. (2006). Treatment of acute lymphoblastic leukemia. *N Engl J Med* 354, 166–178.

Rosenwald, A., Alizadeh, A.A., Widhopf, G., Simon, R., Davis, R.E., Yu, X., Yang, L., Pickeral, O.K., Rassenti, L.Z., Powell, J. et al. (2001). Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia. *J Exp Med* 194, 1639–1647.

Rosenwald, A., Chuang, E.Y., Davis, R.E., Wiestner, A., Alizadeh, A.A., Arthur, D.C., Mitchell, J.B., Marti, G.E., Fowler, D.H., Wilson, W.H. et al. (2004). Fludarabine treatment of patients with chronic lymphocytic leukemia induces a p53-dependent gene expression response. *Blood* 104, 1428–1434.

Ross, M.E., Zhou, X., Song, G., Shurtleff, S.A., Girtman, K., Williams, W.K., Liu, H.C., Mahfouz, R., Raimondi, S.C., Lenny, N. et al. (2003). Classification of pediatric acute lymphoblastic leukemia by gene expression profiling. *Blood* 102, 2951–2959.

Ross, M.E., Mahfouz, R., Onciu, M., Liu, H.C., Zhou, X., Song, G., Shurtleff, S.A., Pounds, S., Cheng, C., Ma, J. et al. (2004). Gene expression profiling of pediatric acute myelogenous leukemia. *Blood* 104, 3679–3687.

Rowley, J.D. (1973). Identificaton of a translocation with quinacrine fluorescence in a patient with acute leukemia. *Ann Genet* 16, 109–112.

Schlenk, R.F., Benner, A., Krauter, J., Buchner, T., Sauerland, C., Ehninger, G., Schaich, M., Mohr, B., Niederwieser, D., Krahl, R. et al.

(2004). Individual patient data-based meta-analysis of patients aged 16 to 60 years with core binding factor acute myeloid leukemia: A survey of the German Acute Myeloid Leukemia Intergroup. *J Clin Oncol* 22, 3741–3750.

Seiler, T., Dohner, H. and Stilgenbauer, S. (2006). Risk stratification in chronic lymphocytic leukemia. *Semin Oncol* 33, 186–194.

Shtivelman, E., Lifshitz, B., Gale, R.P. and Canaani, E. (1985). Fused transcript of ABL and BCR genes in chronic myelogenous leukaemia. *Nature* 315, 550–554.

Slovak, M.L., Kopecky, K.J., Cassileth, P.A., Harrington, D.H., Theil, K.S., Mohamed, A., Paietta, E., Willman, C.L., Head, D.R., Rowe, J.M. et al. (2000). Karyotypic analysis predicts outcome of preremission and postremission therapy in adult acute myeloid leukemia: A Southwest Oncology Group/Eastern Cooperative Oncology Group Study. *Blood* 96, 4075–4083.

Staudt, L.M. (2003). Molecular diagnosis of the hematologic cancers. *N Engl J Med* 348, 1777–1785.

Stegmaier, K., Ross, K.N., Colavito, S.A., O'Malley, S., Stockwell, B.R. and Golub, T.R. (2004). Gene expression-based high-throughput screening(GE-HTS) and application to leukemia differentiation. *Nat Genet* 36, 257–263.

Stegmaier, K., Corsello, S.M., Ross, K.N., Wong, J.S., Deangelo, D.J. and Golub, T.R. (2005). Gefitinib induces myeloid differentiation of acute myeloid leukemia. *Blood* 106, 2841–2848.

Tallman, M.S. (2005). New strategies for the treatment of acute myeloid leukemia including antibodies and other novel agents. *Hematology (Am Soc Hematol Educ Program)*, 143–150.

Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E.S. and Golub, T.R. (1999). Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. *Proc Natl Acad Sci USA* 96, 2907–2912.

Tipping, A.J., Deininger, M.W., Goldman, J.M. and Melo, J.V. (2003). Comparative gene expression profile of chronic myeloid leukemia cells innately resistant to imatinib mesylate. *Exp Hematol* 31, 1073–1080.

Valk, P.J., Verhaak, R.G., Beijen, M.A., Erpelinck, C.A., Barjesteh van Waalwijk van Doorn-Khosrovani, S., Boer, J.M., Beverloo, H.B., Moorhouse, M.J., van der Spek, P.J., Lowenberg, B. et al. (2004). Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med* 350, 1617–1628.

Walgren, R.A., Meucci, M.A. and McLeod, H.L. (2005). Pharmacogenomic discovery approaches: Will the real genes please stand up?. *J Clin Oncol* 23, 7342–7349.

Yeoh, E.J., Ross, M.E., Shurtleff, S.A., Williams, W.K., Patel, D., Mahfouz, R., Behm, F.G., Raimondi, S.C., Relling, M.V., Patel, A. et al. (2002). Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* 1, 133–143.

RECOMMENDED RESOURCES

American Society of Hematology http://www.hematology.org/
European Hematology Association http://www.ehaweb.org/
National Cancer Institute (NCI) http://www.cancer.gov
Gene expression omnibus http://www.ncbi.nlm.nih.gov/geo/

CHAPTER

Genomics in the Diagnosis and Management of Lung Cancer

Hasmeena Kathuria, Avrum Spira and Jerome Brody

INTRODUCTION

In the early 1900s, lung cancer was a rare disease; today it is the most common cause of cancer death in the United States and in the world. Worldwide in 2006, there will be >1.2 million lung cancer deaths. In the United States alone, there will be 174,000 new cases of lung cancer in 2006, with 162,000 deaths, accounting for 31% of all cancer deaths in men and 26% of all cancer deaths in women. The incidence of lung cancer tracks, with a 30–40 year lag, the frequency of cigarette smoking in a population. Smoking incidence and lung cancer rates are similar to the United States in Western Europe, but are 30% higher in Eastern Europe and in China where 70% of men smoke, and an epidemic of lung cancer is predicted in the mid-21st century.

The fact that only 10–15% of smokers develop lung cancer suggests there are genetic factors that influence individual susceptibility to the carcinogenic effects of cigarette smoke. However, the late age at which lung cancer usually develops limits traditional genetic studies, since parents and siblings are not often available to study. Unfortunately, survival rates for newly diagnosed lung cancer remain at ~15%, virtually unchanged in the past 4–5 decades, largely because the disease is most often diagnosed at an advanced stage and there is presently no way to determine which of the 10–15% of current and former smokers are at highest risk for developing lung cancer.

This chapter summarizes the progress that has been made in understanding the molecular mechanisms and pathogenesis of lung cancer and attempts to describe how gene expression profiling on human lung cancer, airway, and peripheral blood specimens is leading to new approaches in lung cancer screening, classification, diagnosis, prognosis, and treatment. This review focuses mainly on the genomic studies of lung cancer, including a brief discussion on the findings of methylation and microRNA (miRNA) arrays in lung tumorigenesis. Studies using single nucleotide polymorphism (SNP) and protein arrays are beyond the scope of this review, as are studies of *in vitro* cell lines and mouse models. Although we have attempted to be as inclusive as possible, our list of studies reviewed is likely to be incomplete.

The section on "Early Diagnosis/Screening of Lung Cancer" provides a brief overview of the hereditary factors that may explain susceptibility to the carcinogenic effects of smoking and describes the polymorphisms in genes involved in the metabolism of the toxic components of cigarette smoke. This section also summarizes the current studies that use genomic profiling to identify the subset of smokers who are at high risk for developing lung cancer and highlights the importance of developing biomarkers and chemoprevention strategies aimed at high-risk current or former smokers. Although this section briefly discusses epigenetic changes that have been associated with smoking, discussions on other genetic changes, including cytogenetic changes and somatic mutations, are beyond the scope of this chapter.

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

The section on "Classification and Prognosis" provides a brief overview of the different categories of genes known to be involved in tumorigenesis, including oncogenes, growth factors and their receptors, and tumor suppressor genes. Given the complexity of lung cancer, these tumors are likely dependent on more than one oncogenic signaling pathway. This section sites examples of how genomics has advanced our ability to predict the dysregulation of various oncogenic pathways in lung cancer (both the mutated gene product itself and its downstream targets), thus potentially offering an opportunity for developing new therapeutic drugs that are pathway-specific. This section focuses in some detail on recent advances in understanding epidermal growth factor receptor (EGFR)-mediated survival signals.

The section on "Pathogenesis and Treatment of Lung Cancer" begins by describing the tumor, node, metastases (TNM) tumor classification system. Although tumor staging is currently the most important prognostic variable for predicting survival, some of the limitations of the current system include the fact that patients diagnosed with similar stage and treated using similar protocols, often respond quite differently and have varying survival rates. This section describes how genomic expression profiling of lung cancer has identified prognostic genes and identified subtypes of adenocarcinomas, thus potentially guiding clinical decision-making in the future by identifying "high risk" lung cancer patients that would benefit from improved diagnostic and treatment options.

Gene expression profiling, combined with genetics, clinical information, proteomics, and imaging studies can be applied to developing tools for risk-assessment, early diagnosis, and new approaches for individualized treatment. With stronger working relationships and collaborations between bench scientists and their clinical counterparts and establishment of large databases with standardized methods for data collection and analyses, modern genomic technology promises continued improvement in diagnostic and therapeutic options for lung cancer patients.

EARLY DIAGNOSIS/SCREENING OF LUNG CANCER

Smoking, Lung Cancer, and Genetics

In the late 1920s, it was first suspected that cigarette smoking caused lung cancer, and in the 1950s and 1960s a variety of studies linking lung cancer to cigarette smoking were published. In the 1960s, over 40% of men in the United States smoked until the Surgeon General's report linked smoking to lung cancer. Since then, smoking rates in men have decreased to 20–25% with a subsequent leveling off and then fall in lung cancer incidence in the 1990s. Smoking rates in women, however, began to rise in the 1950s and lung cancer became the leading cause of death from cancer in women in the 1980s (reviewed by Spiro and Silvestri, 2005). By 2030, it is projected that smoking incidence will plateau at 20% of the adult US population in both men and women, ensuring a constant high rate of lung cancer throughout the first half of the 21st century.

There is controversy about whether lung cancer is now occurring more frequently in never-smokers. Environmental tobacco exposure or second-hand smoke may cause lung cancer in life-long non-smokers. There have been reports of lung cancer in large numbers of non-smoking Chinese women who have been exposed to fumes from cooking with charcoal in poorly ventilated dwellings (Luo et al., 1996; Zhou et al., 2000). These observations raise the question of whether lung cancer in other never-smokers might result from a variety of unrecognized toxic exposures especially in genetically susceptible individuals.

We now know that although 80–85% of patients with lung cancer have a history of smoking, only 10–15% of patients who smoke actually develop lung cancer suggesting that hereditary factors may explain susceptibility to the carcinogenic effects of smoking. It is well established that smokers who are first-degree relatives in families with a history of lung cancer have a two- to threefold increased risk of developing lung cancer (Thun et al., 2002). The risk is further increased with a family history of early onset lung cancer. Hereditary occurrence is also suggested since both smokers and non-smokers with a positive family history are at higher risk for developing lung cancer.

Epidemiological studies show that Native American and African American smokers are more susceptible to lung cancer than whites, while Latinos, and Japanese Americans are less susceptible (Haiman et al., 2006). These differences are accentuated in those who have smoked fewer than 10 cigarettes per day and tend to disappear in those who have smoked more than 30 cigarettes per day, suggesting genetically determined carcinogenic susceptibility that is masked at high levels of smoke exposure (Haiman et al., 2006). The reasons for the observed racial/ethnic differences in response to the carcinogenic effects of cigarette smoke are not yet known, but it is likely that both genetic and environmental differences are involved. African Americans, for example, tend to smoke menthol cigarettes or ones with higher levels of nicotine and tar (Okuyemi et al., 2004).

There have been a large number of studies that have attempted to define heritable causes of lung cancer susceptibility. Polymorphic variants in almost 50 genes have been claimed to be associated with either a reduced or elevated risk of lung cancer (reviewed by Cooper, 2005). Early studies focused on polymorphisms in genes involved in the metabolism of the toxic components of cigarette smoke. Most of the compounds in cigarettes are activated by phase I drug metabolizing enzymes (DME), such as cytochrome p450, to become active and carcinogenic. These phase I genes code for substances that convert smoke constituents to highly reactive intermediates that can bind to and mutate DNA. Phase II DMEs, such as *N*-acetyltransferases, sulfotransferases, and glutathianone *S*-transferases, detoxify carcinogenic products by conjugating these reactive species to less reactive, excretable products. While some studies support the importance of polymorphisms in both phase I and II genes, others have found no relation between polymorphisms in these genes and the incidence of lung cancer

(Perera et al., 2006; Raunio et al., 1999; Taioli et al., 2003; Wikman et al., 2001).

It is likely that there are a number of genes, some involved in interacting pathways that account for individual susceptibility to the carcinogenic effects of smoking. Recent studies have described polymorphisms in cell cycle checkpoint genes in African American smokers with lung cancer and in DNA repair genes in smokers both with and without lung cancer (David-Beabes and London, 2001; Wenzlaff et al., 2005). A promoter polymorphism that increases promoter activity in caspase-9, a cysteine protease involved in the caspase-mediated apoptotic pathway, was recently found to contribute to genetic susceptibility in lung cancer (Park et al., 2006). In addition, a number of studies have demonstrated linkage to lung cancer on chromosome 6q23 (Bailey-Wilson et al., 2004) and on chromosome 12q (Sy et al., 2004). Mouse modeling studies have identified similar linkage between lung cancer and chronic obstructive pulmonary disease (COPD), and many of the genes in the area of linkage are involved in inflammation, a process that is likely central in the pathogenesis of both diseases (Bauer et al., 2004).

Identifying Smokers at Risk

The risk for developing lung cancer increases with accumulated exposure to cigarette smoke, most often expressed as pack-years (calculated by multiplying the number of packs of cigarettes smoked per day by the number of years the person has smoked). As noted earlier, even in a high-risk population of smokers, however, the incidence of lung cancer is only ~15% over a lifetime. The incidence increases slightly in those over the age of 60 with greater than 30 pack-years, previous lung cancer, and/or atypia of bronchial epithelial cells in the sputum.

Because of the lack of effective diagnostic biomarkers that identify which of the current and former smokers are at the greatest risk for developing cancer, lung cancer is most often diagnosed at a late stage after it has spread. In contrast to most cancers, 5-year survival rates for lung cancer, ~15%, have not changed appreciably over the past 4–5 decades. Previous screening trials with frequent chest x-rays and sputum cytology have not demonstrated an effect on lung cancer mortality (reviewed by Jett and Midthun, 2004). Spiral computerized tomography (CT) scan screening can detect lung tumors at an early stage. The I-ELCAP study, a systematic case–control observational study that included more than 31,000 subjects who were at risk for lung cancer, found that screening by CT resulted in a diagnosis of lung cancer in 484 participants, of which 412 (85%) had clinical stage I lung cancer (Henschke et al., 2006). Spiral CTs, however, while highly sensitive, can be non-specific and many newly detected small lesions have proven on resection to be non-malignant scar tissue or old granulomas rather than early lung cancers (Jett, 2005). Thus, it is not yet known if this approach will alter lung cancer mortality or justify the relatively high cost of large-scale screening. Current trials, including the National Lung Cancer Screening Trial (NLST), will help answer these questions.

Developing biomarkers that are highly sensitive, specific, and target individuals with early stages of cancer is clearly necessary to improve lung cancer mortality. Several groups have used genomic profiling to identify the subset of smokers who are at higher risk for developing lung cancer. Powell et al. (2003) compared the gene expression profiles of tumor and matched normal tissues from smokers and non-smokers. Although hierarchical clustering did not separate tumors from smokers versus non-smokers, it did separate tumor and non-tumor tissue. Four times more genes were altered between tumor and lung in non-smokers as compared with smokers. Their findings demonstrate that underlying normal tissues from smokers and non-smokers differ, consistent with the concept of "field cancerization" in smokers where genes are altered in the entire respiratory epithelium.

To begin to understand the mechanisms by which some individuals protect themselves from the carcinogenic effects of smoking, Spira et al. (2004) used high-throughput genomic and bioinformatic tools to define the genome-wide impact of smoking and smoking cessation on bronchial airway epithelium. They demonstrated changes in both antioxidant and drug metabolizing genes in airway epithelial cells as well as increases in putative oncogenes and decreases in tumor suppressor genes. They also noted that the expression level of smoking-induced genes among former smokers began to resemble that of never smokers after 2 years of smoking cessation. Genes that reverted to normal within 2 years of cessation tended to serve metabolizing and antioxidant functions. In addition the authors found that several genes, including putative oncogenes and tumor suppressor genes, failed to revert to never-smoker levels years after smoking cessation, perhaps explaining the continued risk for developing lung cancer many years after individuals have stopped smoking (Spira et al., 2004). This same group has reported a lung cancer-specific diagnostic gene expression profile in histologically normal airway epithelial cells of patients being evaluated for the diagnosis of lung cancer (Spira et al., 2007). These genes also predicted lung cancer in previously published studies of normal and lung cancer tissue, raising the possibility that changes in airway gene expression of smokers might not only serve as biomarker for risk of developing lung cancer, but may reflect to some extend the genomic changes that occur in the actual lung cancers. These studies support the possibility that measuring gene expression in easily accessible airway epithelial cells may provide a biomarker that identifies smokers at high risk of developing lung cancer.

Detection of mutations or aberrant methylation in sputum or serum is a promising approach to the early diagnosis of lung cancer. Detection of promoter hypermethylation in the CDKN2A, DAPK1, RASSF1, CDH1, GSTP1, RAB, CDH13, APC, MLH1, MSH2, and MGMT genes have been demonstrated using sputum or bronchial lavage (reviewed by Cooper, 2005). Aberrant promoter hypermethylation has been shown to occur frequently in patients with resected lung cancers and concordance between methylation patterns in tumor specimens and bronchial epithelial cells has been demonstrated. In addition, recent studies demonstrated that gene promoter hypermethylation in sputum could identify people at high risk for lung cancer incidence. One study demonstrated FHIT methylation in preneoplastic lesions from smoking-damaged bronchial epithelium

(Zochbauer-Muller et al., 2001). Another study demonstrated hypermethylation of the CDKN2A gene promoter prior to clinical evidence of lung carcinoma (Kersting et al., 2000). Lastly, Palmisano et al. (2000) showed that p16 and/or MGMT could be detected in DNA from sputum up to 3 years before clinical diagnosis of squamous cell cancer.

Reports documenting the clinical potential of detecting tumor-related genes, mutations, loss of heterozygosity, polymorphisms, and promoter hypermethylation of circulating tumor DNA in serum are also surfacing. In addition high-throughput methods for analyzing the methylation status of hundreds of genes simultaneously are being applied to the discovery of methylation signatures that distinguish normal from cancer tissue samples (Bibikova et al., 2006; Wilson et al., 2006). However, none of these studies has tested the diagnostic potential in prospective multicenter trials.

Future Directions

Lung cancer is one of the few cancers for which screening is not recommended, even in high-risk individuals. Although spiral CT and autofluorescent bronchoscopy may increase the detection rate of early lesions, these procedures are costly and/or relatively invasive, have not been shown to be specific for the presence of lung cancer, and it is not yet known if screening will alter mortality. Since molecular, genetic, and epigenetic abnormalities precede morphological changes in bronchi and alveoli, biomarkers may help select a group of high-risk patients that would benefit from spiral CT and/or fluorescent bronchoscopy.

To develop screening profiles that could potentially predict those at risk for developing cancer and/or detect lung cancer at an earlier stage, profiles must be capable of identifying the few abnormal cells among many normal cells and samples ideally should be obtained in a relatively non-invasive fashion. The availability of a number of agents that might be effective in reversing the pre-malignant changes in airway epithelial cells (see below) has driven the search for biomarkers that identify individuals at highest risk for developing lung cancer. Genetic abnormalities can be detected from bronchial biopsies, respiratory cells from sputum, and circulating DNA, and gene expression profiles generated from these specimens offer a wide area of investigation for biomarker development. To be applied in a screening program, these biomarkers must be specific and cost effective with a high efficiency: the focus has been on intermediate markers that indicate risk but might be reversible with appropriate treatment.

A recent review has identified the benefits of and requirements for the ideal biomarker (Dalton and Friend, 2006). Benefits include (1) predict who will develop lung cancer and/or detect cancer at an early treatable stage, (2) serve as guides for treatment decisions (3) and serve to identify new targets for drug development. The authors in this review point out the need to develop biomarkers that deal with the molecular diversity of lung cancer that monitor disease progression and that predict and/or monitor response to preventative therapy. A biomarker that meets even the majority of these requirements remains to be discovered but will almost certainly evolve with modern genomic technology.

A relatively new field, "genetical genomics," which attempts to link gene expression in affected tissues with genetic polymorphisms promises to provide insights into heritable factors in lung cancer (Li and Burmeister, 2005). This field is derived from recent technological and computational advances that allow one to measure levels of gene expression in the affected tissue and relate them to high-density SNP discovery in genomic DNA (Pastinen et al., 2006). Patterns of gene expression in smoke-exposed tissue and linkage to SNPs in the affected genes are likely to lead to new insights into heritable cause of lung cancer.

Chemoprevention of high-risk current or former smokers represents one of the most important areas of current research in the prevention of lung cancer. It has been a topic of great interest for over 20 years. Despite numerous trials beginning with high doses of retinoids in the 1980s, there have been no studies to date that have demonstrated a positive outcome in terms of mortality from lung cancer and few studies that have demonstrated an effect of intermediate end points, such as sputum atypia or genomic markers of epithelial cell damage. Several recent publications review the results of previous randomized trials and the rationale for use of new chemopreventative agents (Hirsch and Lippman, 2005; Kelloff et al., 2006; Khuri and Cohen, 2004).

Chemoprevention (chemoprophylaxis) has assumed increasing importance as former smokers now account for 50% of new lung cancer cases in the United States. Since there are approximately 45 million former smokers in the United States, identifying which former smokers are at highest risk for developing lung cancer in the future is the most important first step in any chemoprevention program. Many new chemopreventative agents are being explored using agents that fall into several broad categories: (1) anti-inflammatory/antioxidants, (2) epigenetic modulators of methylation and/or acetylation, and (3) modulators of signal transduction, particularly those that affect the EGFR-signaling pathways. A recent clinical study conducted in 10 patients to assess the potential chemopreventive effect of myo-inositol in smokers with bronchial dysplasia showed a significant increase in the rate of regression of preexisting dysplastic lesions (Lam et al., 2006). The potential impact of chemoprevention is large, but the field awaits the emergence of intermediate markers of cancer risk that must be validated in prospective studies.

CLASSIFICATION AND PROGNOSIS

Histological Classification of Lung Tumors and TNM Staging

Lung cancers are classified as small-cell lung carcinomas (SCLC) or non–small-cell carcinomas (NSCLC). Small-cell lung cancers have neuroendocrine features that are identified by immunohistochemistry and histology. NSCLC are subcategorized as adenocarcinomas (most common), squamous cell carcinomas, and large-cell carcinomas, and are clinically distinct from SCLC. The pathological distinction between SCLC and NSCLC is very important since these tumor types are treated differently (reviewed by Spira and Ettinger, 2004).

Pre-malignant lesions termed low and high-grade dysplasia, carcinoma *in situ* (CIS), and atypical adenomatous hyperplasia (AAH) are associated with increased risk for developing lung cancer (reviewed by Kerr, 2001). When the airways are exposed to carcinogens, cellular and molecular changes occur over broad mucosal surfaces, a concept termed "field cancerization." Accumulation of mutations and epigenic alterations ensue, eventually leading to invasive lung cancer, a process called "multistep carcinogenesis" (reviewed by Cooper, 2005).

Histological analysis is currently used to identify and classify cancers, yet definitive diagnoses are often difficult to make. Furthermore, tumors with the same histological classification behave differently, and morphological classification has not been effective in predicting the aggressiveness of a cancer or how the cancer will respond to therapeutic agents.

In NSLC, tumor staging (*International System for Staging Lung Cancer*) is the most important prognostic variable for predicting survival. TNM staging takes into account the degree of spread of the primary tumor (T), the extent of regional lymph node involvement (N), and the presence or absence of metastases (M) (summarized by Hoffman et al., 2000). Recently adopted revisions to TNM staging include (1) the splitting of stage I into IA and IB; (2) the splitting of stage II into IIA and IIB; (3) reclassifying T3N0M0 from Stage IIIA to IIB; and (4) classifying multiple pulmonary tumor nodules as T4 if the satellite nodule(s) are in the same lobe, or M1 if the ipsilateral nodules(s) are in the non-primary lobe. Despite these changes and improvements in diagnosis including both non-invasive methods (CT/PET), and invasive methods (EUS, EBUS), patients diagnosed with similar stage and treated using similar protocols, often respond quite differently and have varying survival rates. Some tumors, though found at an early stage, will rapidly progress to metastatic disease. Furthermore, in patients who have undergone surgical resection, recurrence rates are high with 5-year survival rates only 40% for stage IIB NSCLC.

Although several combined clinical, histological, and laboratory variables such as age, stage or grade of tumor, and serum protein levels can be used to assess a patient's prognosis with variable accuracy, these criteria are not able to provide important information about the prognostic diversity within each stage such as how aggressive a particular subtype will be, or how a patient will respond to therapy. By combining clinical variables and histopathology with gene expression profiling, predicting a patient's prognosis could theoretically be improved.

Molecular Classification and Prognostic Value of Lung Cancer Genomics

Morphological tumor classification does not always accurately predict the patient's clinical behavior since lung cancer is genetically heterogeneous. For example, patients with stage IA lung cancer resected for cure still have 30% mortality from local recurrence, distant metastases, and/or new occurrence. Within each clinical stage, there is variability in the presence of specific mutations, deletions of tumor suppressor genes, amplifications of oncogenes, and chromosomal abnormalities. Genomics is a powerful tool for classifying tumor subtypes. Lung cancer patients with biomarkers that predict a poor outcome could be selected for adjuvant chemotherapy while those that predict a good prognosis may be able to avoid the toxicity and cost of unnecessary chemotherapy.

Recently, microarray studies have identified and validated specific genes whose expression differs between normal and tumor tissue. Using cDNA arrays, Wikman et al. (2002) compared gene expression profiles of 14 pulmonary adenocarcinoma patients with normal lung tissue. These authors demonstrated marked differences in gene expression level between normal lung and adenocarcinomas. Another study by Yamagata et al. (2003) showed that a profile based on gene expression could be used on blinded samples to differentiate primary NSCLC from normal lung and lung metastases. Furthermore, groups of genes were identified that were able to identify known histological subgroups of NSCLCs.

Genomic high-throughput technologies have been used in many studies to identify gene expression signatures that predict patient survival and/or relapse rates. A considerable number of expression profiling studies have been performed on clinical lung cancer specimens (summarized by Granville and Dennis, 2005; Kopper and Timar, 2005; and Cooper, 2005). Not only can gene expression profiles group tumor samples consistent with classical histology, but can also identify subgroups within histologic subclasses. In an attempt to discover previously unrecognized subtypes of lung cancer, two microarray studies aimed at class discovery by hierarchical clustering using primary lung cancer specimens were published in 2001 (see Figure 34.1). Garber et al. (2001) identified gene subsets that are characteristic of each of the known morphological subtype in NSCLSs. In addition, these authors found that these tumors could be further divided into subgroups with significant differences in patient survival. Another study by Bhattacharjee et al., (2001) showed by gene expression analysis of 186 lung carcinomas that biologically distinct subclasses of lung carcinomas exist. In their study, a subclass of adenocarcinomas defined by having neuroendocrine gene expression had a less favorable outcome, whereas a subset of patients with predominantly type II pneumocyte expression, had a more favorable outcome.

In a study designed to identify whether gene expression patterns could predict survival, Beer et al. (2002) demonstrated that expression profiles based on microarray analysis could be used to predict disease progression and clinical outcome in early-stage lung adenocarcinomas. Subdivision of the lung tumors based on gene expression patterns matched the morphological classification of tumors into squamous, large-cell, adenocarcinoma, and small-cell lung cancer. Furthermore, adenocarcinomas could be further subclassified based on their gene expression profiles that correlated with the degree of tumor differentiation and patient survival. They demonstrated by gene profiling that a list of 50 genes were most effective at dividing patients with stage I lung adenocarcinoma into high- and low-risk groups for mortality, thus potentially identifying a subset of stage I NSCLC cancer patients who would benefit from adjuvant therapy. This group has recently extended their observations to the squamous cell

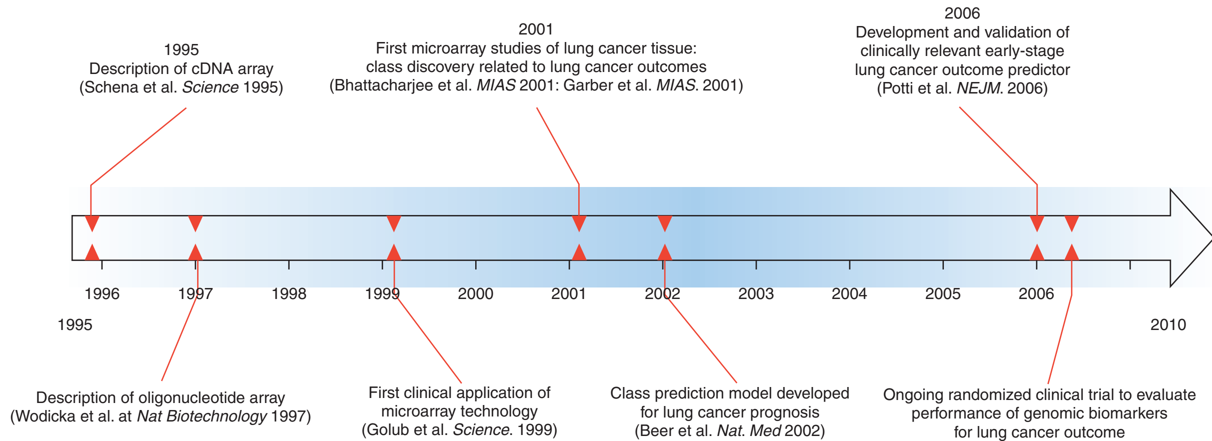


Figure 34.1 Timeline for the development of DNA microarray technology and its application to lung cancer. The key advances in the field of microarray technology and their clinical applications to lung cancer tissue for prognostic purposes are highlighted.

carcinoma subset of NSCLC and have generated a similar risk of progression stratifier which combined with the adenocarcinoma tool, will cover ~80% of all NSCLC (Raponi et al., 2006).

Recently, a clinically relevant prognostic tool that may alter clinical decision-making in early-stage lung cancer was developed using gene expression profiling (Potti et al., 2006a, b). Potti et al. (2006a, b) identified a gene expression profile that predicted the risk of recurrence in 89 patients with early-stage lung cancer (Figure 34.2). The authors then validated the model, termed the "lung metagene" model, in a group of 134 patients. The test predicted recurrence for individual patients significantly better than did clinical prognostic factors. This predictor also identified a subgroup of patients with stage IA lung cancer who were at high risk of recurrence and would potentially benefit from adjuvant therapy. Based on these findings, a multicenter clinical trial has been initiated. This upcoming clinical trial is the first to use a genomic test to select treatment options for individual lung cancer patients.

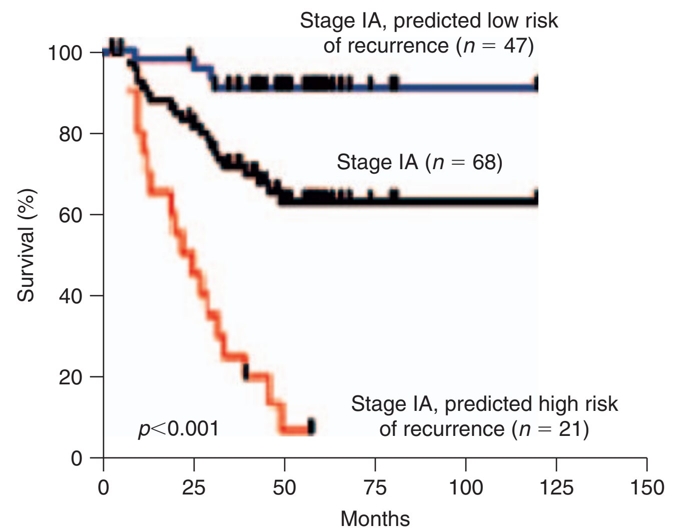


Figure 34.2 Application of the gene expression model to determine prognosis in stage IA NSCLC. Kaplan-Meier survival estimates from three cohorts of patients with stage IA disease. The middle black Kaplan-Meier curve represents the median survival estimate in all cohorts. The top and bottom Kaplan-Meier curves represent survival estimates for subjects predicted to have low and high risk of recurrence respectively by the gene expression model, demonstrating the ability of the genomic model to accurately predict survival in stage IA disease. Figure reproduced from Potti et al. *NEJM*, 2006.

Future Directions

Currently, lung cancer patients within a given clinical stage and tumor type receive the same treatment despite the genomic heterogeneity that exists between patients. Although expression profiling of lung cancer has identified prognostic genes and identified subtypes of adenocarcinomas, until recently, these profiles have not been ready to be incorporated into clinical practice. Comparison of array studies has been difficult because the array platforms, sample preparation, and technical factors have been different. In addition, many molecular classification studies do not match the classification based on tumor histology, perhaps because use of whole lungs for microarray studies may not accurately reflect gene expression in the cancer cells because of contamination and differences in abundance of stromal and surrounding normal cells. Although laser-captured microdissection (LCM) can be used to obtain a homogeneous population of epithelial cancer cells, excluding surrounding stroma may eliminate cells that contribute to the tumor environment.

Redefining tumor classification from strictly morphology-based schemes to molecular-based classifications using a variety of parameters including histological patterns, gene expression profiles including the presence or absence of oncogenes, tumor

suppressor genes, and miRNAs, and cell surface protein and receptor status, promises to provide clinically important information on tumor subsets within morphological classes. A recent expression profile analysis of lung adenocarcinomas, for example, classified tumors into two major subtypes, terminal respiratory unit (TRU) and non-TRU subtypes (Takeuchi et al., 2006). TRU type tumors retained features of normal peripheral lung features and had higher frequency of EGFR mutations. The presence of EGFR mutations in this adenocarcinoma subtype predicted a shorter post-operative survival.

A large study is underway in which participating centers are pooling and analyzing lung adenocarcinoma specimens, the results of which will hopefully lead to a gene expression signature that will differentiate the adenocarcinoma subtypes (Dobbin et al., 2005). Prognostic profiles and eventually, a prognostic-specific gene chip, may help guide clinical decision-making by identifying "high risk" lung cancer patients that would benefit from improved diagnostic and treatment options.

PATHOGENESIS AND TREATMENT OF LUNG CANCER

Molecular Alterations in Lung Cancer

Proto-oncogenes

Proto-oncogenes such as *Ras* and *Myc* were first discovered in the 1970s and 1980s, and it was only in the 1980s and 1990s that these genes were integrated into signaling pathways and that mutations in these genes were linked to human cancers. The *Ras* genes (*Hras*, *Kras*, and *Nras*) encode GTPase proteins that help transduce survival and growth-promoting signals. When oncogenic mutations occur, the normal abrogation of RAS signaling by hydrolysis of bound GTP to GDP is impaired, resulting in persistent signaling (reviewed by Singhal et al., 2005). Point mutations (found most frequently in codon 12, 13, and 61) are detected in 20–30% of lung adenocarcinomas (Slebos et al., 1990) and 90% of these mutations are found in Kras (see Table 34.1). Mutations in Kras are markers for poor prognosis in NSCLCs (Graziano et al., 1999) as mutations rarely occur in SCLC. *Myc* genes (*MYCL*, *MYCN*, and *CMYC*), which encode transcription factors that regulate genes involved in cell cycle regulation, DNA synthesis, and RNA metabolism, become activated by loss of transcriptional control or by gene amplification, resulting in MYC protein overexpression. C-Myc amplification occurs in 5–10% of NSCLCs (Richardson and Johnson, 1993).

Growth Factors and Their Receptors

Lung tumors often express growth factors and their receptors, and the resulting regulatory loops can stimulate tumor growth. EGFR, also known as ERBB-1, is highly expressed in many epithelial tumors, including a subset of lung adenocarcinomas (Reissman et al., 1999). When ligand (EGF or TGF-α) binds to EGFR, there is receptor tyrosine kinase activation and a series of downstream signaling events, including the mitogen-activated protein kinase (MAPK), PI3/Akt, and Jak/Stat pathways (see Figure 34.3) that can result in cellular proliferation, increased cell motility, tumor invasion, anti-apoptosis, and resistance to chemotherapy summarized by Baselga, 2006). EGFR tyrosine kinase, therefore, was proposed as a target for cancer therapy 20 years ago.

TABLE 34.1 Common molecular alterations in lung cancer

	Type of mutation	Frequency
Proto-oncogenes		
k-Ras	Point mutation	NSCLC: 20–30% (mostly adeno)
MYC	Translocation amplification	SCLC: 30–40% NSCLC: 5–10%
Bcl-2	Translocation	NSCLC: (30%) Squamous: 25%; Adeno: 10%
Growth factors		
EGFR (ERBB-1)	Deletion/Mutation amplification	SCLC: 0% NSCLC: 10% (BAC: 25%)
HER2 (erbB2)	Translocation amplification	NSCLC 30–40%
Tumor suppressors		
p53	Deletion/LOH	SCLC: 75% NSCLC: 50%
Rb gene	Point mutation	SCLC: 90% NSCLC: 15–30%
P16 (CDKN2)	Deletion/LOH	SCLC: >80% NSCLC: 30–50%
FHIT	Deletion/LOH	SCLC: 100% NSCLC: 60%

Anti-EGFR drugs approved for cancer treatment include the monoclonal antibodies directed against the extracellular domain of the receptor (anti-EGFR Mabs) and small-molecule inhibitors of EGFR's tyrosine kinase activity (TKIs). In NSCLCs, two trials have demonstrated that in unselected patient populations, anti-EGFR TKIs have modest levels of antitumor activity. In 10% of patients, however, tumor response is often dramatic, but responses tend to be transient. Responsive patients are most often non-smoking Asian woman with adenocarcinomas, often bronchioloalveolar carcinoma (BAC) (reviewed by Baselga, 2006; Dziadziuszko et al., 2006). Subsequent analyses demonstrated that lung cancers with somatic mutations in EGFR correlate with a positive clinical response to EGFR inhibitors (reviewed by Janne et al., 2005; Lynch et al., 2006; and Thomas et al., 2006). Unfortunately, these initial responders often develop resistance to these drugs. A second EGFR somatic mutation, T790M, occurs in some cases of NSCLC that recur after an initial response to TKIs (Kobayashi

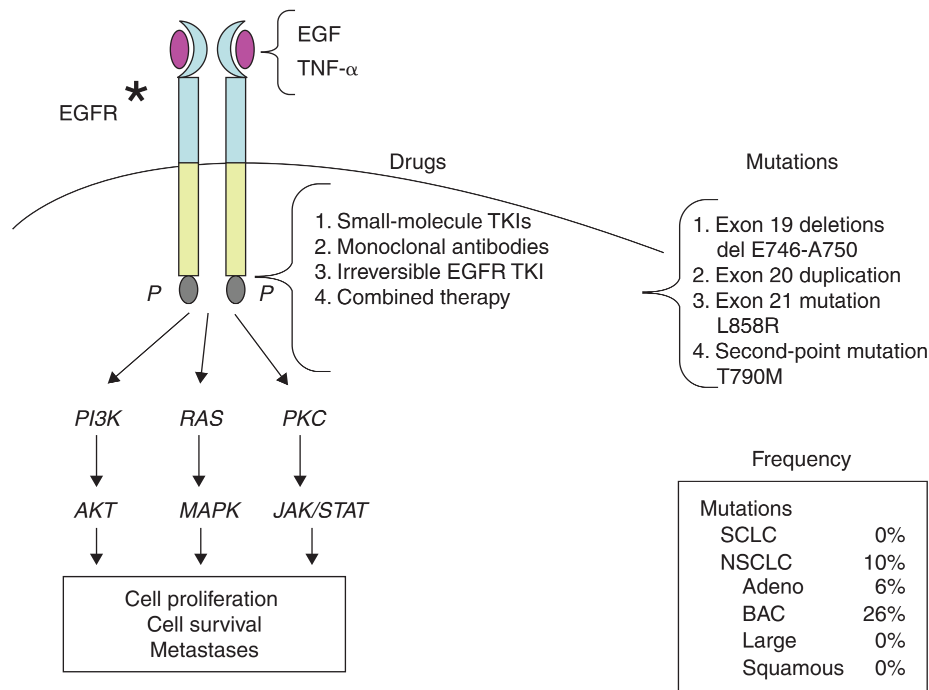


Figure 34.3 When ligand (EGF or TGF-α) binds to EGFR, there is receptor tyrosine kinase activation and a series of downstream signaling events, including the MAPK, PI3/Akt, and Jak/Stat pathways, that can result in cellular proliferation, increased cell motility, tumor invasion, anti-apoptosis, and resistance to chemotherapy. Receptor expression can be increased by gene duplication (*). Most of the tyrosine kinase domain and all activating EGFR mutations thus far are in exons 18–21. Deletions in exon 19 and substitution mutations in exon 21 account for the majority of mutations. Mutations occur most frequently in bronchoalveolar carcinoma (BAC), but not in squamous cell carcinoma or SCLC (Marchetti et al., 2005; Shigematsu et al., 2005). Anti-EGFR drugs approved for cancer treatment include the monoclonal antibodies directed against the extracellular domain of the receptor (anti-EGFR Mabs) and small-molecule inhibitors of EGFR's tyrosine kinase activity (TKIs). Examples of anti-EGFR Mabs include Cetuximab (Erbitux), Panitumumab, and Pertuzumab. Examples of TKIs include Gefitinib (Iressa) and Erlotinib (Tarceva).

et al., 2005). Germ line T790M mutations have been described in a family with multiple cases of lung cancer (Bell et al., 2005).

Clinical trials using alternate EGFR inhibitors, such as irreversible EGFR inhibitors, are underway in NSCLC patients. Some centers have begun to sequence EGFR in all resected tumors in order to tailor drug therapy to specific mutations, although further research is needed to more completely define the appropriate population for EGFR testing (reviewed by Sequist et al., 2006). The complexity of EGFR mutations is highlighted by a recent report identifying different EGFR mutations in the primary tumor versus lesions metastatic from the primary site (Italiano et al., 2006). Furthermore, Engelman et al. (2006) recently reported that allelic dilution of a biologically significant resistance mutation in EGFR-amplified lung cancer may be undetected by direct sequencing.

Tumor Suppressor Genes

The role of p53, a tumor suppressor gene, is to help maintain genomic integrity after DNA damage. When cells undergo stress, such as from carcinogen exposure, UV radiation, and/or hypoxia, p53 becomes upregulated and then acts as a transcription factor to increase genes such as p21 (which in turn controls G1/2 cell cycle transition) and induces apoptosis by activating genes such as BAX, PERP, and others (reviewed by Fong et al., 2003; Singhal et al., 2005). In lung cancer, missense mutations can occur and cause loss of p53 function in >75% of SCLCs and 50% of NSCLCs (Toyooka et al., 2003).

Decreased expression of p16 by promoter hypermethylation, mutations, or allelic loss occurs in 30–50% of NSCLCs. The *p16*^{INK4A}*-cyclin D1-CDK4-RB* pathway controls G1/S cell cycle transition, and loss of p16 releases the tumor cells from RB-mediated cell cycle arrest (reviewed by Fong et al., 2003). Alternatively, the Rb gene can be inactivated directly by deletions, point mutations, or alternative splicing, and are more commonly found in SCLCs (>90%), than NSCLCs (15–30%) (see Table 34.1).

Oncogenic Pathway Signatures

The malignant process is complex, involving multiple mutations leading to deregulation of signaling pathways (see above). In addition, the role of miRNA, very small non-coding RNA products that play a key role in regulatory networks by regulating the translation and regulation of mRNAs, are emerging

(reviewed by Calin and Croce, 2006). Recent studies have demonstrated altered miRNA gene expression in lung cancers. Dicer, a component of the miRNA machinery has been shown to be downregulated in NSCLCs (Karube et al., 2005). Reduced expression of *let*-7 miRNA family members have been found to correlate with shorter post-operative survival in potentially curative lung cancer resection (Takamizawa et al., 2004).

The feasibility and utility of monitoring miRNAs has been recently demonstrated in a study by Lu et al. (2005). These authors demonstrated, using a bead-based miRNA detection method, that a relatively small number of miRNA genes (~200 miRNAs) may be sufficient to classify human tumors. A study by Yanaihara et al. (2006) showed that lung cancer specimens have extensive alterations of miRNA expression that may dysregulate cancer-related genes. Furthermore, miRNA microarray analysis using a microchip for genome-wide miRNA profiling (developed by Liu et al., 2004) identified a signature that could discriminate lung cancer from non-cancer tissue, and could also correlate lung cancer specimens with patient survival. In their study, high *miR-155* and low *let-7a-2* correlated with poor survival in lung adenocarcinomas (Yanaihara et al., 2006).

let-7 miRNA family members have been shown to directly regulate Ras genes (Johnson et al., 2005). These combined studies suggest that *let*-7 may be a promising therapeutic agent to treat lung cancers caused by activating mutations in Ras genes (reviewed by Esquela-Kerscher and Slack, 2006).

DNA microarray–based gene expression signatures have been developed that have the ability to define oncogenic pathways. The ability to predict the deregulation of various oncogenic pathways (both the mutated gene product itself and its downstream targets) offers an opportunity for new therapeutic drugs that are pathway-specific. Sweet-Cordero et al. (2005) recently developed a KRAS-mutant signature from analysis of expression array data using the KRAS-mutant mouse model of lung adenocarcinoma, which was then transferred to a human KRAS-mutant signature. The signature accurately identified KRAS-mutant lung tumors. These results were confirmed using siRNA and real-time RT-PCR.

Another study by Bild et al. (2006a, b) demonstrated that DNA microarray–based gene expression signatures can not only predict cells expressing oncogenic activity from control cells, but can also predict deregulation of various oncogenic pathways in specific tumor types derived from mouse cancer models and human cancer specimens. These authors used multiple experimental models including functional cell-based assays, mouse lung cancer models, and human cancer specimens to demonstrate that the Ras pathway status clearly correlates with lung adenocarcinoma relative to squamous cell carcinomas subtypes. Furthermore, independent of tumor histology, patients displaying deregulation of multiple pathways (Ras with Src, Myc, β-catenin) had a poor survival.

Future Directions

Molecularly targeted therapy to reduce lung cancer mortality biomarkers provide an opportunity to identify subpopulations of patients who are most likely to respond to a given therapy and identify new targets for drug development. A study by Olaussen et al. (2006) for example, showed that in lung tumor specimens, the absence of ERCC1 (an enzyme that participates in the repair of DNA damage caused by cisplatin) was associated with a survival benefit from cisplatin-based adjuvant chemotherapy, whereas patients whose tumor expressed the enzyme failed to benefit from chemotherapy.

Several clinical studies have now shown the therapeutic efficacy and safety of tyrosine kinase inhibitors for specific tumor subtypes. Many of the 90 proteins with tyrosine kinase domains encoded by the human genome are aberrantly activated in human tumors. Intriguingly, after treatment with tyrosine kinase inhibitors in some tumor subtypes, there is marked reduction not only in tumor growth, but also in tumor viability. This concept of dependence of the tumor mutated oncogenes and/or on EGFR-mediated survival signals is referred to as "oncogene dependence or oncogene addiction" (Varmus, 2006; Weinstein, 2002). Oncogene dependence is also demonstrated in mouse models overexpressing constitutively active K-ras, where the mutant K-ras is then silenced. In this model, lung tumors rapidly regress as a result of apoptosis when K-ras is silenced even in the absence of important tumor suppressor genes (Fisher et al., 2001). An alternative hypothesis to explain oncogene addiction is through differential signal attenuation of multiple pro-apoptotic and pro-survival signals, a term coined "oncogenic shock" (Sharma et al., 2006).

The improved survival in NSCLC patients taking EGFR TKIs seems not to only be limited to patients with EGFR mutations. Markers such as EGFR gene amplification, ErbB3 levels, and Her2 mutations may also be predictors of responsiveness (reviewed by Engelman and Cantley, 2006). Ongoing studies to investigate EGFR TKIs as first line therapy in selected patients with mutations in exons 18–21, increased EGFR copy number or protein expression, and with clinical characteristics associated with response are needed (reviewed by Johnson, 2006). Although the majority of lung tumor cells acquire resistance mutations during therapy with EGFR TKIs, these cells seem to be to be sensitive to a new group of TKIs that covalently cross-link the receptor (Carter et al., 2005). In addition, a recent study showed that microarray gene expression profiling demonstrated a pattern of gene expression associated with sensitivity to EGFR (Coldren et al., 2006). Mouse lung models with inducible expression of mutations in EGFR have been generated and may help in further understanding the pathogenesis of human lung cancer and aid in the validation of cancer therapeutics (reviewed by Dutt and Wong, 2006).

Given the complexity of NSCLC, it is likely that these tumors are dependent on more than one oncogenic signaling pathway. Alterations in cancer-specific copy number and loss of heterozygosity (LOH) are important changes found in cancer cells and SNP array analyses may reveal pathways disrupted in tumorigenesis. SNP arrays have been used in genomic studies for detecting LOH in lung cancer and have both detected previously unknown regions of copy number change as well as

known regions of both amplification and homozygous deletion. It is likely therefore that combination targeted therapy directed at multiple oncogenic pathways may not only prove more effective than single agents alone, but may also prevent or delay secondary resistance.

Most tumors express multiple mutant genes, but the relative importance of mutated genes in maintaining the cancer phenotype is not yet known, nor is the role of growth-promoting signals by non-cancerous stromal cells. The success of anti-angiogenic strategies in NSCLC suggests that targeting specific proteases and growth signals supplied by the tissue environment may prove useful as therapeutics.

Establishing tumor banks to study the molecular phenotype of patients who are sensitive and resistant to therapy in order to identify the appropriate patients to treat is therefore crucial. Recently, gene expression signatures from tumor biopsy specimens have been developed that can predict sensitivity to individual chemotherapeutics. Furthermore, these chemotherapy response signatures are being integrated with signatures of oncogenic pathway deregulation to potentially identify new therapeutic strategies (Potti et al., 2006a, b). The considerable potential of using gene expression profiling of tumors to define causal molecular pathways and potential therapeutic targets for individuals is summarized in a recent review from Nevins' group (Bild et al., 2006a).

2009 UPDATE

Early Diagnosis/Screening of Lung Cancer

Developing biomarkers that identify smokers at high risk for developing lung cancer represents an important approach to improving lung cancer mortality. A study using DNA microarrays to profile the gene expression patterns of cytologically normal airway epithelial cells in current and former smokers undergoing bronchoscopy for the clinical suspicion of lung cancer found an 80-probeset lung cancer-specific biomarker that distinguishes smokers with and without lung cancer (Spira et al., 2007). The biomarker was sensitive and specific when tested on an independent test set and on an additional prospectively collected set of samples. In addition, this biomarker provides information about the likelihood of lung cancer that is independent of clinical risk factors for lung cancer among patients with non-diagnostic bronchoscopies (Beane et al., 2008). Since smoking-induced cancer-specific alterations in gene expression may precede the development of lung cancer, these biomarkers may prove useful in identifying high-risk lung cancer patients.

Classification and Prognosis

Accurate molecular classification of the subtypes of NSCLC, for tumor staging and prognosis, may provide a new approach to tumor-specific therapies. Chen et al. reported a 5-gene signature that predicts survival among patients with NSCLC (Chen et al., 2007). Sixteen genes were found to be associated with survival across training and test sets using DNA microarrays measuring 672 previously identified genes associated with invasive activity in invasive NSCLC cell lines (Chen et al., 2001).

A meta-analysis from seven different datasets to predict survival in patients with stage I NSCLC identified genes that were common to the microarray platforms used in all seven datasets (Lu et al., 2006). A total of 197 samples with stage I NSCLC from five of the seven datasets were used to identify a gene expression signature of 64 genes predictive of survival. The signature had higher classification power compared to stage and was accurately predicted survival in the two datasets not used to develop the signature.

Whole-genome gene expression microarrays on frozen tumor samples from 172 NSCLC patients (pT1-2, N0-1, M0) in five European institutions were used to develop a 72-gene expression prognostic NSCLC classifier (Roepman et al., 2009). Based on the classifier score, patients were classified as either high or low risk of disease recurrence. Patients classified as low risk showed a significantly better recurrence-free survival, both in the training set and in the independent validation set. Genes in the prognostic signature were strongly enriched for genes associated with immune response.

Recently in a large retrospective, multi-site, blinded study, Shedden et al. analyzed the performance of several prognostic models in 442 lung adenocarcinoma specimens collected from four different institutions (Shedden et al., 2008). Eight prognostic classifiers and classifiers based on the work of Potti et al. (2006) and Chen et al. (2007) were developed and evaluated on training and test subsets. The results varied, but the inclusion of clinical covariates improved the performance of most classifiers, and a small subset of the classifiers performed well across both tests sets when trained across all stages or stage I samples. The study presents a framework for conducting large studies to test gene expression signatures and illustrates the challenges associated with developing a prognostic gene expression signature for clinical application.

Pathogenesis and Treatment

Chemotherapy-resistance is common in cancer and biomarkers that identify chemotherapeutic-specific lung cancer subtypes may improve prognosis through use of individualized treatments. Hsu et al. (2007) developed predictors of cisplatin (a first line agent) and pemetrexed (a second line agent) sensitivity using the NCI-60 data and data from Gyorffy et al. (2006). The authors found an inverse correlation between *in vitro* cisplatin and pemetrexed sensitivity, and between the likelihood of cisplatin and pemetrexed response in patients.

In a recent study, gene set enrichment analysis was used to associate pathways with chemotherapy resistance in two data

sets: the NCI-60 cancer cell lines deemed sensitive and resistant to specific chemotherapeutic agents and a series of 40 lung cancer cell lines for which sensitivity to cisplatin and docetaxel was determined. Analysis of the lung cancer cell lines identified the bcl-2 pathway to be associated with cisplatin resistance and the AKT pathway enriched in cisplatin- and docetaxel-resistant cell lines. Results from Connectivity Map supported an association between phosphatidylinositol 3-kinase/AKT and docetaxel resistance and targeted inhibition of the phosphatidylinositol 3-kinase/AKT pathway, in combination with docetaxel, resulted in a synergistic effect in previously docetaxel-resistant cell lines. These results support the use of a genomic approach to identify drug-specific targets associated with the development of chemotherapy resistance (Riedel et al., 2008).

Conclusion

The contributions of modern genomic technologies, particularly those that provide measures of global gene expression, are depicted in Figure 34.4. Gene expression profiling provides a new approach to assessing risk of developing lung cancer, potential tools for early diagnosis, new approaches to determining prognosis and oncogenic pathways that will lead to individually targeted therapies. Combined with genome-wide SNP screens that hold the promise of determining heritable predispositions to developing lung cancer and defining pathway-based phamacogenomics, there is promise that lung cancer may no longer be the number one cause of cancer death in the world.

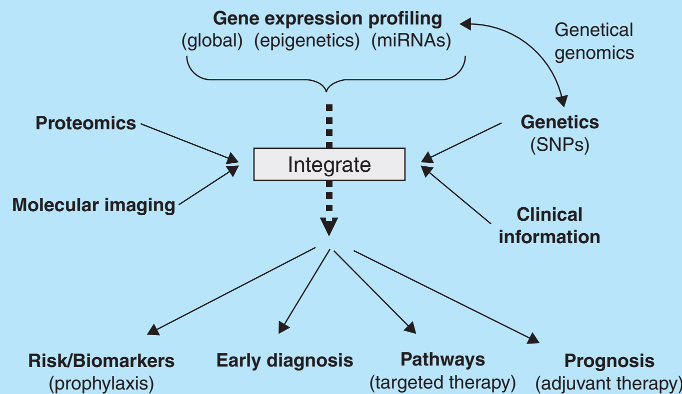


Figure 34.4 Diagram depicting contributions of gene expression profiling, combined with genetics, clinical information, proteomics, and imaging studies that can be applied to developing (a) risk-assessment tools that can be used to identify current and former smokers at highest risk for developing lung cancer and who might benefit from chemopreventative therapy; (b) tools for early diagnosis of current and former smokers with lung stage I-potentially resectable lung cancer; (c) tools that define molecular pathways that have lead to individual lung cancers and that predict what pharmacological approaches might be used to define best approaches to treatment of individual lung cancers; and (d) prognosis of resected cancers defining which patients have a high probability of recurrence or metastases and therefore should receive adjuvant therapy.

2009 UPDATE REFERENCES

Beane, J., Sebastiani, P., Whitfield, T.H. et al. (2008). A prediction model for lung cancer diagnosis that integrates genomic and clinical features. *Cancer Prev Res* 1, 65–76.

Chen, H.Y., Yu, S.L., Chen, C.H. et al. (2007). A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med* 356, 11–20.

Chen, J.J., Peck, K., Hong, T.M. et al. (2001). Global analysis of gene expression in invasion by a lung cancer model. *Cancer Res* 61, 5223–5230.

Gyorffy, B., Surowiak, P., Kiesslich, O. et al. (2006). Gene expression profiling of 30 cancer cell lines predicts resistance towards 11 anticancer drugs at clinically achieved concentrations. *Int J Cancer* 118, 1699–1712.

Hsu, D.S., Balakumaran, B.S., Acharya, C.R. et al. (2007). Pharmacogenomic strategies provide a rational approach to the treatment of cisplatin-resistant patients with advanced cancer. *J Clin Oncol* 25, 4350–4357.

Lu, Y., Lemon, W., Liu, P.Y. et al. (2006). A gene expression signature predicts survival of patients with stage I non small cell lung cancer. *PLoS Med* 3, e467.

Potti, A., Mukherjee, S., Petersen, R. et al. (2006). A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. *N Engl J Med* 355, 570–580.

Riedel, R.F., Porrello, A., Pontzer, E., Chenette, E.J., Hsu, D.S., Balakumaran, B., Potti, A., Nevins, J. and Febbo, P.G. (2008). A genomic approach to identify molecular pathways associated with chemotherapy resistance. *Mol Cancer Ther* 7, 3141–3149.

Roepman, P., Jassem, J., Smit, E.F., Muley, T., Niklinski, J., van de Velde, T., Witteveen, A.T., Rzyman, W., Floore, A., Burgers, S. et al. (2009). An immune response enriched 72-gene prognostic profile for early-stage non-small-cell lung cancer. *Clin Cancer Res* 15, 284–290.

Shedden, K., Taylor, J.M., Enkemann, S.A. et al. (2008). Gene expression-based survival prediction in lung adenocarcinoma: A multi-site, blinded validation study. *Nat Med* 14(8), 822–827.

Spira, A., Beane, J.E., Shah, V. et al. (2007). Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat Med* 13, 361–366.

REFERENCES

Bailey-Wilson, J.E., Amos, C.I., Pinney, S.M., Petersen, G.M., de Andrade, M., Wiest, J.S., Fain, P., Schwartz, A.G., You, M., Franklin, W. et al. (2004). A major lung cancer susceptibility locus maps to chromosome 6q23-25. *Am J Hum Genet* 75(3), 460–474.

Baselga, J. (2006). Targeting tyrosine kinases in cancer: The second wave. *Science* 312(5777), 1175–1178.

Bauer, A.K., Malkinson, A.M. and Kleeberger, S.R. (2004). Susceptibility to neoplastic and non-neoplastic pulmonary diseases in mice: Genetic similarities. *Am J Physiol Lung Cell Mol Physiol* 287(4), L685–L703.

Beer, D.G., Kardia, S.L., Huang, C.C., Giordano, T.J., Levin, A.M., Misek, D.E., Lin, L., Chen, G., Gharib, T.G., Thomas, D.G. et al. (2002). Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* 8(8), 816–824.

Bell, D.W., Gore, I., Okimoto, R.A., Godin-Heymann, N., Sordella, R., Mulloy, R., Sharma, S.V., Brannigan, B.W., Mohapatra, G., Settleman, J. et al. (2005). Inherited susceptibility to lung cancer may be associated with the T790M drug resistance mutation in EGFR. *Nat Genet* 37(12), 1315–1316.

Bhattacharjee, A., Richards, W.G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M. et al. (2001). Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci USA* 98(24), 13790–13795.

Bibikova, M., Lin, Z., Zhou, L., Chudin, E., Garcia, E.W., Wu, B., Doucet, D., Thomas, N.J., Wang, Y., Vollmer, E. et al. (2006). High-throughput DNA methylation profiling using universal bead arrays. *Genome Res* 16(3), 383–393.

Bild, A.H., Potti, A. and Nevins, J.R. (2006a). Linking oncogenic pathways with therapeutic opportunities. *Nat Rev Cancer* 6(9), 735–741.

Bild, A.H., Yao, G., Chang, J.T., Wang, Q., Potti, A., Chasse, D., Joshi, M.B., Harpole, D., Lancaster, J.M., Berchuck, A. et al. (2006b). Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 439(7074), 353–357.

Calin, G.A. and Croce, C.M. (2006). MicroRNA-cancer connection: The beginning of a new tale. *Cancer Res* 66(15), 7390–7394.

Carter, T.A., Wodicka, L.M., Shah, N.P., Velasco, A.M., Fabian, M.A., Treiber, D.K., Milanov, Z.V., Atteridge, C.E., Biggs, W.H., 3rd, Edeen, P.T. et al. (2005). Inhibition of drug-resistant mutants of ABL, KIT, and EGF receptor kinases. *Proc Natl Acad Sci USA* 102(31), 11011–11016.

Coldren, C.D., Helfrich, B.A., Witta, S.E., Sugita, M., Lapadat, R., Zeng, C., Baron, A., Franklin, W.A., Hirsch, F.R., Geraci, M.W. and Bunn, P.A., Jr (2006). Baseline gene expression predicts sensitivity to gefitinib in non-small cell lung cancer cell lines. *Mol Cancer Res* 4(8), 521–528.

Cooper, D.N. (2005). *The Molecular Genetics of Lung Cancer*. Springer, Heidelberg, Germany.

Dalton, W.S. and Friend, S.H. (2006). Cancer biomarkers – an invitation to the table. *Science* 312(5777), 1165–1168.

David-Beabes, G.L. and London, S.J. (2001). Genetic polymorphism of XRCC1 and lung cancer risk among African-Americans and Caucasians. *Lung Cancer* 34(3), 333–339.

Dobbin, K.K., Beer, D.G., Meyerson, M., Yeatman, T.J., Gerald, W. L., Jacobson, J.W., Conley, B., Buetow, K.H., Heiskanen, M., Simon, R.M. et al. (2005). Interlaboratory comparability study of cancer gene expression analysis using oligonucleotide microarrays. *Clin Cancer Res* 11, 565–572.

Dutt, A. and Wong, K.K. (2006). Mouse models of lung cancer. *Clin Cancer Res* 12, 4396s–4402s.

Dziadziuszko, R., Hirsch, F.R., Varella-Garcia, M. and Bunn, P.A., Jr (2006). Selecting lung cancer patients for treatment with epidermal growth factor receptor tyrosine kinase inhibitors by immunohistochemistry and fluorescence in situ hybridization – why, when, and how? *Clin Cancer Res* 12, 4409s–4415s.

Engelman, J.A. and Cantley, L.C. (2006). The role of the ErbB family members in non-small cell lung cancers sensitive to epidermal growth factor receptor kinase inhibitors. *Clin Cancer Res* 12, 4372s–4376s.

Engelman, J.A., Mukohara, T., Zejnullahu, K., Lifshits, E., Borras, A.M., Gale, C.M., Naumov, G.N., Yeap, B.Y., Jarrell, E., Sun, J. et al. (2006). Allelic dilution obscures detection of a biologically significant resistance mutation in EGFR-amplified lung cancer. *J Clin Invest* 116(10), 2695–2706.

Esquela-Kerscher, A. and Slack, F.J. (2006). Oncomirs-microRNAs with a role in cancer. *Nat Rev Cancer* 6(4), 259–269.

Fisher, G.H., Wellen, S.L., Klimstra, D., Lenczowski, J.M., Tichelaar, J.W., Lizak, M.J., Whitsett, J.A., Koretsky, A. and Varmus, H.E. (2001). Induction and apoptotic regression of lung adenocarcinomas by regulation of a K-Ras transgene in the presence and absence of tumor suppressor genes. *Genes Dev* 15(24), 3249–3262.

Fong, K.M., Sekido, Y., Gazdar, A.F. and Minna, J.D. (2003). Lung cancer. 9: Molecular biology of lung cancer: Clinical implications. *Thorax* 58(10), 892–900.

Garber, M.E., Troyanskaya, O.G., Schluens, K., Petersen, S., Thaesler, Z., Pacyna-Gengelbach, M., van de Rijn, M., Rosen, G.D., Perou, C.M., Whyte, R.I. et al. (2001). Diversity of gene expression in adenocarcinoma of the lung. *Proc Natl Acad Sci USA* 98(24), 13784–13789.

Granville, C.A. and Dennis, P.A. (2005). An overview of lung cancer genomics and proteomics. *Am J Respir Cell Mol Biol* 32(3), 169–176.

Graziano, S.L., Gamble, G.P., Newman, N.B., Abbott, L.Z., Rooney, M., Mookherjee, S., Lamb, M.L., Kohman, L.J. and Poiesz, B.J. (1999). Prognostic significance of K-ras codon 12 mutations in patients with resected stage I and II non-small-cell lung cancer. *J Clin Oncol* 17(2), 668–675.

Haiman, C.A., Stram, D.O., Wilkens, L.R., Pike, M.C., Kolonel, L.N., Henderson, B.E. and Le Marchand, L. (2006). Ethnic and racial differences in the smoking-related risk of lung cancer. *N Engl J Med* 354(4), 333–342.

Henschke, C.I., Yankelevitz, D.F., Libby, D.M., Pasmantier, M.W., Smith, J.P. and Miettinen, O.S. (2006). Survival of patients with stage I lung cancer detected on CT screening. *N Engl J Med* 355, 1763–1771.

Hirsch, F.R. and Lippman, S.M. (2005). Advances in the biology of lung cancer chemoprevention. *J Clin Oncol* 23(14), 3186–3197.

Hoffman, P.C., Mauer, A.M. and Vokes, E.E. (2000). Lung cancer. *Lancet* 355(9202), 479–485.

Italiano, A., Vandenbos, F.B., Otto, J., Mouroux, J., Fontaine, D., Marcy, P.Y., Cardot, N., Thyss, A. and Pedeutour, F. (2006). Comparison of the epidermal growth factor receptor gene and protein in primary

non-small-cell-lung cancer and metastatic sites: Implications for treatment with EGFR-inhibitors. *Ann Oncol* 17(6), 981–985.

Janne, P.A., Engelman, J.A. and Johnson, B.E. (2005). Epidermal growth factor receptor mutations in non-small-cell lung cancer: Implications for treatment and tumor biology. *J Clin Oncol* 23(14), 3227–3234.

Jett, J.R. (2005). Limitations of screening for lung cancer with low-dose spiral computed tomography. *Clin Cancer Res* 11, 4988s–4992s.

Jett, J.R. and Midthun, D.E. (2004). Screening for lung cancer: Current status and future directions: Thomas A. Neff lecture. *Chest* 125(5 Suppl), 158S–162S.

Johnson, D.H. (2006). Targeted therapies in combination with chemotherapy in non-small cell lung cancer. *Clin Cancer Res* 12, 4451s–4457s.

Johnson, S.M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., Labourier, E., Reinert, K.L., Brown, D. and Slack, F.J. (2005). RAS is regulated by the let-7 microRNA family. *Cell* 120(5), 635–647.

Karube, Y., Tanaka, H., Osada, H., Tomida, S., Tatematsu, Y., Yanagisawa, K., Yatabe, Y., Takamizawa, J., Miyoshi, S., Mitsudomi, T. et al. (2005). Reduced expression of Dicer associated with poor prognosis in lung cancer patients. *Cancer Sci* 96(2), 111–115.

Kelloff, G.J., Lippman, S.M., Dannenberg, A.J., Sigman, C.C., Pearce, H.L., Reid, B.J., Szabo, E., Jordan, V.C., Spitz, M.R., Mills, G.B. et al. (2006). Progress in chemoprevention drug development: The promise of molecular biomarkers for prevention of intraepithelial neoplasia and cancer – a plan to move forward. *Clin Cancer Res* 12(12), 3661–3697.

Kerr, K.M. (2001). Pulmonary preinvasive neoplasia. *J Clin Pathol* 54(4), 257–271.

Kersting, M., Friedl, C., Kraus, A., Behn, M., Pankow, W. and Schuermann, M. (2000). Differential frequencies of p16(INK4a) promoter hypermethylation, p53 mutation, and K-ras mutation in exfoliative material mark the development of lung cancer in symptomatic chronic smokers. *J Clin Oncol* 18(18), 3221–3229.

Khuri, F.R. and Cohen, V. (2004). Molecularly targeted approaches to the chemoprevention of lung cancer. *Clin Cancer Res* 10, 4249s–4253s.

Kobayashi, S., Boggon, T.J., Dayaram, T., Janne, P.A., Kocher, O., Meyerson, M., Johnson, B.E., Eck, M.J., Tenen, D.G. and Halmos, B. (2005). EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N Engl J Med* 352(8), 786–792.

Kopper, L. and Timar, J. (2005). Genomics of lung cancer may change diagnosis, prognosis and therapy. *Pathol Oncol Res* 11(1), 5–10.

Lam, S., McWilliams, A., LeRiche, J., MacAulay, C., Wattenberg, L. and Szabo, E. (2006). A phase I study of myo-inositol for lung cancer chemoprevention. *Cancer Epidemiol Biomarkers Prev* 15(8), 1526–1531.

Li, J. and Burmeister, M. (2005). Genetical genomics: Combining genetics with gene expression analysis. *Hum Mol Genet* 14(Spec No. 2), R163–R169.

Liu, C.G., Calin, G.A., Meloon, B., Gamliel, N., Sevignani, C., Ferracin, M., Dumitru, C.D., Shimizu, M., Zupo, S., Dono, M. et al. (2004). An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. *Proc Natl Acad Sci USA* 101(26), 9740–9744.

Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A. et al. (2005). MicroRNA expression profiles classify human cancers. *Nature* 435(7043), 834–838.

Luo, R.X., Wu, B., Yi, Y.N., Huang, Z.W. and Lin, R.T. (1996). Indoor burning coal air pollution and lung cancer – a case-control study in Fuzhou, China. *Lung Cancer* 14(Suppl 1), S113–S119.

Lynch, T.J., Sequist, L.V., Joshi, V.A., Jänne, P.A., Bell, D.W., Fidias, P., Lindeman, N.I., Louis, D.N., Lee, J.C., Mark, E.J., Longtine, J., Verlander, P., Kucherlapati, R., Meyerson, M., Haber, D.A. and Johnson, B.E. (2006). Epidermal growth factor receptor mutation testing in the care of lung cancer patients. *Clin Can Res* 12, 4403s–4408s.

Marchetti, A., Martella, C., Felicioni, L., Barassi, F., Salvatore, S., Chella, A., Camplese, P.P., Iarussi, T., Mucilli, F., Mezzetti, A. et al. (2005). EGFR mutations in non-small-cell lung cancer: Analysis of a large series of cases and development of a rapid and sensitive method for diagnostic screening with potential implications on pharmacologic treatment. *J Clin Oncol* 23(4), 857–865.

Okuyemi, K.S., Ebersole-Robinson, M., Nazir, N. and Ahluwalia, J.S. (2004). African-American menthol and nonmenthol smokers: Differences in smoking and cessation experiences. *J Natl Med Assoc* 96(9), 1208–1211.

Olaussen, K.A., Dunant, A., Fouret, P., Brambilla, E., André, F., Haddad, V., Taranchon, E., Filipits, M., Pirker, R., Popper, H.H. et al. (2006). DNA repair by ERCC1 in non-small-cell lung cancer and cisplatin-based adjuvant chemotherapy. *N Engl J Med* 355(10), 983–991.

Palmisano, W.A., Divine, K.K., Saccomanno, G., Gilliland, F.D., Baylin, S.B., Herman, J.G. and Belinsky, S.A. (2000). Predicting lung cancer by detecting aberrant promoter methylation in sputum. *Cancer Res* 60(21), 5954–5958.

Park, J.Y., Park, J.M., Jang, J.S., Choi, J.E., Kim, K.M., Cha, S.I., Kim, C.H., Kang, Y.M., Lee, W.K., Kam, S. et al. (2006). Caspase 9 promoter polymorphisms and risk of primary lung cancer. *Hum Mol Genet* 15(12), 1963–1971.

Pastinen, T., Ge, B. and Hudson, T.J. (2006). Influence of human genome polymorphism on gene expression. *Hum Mol Genet* 15(Spec No 1), R9–R16.

Perera, F.P., Tang, D., Brandt-Rauf, P., Santella, R.M., Mooney, L.V., Tu, Y.H., Bendkowska, I. and Bell, D.A. (2006). Lack of associations among cancer and albumin adducts, ras p21 oncoprotein levels, and CYP1A1, CYP2D6, NAT1, and NAT2 in a nested case-control study of lung cancer within the physicians' health study. *Cancer Epidemiol Biomarkers Prev* 15(7), 1417–1419.

Potti, A., Dressman, H.K., Bild, A., Riedel, R.F., Chan, G., Sayer, R., Cragun, J., Cottrill, H., Kelley, M.J., Petersen, R. et al. (2006a). Genomic signatures to guide the use of chemotherapeutics. *Nat Med* 12(11), 1294–1300.

Potti, A., Mukherjee, S., Petersen, R., Dressman, H.K., Bild, A., Koontz, J., Kratzke, R., Watson, M.A., Kelley, M., Ginsburg, G.S. et al. (2006b). A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. *N Engl J Med* 355(6), 570–580.

Powell, C.A., Spira, A., Derti, A., DeLisi, C., Liu, G., Borczuk, A., Busch, S., Sahasrabudhe, S., Chen, Y., Sugarbaker, D. et al. (2003). Gene expression in lung adenocarcinomas of smokers and nonsmokers. *Am J Respir Cell Mol Biol* 29(2), 157–162.

Raponi, M., Zhang, Y., Yu, J., Chen, G., Lee, G., Taylor, J.M., Macdonald, J., Thomas, D., Moskaluk, C., Wang, Y. et al. (2006). Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. *Cancer Res* 66(15), 7466–7472.

Raunio, H., Hakkola, J., Hukkanen, J., Lassila, A., Paivarinta, K., Pelkonen, O., Anttila, S., Piipari, R., Boobis, A. and Edwards, R.J. (1999). Expression of xenobiotic-metabolizing CYPs in human pulmonary tissue. *Exp Toxicol Pathol* 51(4–5), 412–417.

Reissmann, P.T., Koga, H., Figlin, R.A., Holmes, E.C. and Slamon, D.J. (1999). Amplification and overexpression of the cyclin D1 and epidermal growth factor receptor genes in non-small-cell lung cancer. Lung Cancer Study Group. *J Cancer Res Clin Oncol* 125(2), 61–70.

Richardson, G.E. and Johnson, B.E. (1993). The biology of lung cancer. *Semin Oncol* 20(2), 105–127.

Sequist, L.V., Joshi, V.A., Janne, P.A., Bell, D.W., Fidias, P., Lindeman, N.I., Louis, D.N., Lee, J.C., Mark, E.J., Longtine, J. et al. (2006). Epidermal growth factor receptor mutation testing in the care of lung cancer patients. *Clin Cancer Res* 12, 4403s–4408s.

Sharma, S.V., Fischbach, M.A., Haber, D.A. and Settleman, J. (2006). Oncogenic shock: Explaining oncogene addiction through differential signal attenuation. *Clin Cancer Res* 12, 4392s–4395s.

Shigematsu, H., Lin, L., Takahashi, T., Nomura, M., Suzuki, M., Wistuba, I.I., Fong, K.M., Lee, H., Toyooka, S., Shimizu, N. et al. (2005). Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J Natl Cancer Inst* 97(5), 339–346.

Singhal, S., Vachani, A., Antin-Ozerkis, D., Kaiser, L.R. and Albelda, S.M. (2005). Prognostic implications of cell cycle, apoptosis, and angiogenesis biomarkers in non-small cell lung cancer: A review. *Clin Cancer Res* 11(11), 3974–3986.

Slebos, R.J., Kibbelaar, R.E., Dalesio, O., Kooistra, A., Stam, J., Meijer, C.J., Wagenaar, S.S., Vanderschueren, R.G., van Zandwijk, N., Mooi, W. J. et al. (1990). K-ras oncogene activation as a prognostic marker in adenocarcinoma of the lung. *N Engl J Med* 323(9), 561–565.

Spira, A., Beane, J., Shah, V., Steiling, K., Liu, G., Schembri, F., Gilman, S., Dumas, Y., Calner, P., Sebastiani, P., Sridhar, S., Beamis, J., Lamb, C., Keane, J., Lenburg, M. and Brody, J. (2007). Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat Med* 13, 361–366.

Spira, A. and Ettinger, D.S. (2004). Multidisciplinary management of lung cancer. *N Engl J Med* 350(4), 379–392.

Spira, A., Beane, J., Shah, V., Liu, G., Schembri, F., Yang, X., Palma, J. and Brody, J.S. (2004). Effects of cigarette smoke on the human airway epithelial cell transcriptome. *Proc Natl Acad Sci USA* 101(27), 10143–10148.

Spiro, S.G. and Silvestri, G.A. (2005). One hundred years of lung cancer. *Am J Respir Crit Care Med* 172(5), 523–529.

Sweet-Cordero, A., Mukherjee, S., Subramanian, A., You, H., Roix, J.J., Ladd-Acosta, C., Mesirov, J., Golub, T.R. and Jacks, T. (2005). An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis. *Nat Genet* 37(1), 48–55.

Sy, S.M., Fan, B., Lee, T.W., Mok, T.S., Pang, E., Yim, A. and Wong, N. (2004). Spectral karyotyping indicates complex rearrangements in lung adenocarcinoma of nonsmokers. *Cancer Genet Cytogenet* 153(1), 57–59.

Taioli, E., Gaspari, L., Benhamou, S., Boffetta, P., Brockmoller, J., Butkiewicz, D., Cascorbi, I., Clapper, M.L., Dolzan, V., Haugen, A. et al. (2003). Polymorphisms in CYP1A1, GSTM1, GSTT1 and lung cancer below the age of 45 years. *Int J Epidemiol* 32(1), 60–63.

Takamizawa, J., Konishi, H., Yanagisawa, K., Tomida, S., Osada, H., Endoh, H., Harano, T., Yatabe, Y., Nagino, M., Nimura, Y., Mitsudomi, T. and Takahashi, T. (2004). Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Can Res* 64(11), 3753–3756.

Takeuchi, T., Tomida, S., Yatabe, Y., Kosaka, T., Osada, H., Yanagisawa, K., Mitsudomi, T. and Takahashi, T. (2006). Expression profile-defined classification of lung adenocarcinoma shows close relationship with underlying major genetic changes and clinicopathologic behaviors. *J Clin Oncol* 24(11), 1679–1688.

Thomas, R.K., Weir, B. and Meyerson, M. (2006). Genomic approaches to lung cancer. *Clin Cancer Res* 12, 4384s–4391s.

Thun, M.J., Henley, S.J. and Calle, E.E. (2002). Tobacco use and cancer: An epidemiologic perspective for geneticists. *Oncogene* 21(48), 7307–7325.

Toyooka, S., Tsuda, T. and Gazdar, A.F. (2003). The TP53 gene, tobacco exposure, and lung cancer. *Hum Mutat* 21(3), 229–239.

Varmus, H. (2006). The new era in cancer research. *Science* 312(5777), 1162–1165.

Weinstein, I.B. (2002). Cancer. Addiction to oncogenes – the Achilles heal of cancer. *Science* 297(5578), 63–64.

Wenzlaff, A.S., Cote, M.L., Bock, C.H., Land, S.J., Santer, S.K., Schwartz, D.R. and Schwartz, A.G. (2005). CYP1A1 and CYP1B1 polymorphisms and risk of lung cancer among never smokers: A population-based study. *Carcinogenesis* 26(12), 2207–2212.

Wikman, H., Thiel, S., Jager, B., Schmezer, P., Spiegelhalder, B., Edler, L., Dienemann, H., Kayser, K., Schulz, V., Drings, P. et al. (2001). Relevance of N-acetyltransferase 1 and 2 (NAT1, NAT2) genetic polymorphisms in non-small cell lung cancer susceptibility. *Pharmacogenetics* 11(2), 157–168.

Wikman, H., Kettunen, E., Seppanen, J.K., Karjalainen, A., Hollmen, J., Anttila, S. and Knuutila, S. (2002). Identification of differentially expressed genes in pulmonary adenocarcinoma by using cDNA array. *Oncogene* 21(37), 5804–5813.

Wilson, I.M., Davies, J.J., Weber, M., Brown, C.J., Alvarez, C.E., MacAulay, C., Schubeler, D. and Lam, W.L. (2006). Epigenomics: Mapping the methylome. *Cell Cycle* 5(2), 155–158.

Yamagata, N., Shyr, Y., Yanagisawa, K., Edgerton, M., Dang, T.P., Gonzalez, A., Nadaf, S., Larsen, P., Roberts, J.R., Nesbitt, J.C. et al. (2003). A training-testing approach to the molecular classification of resected non-small cell lung cancer. *Clin Cancer Res* 9(13), 4695–4704.

Yanaihara, N., Caplen, N., Bowman, E., Seike, M., Kumamoto, K., Yi, M., Stephens, R.M., Okamoto, A., Yokota, J., Tanaka, T. et al. (2006). Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* 9(3), 189–198.

Zhou, B.S., Wang, T.J., Guan, P. and Wu, J.M. (2000). Indoor air pollution and pulmonary adenocarcinoma among females: A case-control study in Shenyang, China. *Oncol Rep* 7(6), 1253–1259.

Zochbauer-Muller, S., Fong, K.M., Maitra, A., Lam, S., Geradts, J., Ashfaq, R., Virmani, A.K., Milchgrub, S., Gazdar, A.F. and Minna, J.D. (2001). 5' CpG island methylation of the FHIT gene is correlated with loss of gene expression in lung and breast cancer. *Cancer Res* 61(9), 3581–3585.

RECOMMENDED RESOURCE

Cooper, D.N. (2005). *The Molecular Genetics of Lung Cancer*. Springer, Heidelberg, Germany.

CHAPTER

Genomics in the Diagnosis and Management of Breast Cancer

Erich S. Huang and Andrew T. Huang

INTRODUCTION

Patient Presentation

Soon after her 55th birthday Joan Smith has her first screening mammogram. Her radiologist notes a small, 1 cm radiodensity associated with microcalcifications in the upper outer quadrant of her right breast. She cancels a planned vacation and spends her next week in and out of appointments with surgical, medical, and radiation oncologists. During the many hours she spends in clinic waiting rooms, she sees many women who share her fears and uncertainty. Some are younger, some older. Some are black, some white. The women she sees cover the range of size, shape, and race, but all share similar apprehensions about what is in store for them. Later, her core needle biopsy reveals an invasive carcinoma. From this point, Mrs. Smith is funnelled into a choreographed series of steps. Her surgeon performs a lumpectomy, the radiation oncologist begins planning her outpatient radiotherapy, and the medical oncologist initiates her chemotheraphy Herceptin treatments.

In spite of the variety of women Mrs. Smith sees in the clinic waiting rooms, the treatment they undergo are generally variations on a theme. For the most part, the diagnostic and therapeutic maneuvers they undergo are validated by large studies such as NSABP B-13, B-14, and B-21 or HERA (Table 35.1). All of these studies are established "dogma." They represent the best available evidence in treating women with breast cancer. The approach her doctors take is algorithmic; she is carried stepwise through a sequence of steps according to what is best supported in medical literature.

A patient support group becomes a source of great comfort for Mrs. Smith. There she learns of the revolutionary impact of estrogen blockade for hormone receptor-positive tumors, and Herceptin for HER2-positive tumors. Each member has a unique story to tell about her struggle with the disease; yet ironically, many of them have followed the same path in diagnosis and treatment. How is that, Mrs. Smith wonders? There is one woman in the group, Mrs. Brown, who was diagnosed with a small cancer 6 years ago, had a lumpectomy with clear margins and a negative axillary node dissection, yet she is now struggling with metastasis to her spine. She was initially told that her prognosis was quite good. Having been told the same thing, Mrs. Smith wonders what are the odds that she will find herself in the same situation at a time she will be looking forward to grandchildren and retirement? If Mrs. Brown's physicians could have foreseen her recurrence, how might they have used this information? Would they have treated her differently? In 8 cases out of 10, the standard treatment would have sufficed and cured her of disease, but this is little comfort to Mrs. Brown who proved to be among the 2 out of 10 exceptional cases.

THE PROMISE

Experiences like Mrs. Smith's are shared by more than 200,000 women in the United States yearly. And every year more than 40,000 of such women die from breast cancer – the second

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

TABLE 35.1 Large Clinical Trials for Management of Breast Cancer.

Trial	Treatment	Outcome
NSABP B-04	Total mastectomy versus total mastectomy with XRT versus radical mastectomy	No significant difference in disease-free or overall survival rates
NSABP B-06	Total mastectomy versus lumpectomy versus lumpectomy with XRT	No significant difference in disease-free or overall survival rates; addition of XRT to lumpectomy reduced local recurrence rate from 39% to 10%
NSABP B-13	Surgery alone versus surgery plus adjuvant chemotherapy in node-negative patients with estrogen receptor-negative tumors	Improved disease-free survival rate for adjuvant chemotherapy group
NSABP B-14	Surgery alone versus surgery plus adjuvant tamoxifen	Improved disease-free survival rate with adjuvant tamoxifen group
NSABP B-18	Neoadjuvant chemotherapy with doxorubicin, cyclophosphamide, or both for 4 cycles versus the same regimen given postoperatively	No significant difference in overall survival or disease-free survival rates (53% and 70% at 9 years in postoperative group and 69% and 55% in the preoperative group)
NSABP B-21	Lumpectomy plus tamoxifen versus lumpectomy plus tamoxifen plus XRT versus lumpectomy plus XRT for none-negative tumors <1 cm	Combination of XRT and tamoxifen was more effective than either alone in reducing ipsilateral breast tumor recurrence
NSABP B-27	Neoadjuvant chemotherapy comparing AC × 4 cycles then surgery versus AC × 4 cycles, docetaxel × 4 cycles then surgery versus surgery between 4 cycles of AC and 4 cycles of docetaxel	Groups I and III were combined and compared with group II; clinical and pathological complete response rates increased significantly among patients who received preoperative AC and docetaxel
NSABP B-32	SLN biopsy followed by axillary dissection versus SLN biopsy alone for clinical node-negative patients	SLN indentification rate was similar in both groups, accuracy was high for both, negative predictive value was high for both

SLN: sentinel lymph node; Reproduced from Anderson, *Surgical Oncology Handbook, 4th edition*, 39–40.

leading cause of death among females (Jemal et al., 2006). As the most common cancer diagnosis in women, one out of every six American women will confront breast cancer during her lifetime. The disease is more common in industrialized Northern America and Europe and is rising in developed Asia (Cheng et al., 2000).

In the past 30 years, mortality has decreased as screening mammography has become common practice. Screening mammography is a surveillance measure for clinically occult disease consisting of two views of each breast for asymptomatic women. Standardized criteria known as the BI-RADS (breast imaging reporting and data system) implemented by the American College of Radiology aid clinical decisions based on identified masses, their morphology and calcifications, their size, number, and morphology.

Not without some controversy, studies best support the notion that screening impacts mortality for women 50 years-old and greater, decreasing mortality by 20–30%, while strong evidence suggests similar mortality reduction in 40–49-year-old women. Current recommendations from the National Cancer Institute, American Cancer Society, and the US Preventative Services Task Force are that women between 40 and 50 years seek screening every 1–2 years with annual exams after 50.

Concomitantly, as more women have historically sought screening, incidence has increased due to broader detection, with the largest component of this increase being early Stage I cancers (Chu et al., 1996).

Risk factors include (1) Gender: 99% of breast cancers are found in women. (2) Age: incidence increases with age until menopause, when it plateaus (Peto et al., 2000). (3) Race: the diagnosis is more common among Caucasians, followed by African Americans, then Hispanic and Asian Americans (Jemal et al., 2006). (4) Family history, for which two factors dominate: number of first degree relatives and age. The presence of a first degree relative diagnosed at a younger age (<40 years) essentially doubles one's risk compared to having an elderly relative with breast cancer (Collaborative Group on Hormonal Factors on Breast Cancer, 2001). Familial breast cancers are estimated to comprise approximately 10% of breast cancers; hence, the large majority of breast cancers are sporadic events. (5) Hormonal status: a well-accepted risk modifier. Patients such as women with early menarche and late first full term pregnancy who have longer exposure to estrogen demonstrate higher risk, while those with shorter lifetime exposure are at reduced risk (Clemons et al., 2001).

GENETIC BASES

Genes implicated in familial breast cancers are classically the BRCA1 and 2 tumor suppressors, as well as the ubiquitously significant p53 proto-oncogene. Ashkenazi Jews, for example, are known to have specific BRCA1 mutations at a much higher frequency than other populations. This is attributed to a founder effect when the population of Ashekenazim was small. As groundbreaking their discovery has been, the BRCA genes account for a narrow segment of the disease.

BRCA1 and 2 gene products are involved in DNA repair of double-stranded breaks. This normally occurs via homologous recombination. In the absence of normal BRCA1 and 2 activity, cells demonstrate morphologically abnormal chromosomes attributed to impaired replication. Clinically, mutation of these genes is responsible for 5–6% of all breast cancers and a significant portion of inherited breast and ovarian cancers (Venkitaraman, 2002).

Unfortunately, multiple mutations of these loci are implicated in breast cancer, and widespread testing is controversial in its cost-effectiveness even for patients with suggestive family histories. With a cost of over $3000, most recommendations for primary care providers involve genetic counseling before ordering a test.

If testing is performed and a patient demonstrates a significant mutation, the options include: (1) prophylactic mastectomy, which is estimated to reduce risk of developing breast cancer between 90% and 95% (Hartmann et al., 1999; Rebbeck et al., 2004). Or (2) surveillance with yearly mammograms and MRI scanning. Clearly, when indicated and when providing useful results, BRCA testing provides information that may be acted upon and impacts developing breast cancer.

Other Mendelian disorders include mutation of the p53 tumor suppressor. Familial mutations are associated with Li-Fraumeni syndrome for which there is high penetrance for breast cancer. Epistatically related is mutatiom of an activator of p53, ATM, which results in Ataxia-Telangiectasia, a syndrome of cerebellar ataxia, immune deficiencies, and oculocutaneous telangiectasias associated with elevated propensity for malignancies, including breast cancer. Other germline mutations include those of the PTEN gene, a tumor suppressor whose mutation results in Cowden syndrome manifest in mucocutaneous hamartomas, dermatologic abnormalities and early onset thyroid and breast cancer. And Peutz-Jeghers syndrome, secondary to mutation of STK11, results in gastrointestinal hamartomatous polyps and mucocutaneous pigmentation and increased risk for malignancies, including breast.

The sporadic derangements that contribute to 90% of the remaining breast cancer cases are little understood.

MOLECULAR BASES

Molecular characteristics of breast tumors that have accepted predictive and prognostic impact in breast cancer include hormonal status and HER2.

Observations of the linkage between hormonal status and breast cancer date back to the late 19th century. This, in combination with the modern concept and discovery of the estrogen receptor has led to the paradigm of estrogen modulation in managing receptor-positive breast cancers. Hormonal status is predictive and arguably prognostic. Multiple studies indicate that the presence of estrogen receptors in breast tumors is strongly predictive of response to hormonal therapy, while an overwhelming minority of receptor-negative tumors demonstrate response (Bezwoda et al., 1991; Manni et al., 1980). Estrogen receptor-positive tumors tend to be more well-differentiated by histopathologic criteria and demonstrate lower rates of recurrence at 5 years. This paradigm demonstrates the profound clinical impact that a single clinical marker can have on managing a disease.

The importance of molecular markers is reinforced by the finding that amplification of HER2 in breast tumors has negative impact on survival and relapse (Slamon et al., 1987). Approximately one-fifth of tumors are found to overexpress this member of the epidermal growth factor receptor family. Studies also suggest that such overexpression predicts improved response to anthracycline-based chemotherapy (Thor et al., 1998; Pritchard et al., 2006). Though there is laboratory-based evidence that there are intersections between hormone-mediated pathways and HER2-based pathways, the clinical impact of overexpression in predicting response to hormone modulation is equivocal.

In any case, directed therapy against HER2 with Trastuzumab (Herceptin), a humanized monoclonal antibody to the receptor, represents a significant advance in identifying key molecular characteristics of a tumor and designing a therapy based on these data.

The apprehension that Mrs. Smith feels is reflective of the current state of breast cancer treatment. Accepted treatment modalities have gone a great distance to improve the care and outlook of breast cancer patients, but they do not take account of the heterogeneity of individual tumor biology. While AJCC cancer staging is undoubtedly useful for making treatment decisions, there are many cases where a presumably low-risk patient who receives definitive therapy develops recurrent disease, and cases where patients with presumably advanced tumors survive longer than expected. HER2 and estrogen and progesterone receptor status prove that understanding molecular aspects of someone's disease can dramatically impact their treatment; yet therapy tailored to the full spectrum of molecular biology in a patient's individual tumor is still a promise rather than a reality. The full complexity of biologic response to treatment does not stop with individual tumors, superimposed on these are the differences in how a patient responds to treatment. One to 4% of women who receive the Herceptin monoclonal antibody develop congestive heart failure, and 10% experience significant decrease in cardiac function (Chien, 2006).

The ultimate goal of all medical interventions, whether for cancer or essential hypertension, is to tailor therapy to the individual patient and her manifestation of disease. The current state

of the art is in designating a patient as being part of a population rather than as a unique individual. Our patient Mrs. Smith, by St Gallen's criteria, is a member of a low-risk population when using anatomic and pathologic parameters such as tumor size or axillary node status, but how she will fare based on the inherent properties of her particular manifestation of disease is unknown to us. As she enters a treatment system that is protocol-bound, Mrs. Smith wants desperately to feel as if she were being treated based on the behavior of her own cancer – that her treatment is calibrated according to her biology. Unfortunately, cancer care as it stands today is not amenable to fine gradations in biology because there are no technical means to identify such molecular particularities. In theory, the richness and diversity of information about a tumor that genome-scale technology provides is truly reflective of the biology of a particular tumor. This will provide an avenue, among others such as proteomics, to reaching the ultimate goal of individualized cancer treatment.

PROGNOSIS AND PREDICTION

Oncologic therapy is an exercise in minimizing morbidity and maximizing efficacy. In order to accomplish this, one needs to calibrate treatment, which is often morbid, with extent or aggressiveness of disease. Consequently, staging methods provide standardized criteria for measuring the anatomic extent of disease. National and international cancer organizations embrace these clinicopathologic yardsticks with the hope that clinicians can use them as universal standards for prognosis, treatment, and outcomes. Staging is anatomic and pathologic: tumor size, number of regional nodes, and the presence of metastasis are the sole components of the TNM system. This reflects a historical understanding of cancer beginning in one location, spreading locally, then accessing lymphatics or invading vascular structures. Implicit in this is that anatomic extent is a surrogate for a tumor's behavior. It is generally accepted as a first approximation that more aggressive tumors will manifest themselves at the time of diagnosis by the presence of lymph node metastasis, invasion into surrounding tissues or by distant metastases; while more indolent tumors are likely small and restricted to their site of origin.

Obviously, the weakness of TNM staging is this very assumption. While the TNM system serves admirably for allowing clinicians to gauge how aggressive treatment should be to how advanced or aggressive a particular patient's disease. Most clinicians acknowledge that it is a surrogate, an approximation of tumor behavior. For instance, modern practice has discarded removal of the pectoral muscle from Halsted's radical mastectomy as originally described, because this aggressive approach to local control ultimately does not save any more lives than the "modified" radical procedure that spares the muscle. The biology of breast cancer suggests that some point of its natural history involves leapfrogging past adjacent tissue – that it becomes a systemic disease. And at the time of surgery, clinicians simply concede their inability to gauge whether micrometastasis has already occurred when there is no gross evidence of invasion. Likely for this reason, while most women with curative resections of small tumors and negative axillary nodes are considered cured, a significant fraction returns years later with metastatic disease.

MOLECULAR MARKERS

Hormone receptor status and presence of HER2 represent molecular markers that undoubtedly impact breast cancer clinical management. The power of adjuvant and neo-adjuvant therapies directed against these entities suggests that therapies specifically directed against growth pathways, whether actuated by hormones or oncogenes (or both) can provide individualized therapy for the characteristics of a specific tumor. A natural outgrowth of this thinking is that identifying more molecular markers will highlight more pathways and more avenues for developing specific therapies. Until recently, there were no technically facile methods to survey and identify differentially expressed genes. Classical molecular biology required a hypothesis and painstaking analysis on a gene-by-gene basis. A graduate or post-doctoral fellow might spend a year assaying the acti-vity of a handful of genes by transferring them to nylon membranes and "blotting" them with radioisotopes to generate a hundred datapoints. With the draft completion of the Human Genome Project, microarrays represent a viable technology for analyzing the activity of several orders of magnitude more genes.

Microarrays are technological descendents of the Southern blot. Fixing a nucleic acid to a substrate provides a method for taking advantage of complimentary hybridization and various fluorescent or radioactive tagging schemes to establish the presence and abundance of a particular nucleic acid sequence from a sample. What a Southern or Northern blot can do for a handful of sequences, a microarray can do in multiplicity. The greater the number of distinct nucleic acids sequences represented on an array, the broader the survey of the genes being expressed in a tissue at the time it is harvested. Therefore tagged substrate derived from RNA harvested from a particular tumor when hybridized to an array can provide an estimate of how actively genes are being expressed at the time of harvesting. Technology has merely provided a complete genome sequence and the means to increase the physical density of sequences to a point where many thousands of sequences can be assayed in a timely manner.

Current generation arrays encompass most of the human genome. Therefore the scale of data provided by microarray experiments expands by orders of magnitude from the 100s available with traditional molecular biology, to the hundreds of thousands or even millions afforded by genome-scale assays. This explosion of data requires methods to analyze it. Researchers are embracing many methodologies for contending with data on such a large scale.

GENOMIC INSIGHTS

Individual Tumors are Identifiable in Principle

A variety of studies in the late 1990's and early 2000's demonstrated that microarray data from neoplastic tissues provided sufficient information to distinguish biological differences. Work from investigators such as Golub or Alizadeh respectively demonstrated not only that hematologic malignancies such as AML and ALL might be distinguished by "class prediction" tools, but that genome-scale microarray data could differentiate between chemotherapy-responsive and unresponsive large B cell lymphomas even if they were microscopically indistinguishable (Alizadeh et al., 2000; Golub et al., 1999).

Extending on this work, investigators soon realized that the richness and complexity of microarray data could provide "molecular portraits" of breast tumors. These portraits were so variegated and distinct that patterns of gene expression data from tumors after the presumed seismic changes of chemotherapy possessed enough of the characteristics of the tumors before therapy to be traced back to original patient. This finding held true when comparing a single individual's primary tumor and tissue from a lymph node metastasis. When using hierarchical clustering schemes, gene expression patterns from an individual, whether from tumor samples taken before chemotherapy or after, or from an individual's primary tumor and lymph node metastases, were more similar to one another than to samples between individuals, suggesting that genome-scale expression data possesses enough information to characterize an individual's particular tumor, even after the changes in that particular tissue's biology represented by exposure to doxorubicin or metastasis to a new microenvironment (Sorlie et al., 1999).

From a technical standpoint, this finding supports that varying expression levels of RNA even when derived from samples taken at different times from the same patient represent genuine biology about that patient. It is how to interpret these data and rigorously link biology with these "pervasive" gene expression patterns, as the authors put it, that continues to be the subject of active investigation.

Luminal and Basal Tumors

A follow-up study from the same investigators demonstrated the expression data derived from ER-positive tumors fell into two categories, Luminal A and B, while estrogen receptor-negative tumors also possessed distinct patterns and could be classified as Normal, Basal-like and Her2-related. More importantly, these classifications corresponded to clinical differences between patients.

The term "luminal" was applied to ER-positive tumors because ER-responsive genes such as trefoil factor 3, LIV-1, and GATA-binding protein thought to typify luminal epithelium were well represented. The distinction between Luminal A and B lay in lower expression of ER-responsive genes in the B subgroup. Particularly compelling were differences in survival between women whose breast cancers fell into either

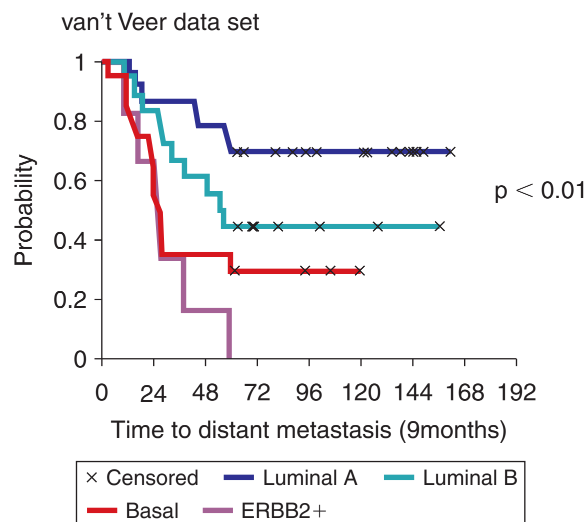


Figure 35.1 Kaplan–Meier analysis of disease outcome. Time to development of distant metastasis in the 97 sporadic cases from van't Veer et al. Patients were stratified according to the subtypes of Luminal A, Luminal B, Basal and ERBB2 (adapted from Sorlie et al., 2001).

of these taxonomies. Women classified as having Luminal A tumors tended to have better prognosis than those with Luminal B. Carrying the analysis further, patients with Basal-like and Her2-related tumors had poorer prognosis than either Luminal subtype with Her2-related tumors possessing the steepest Kaplan-Meier curves.

This type of study suggested that gene expression data could highlight subtle and clinically significant differences between breast tumors in a manner that can extend on conventional histopathologic diagnosis. It also demonstrated that a link might be drawn between the clinical behavior of a tumor and the patterns of gene expression within it (Sorlie et al., 2001) (Figure 35.1).

NETHERLANDS CANCER INSTITUTE STUDY

A prominent application of genome-scale expression studies to prognostication is a study of 78 tumors from lymph node-negative patients younger than 55 from the Netherlands Cancer Institute. A 70-gene classification signature was developed for distinguishing patients with poor prognosis defined as recurrence or metastasis within 5 years from patients with good prognosis (disease-free survival up to 5 years).

This study was promptly followed with a validation of this same signature in a more diverse group of tumor samples from a patient group with a mixture of lymph node positive and negative stage I and II disease. The classifier performed very well in patients with good prognosis; at 5 years it correctly identified 95% with distant metastasis-free disease. When taken out to 10

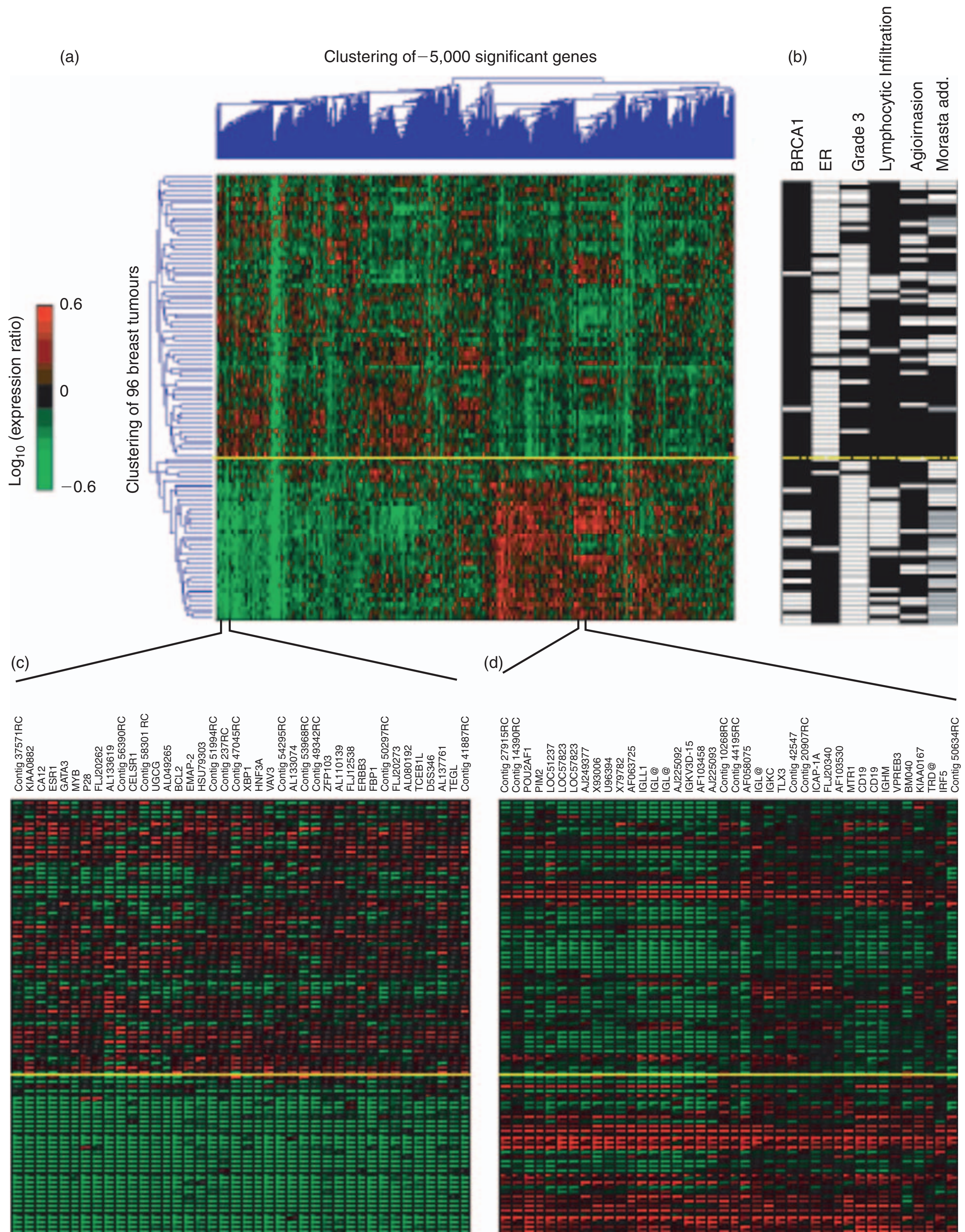


Figure 35.2 The Mammaprint 70 "prognosis classifier genes" across 295 consecutive patients.

years, 85% of patients were correctly classified. When applied to poor prognosis, however, its predictive capabilities were less satisfying. Five year accuracy was only 61%. By 10 years, this rate had declined to 51% (van de Vijver et al., 2002).

Accordingly, an EORTC-sponsored prospective randomized clinical trial based on this classifier, titled MINDACT (Microarray for Node Negative Disease May Avoid Chemotherapy), is in accrual in Europe. Its primary focus is assessing whether patients with clinically high-risk disease, whose 70-gene classifier suggests that they should actually be in a low-risk category can be "safely spared chemotherapy without affecting distant metastasis-free survival." Such a study represents the logical next step after numerous retrospective studies of relatively limited numbers of patients. It should be noted in this particular approach is that gene expression data and clinical parameters are placed in opposing camps.

In the United States, the 70-gene signature is now approved for marketing by the Food and Drug Administration as MammaPrint. It is not yet available for clinical use (Figure 35.2).

DUKE-TAIPEI STUDY

In a similar vein, a collaborative study involving investigators at Duke University and the Koo Foundation Sun Yat-Sen Cancer Center takes a different approach to using gene expression data to develop classifiers for understanding lymph node metastasis and 3-year recurrence in 89 cases. Lymph node status being the strongest clinical predictor of prognosis in breast cancer, it was used as a proof-of-principle for using gene expression data in breast cancer.

Rather than utilizing a single classifier, their analysis is based on multiple classifiers or "metagenes" that are used in aggregate via recursive partitions of the sample set. In other words, a single classifier metagene may correctly separate a large part, but not all of a dataset into lymph node positive or negative patients, and by successively "partitioning" the data with additional classifiers, the prediction of lymph node status is sharpened. A final predictive model is ultimately a composite of many different metagenes or classifiers. An additional advantage to such an approach is that metagenes provide a basis to understand the disparate biological processes in lymph node metastasis rather than attempting to compress a complex biological phenomenon into a single classifier.

Using this methodology, appropriately identifying whether a patient was likely to be lymph node positive or negative was achieved 90% of the time. Extending the analysis to recurrence at 3 years after diagnosis demonstrated similar accuracy (Huang et al., 2003a, b) (Figure 35.3).

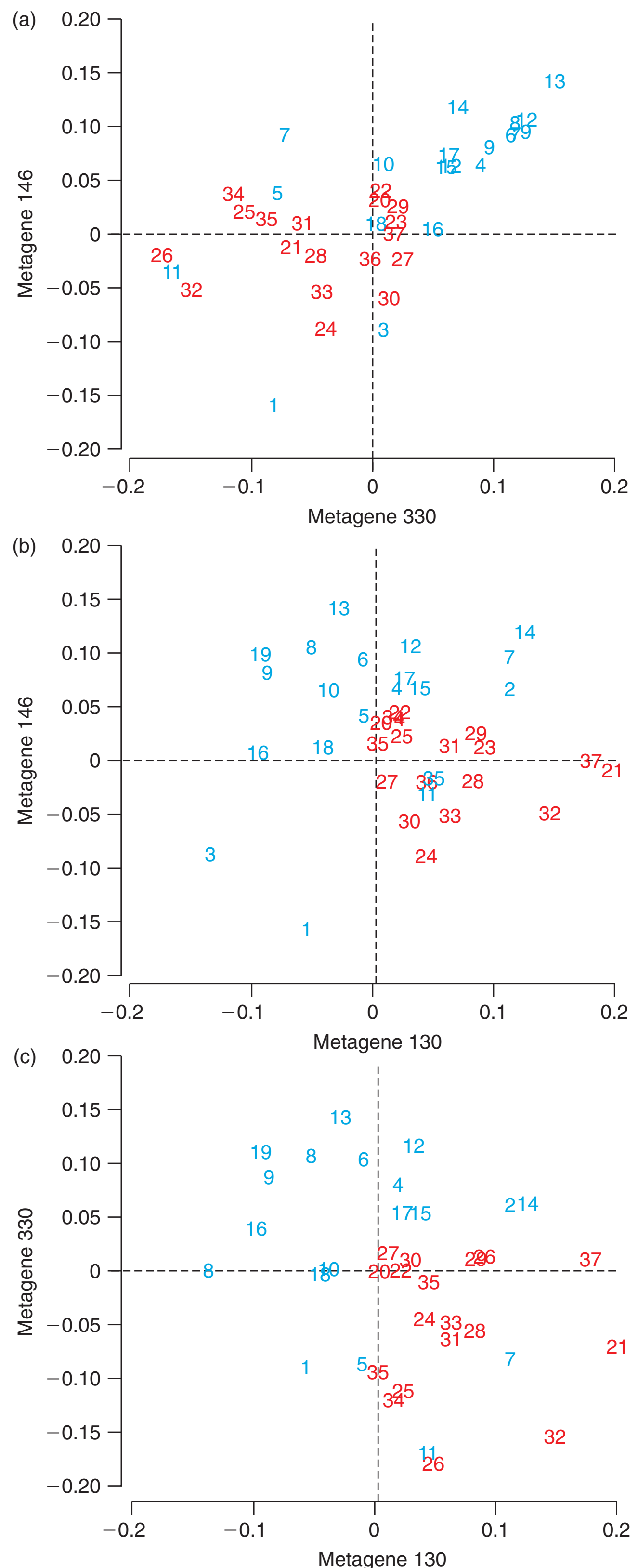


Figure 35.3 Differential contributions of separate metagenes in classifying high from low-risk patients. High risk are labeled in red, while low risk are in blue. Note how different metagenes segregate these patients in different manners.

NSABP STUDY

Members of the NSABP focused their attention on understanding distant recurrence in patients with node-negative,

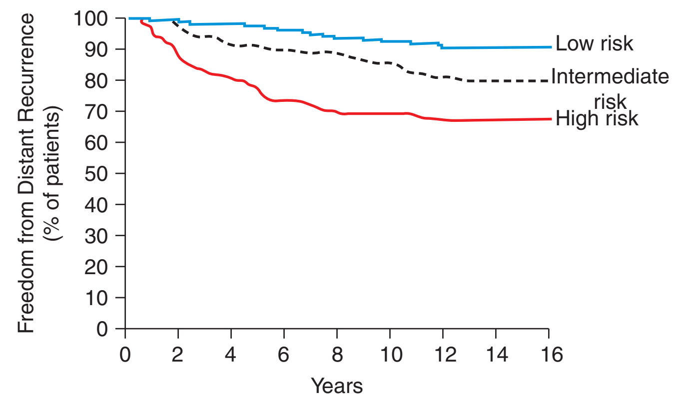
estrogen-receptor positive breast cancers. Successfully making this prediction could potentially obviate chemotherapy for the 85% of patients for whom tamoxifen after surgery is sufficient for preventing 10-year distant recurrence. In a departure from many such studies, the investigators use RT-PCR from paraffin-embedded samples up to 10 years old. Their assay consists of a set of 21 genes (including 5 reference genes) derived from an original set of 250 genes studied across three previous investigations. Their selection was based on their technical robustness in the process of RT-PCR and their performance as predictors. Genes within this set included known proliferation and cell cycle entities, as well as genes linked to tissue invasion and estrogen. An empirically derived "recurrence score" from these RT-PCR data determines whether a patient is low, intermediate, or high risk based on data from three training studies including samples from NSABP B-20 (Figure 35.4). Clinical data from this last study were used to establish the cutoff points for each of these recurrence categories. As applied to 668 samples in the NSABP collection, the investigators tested several hypotheses: first, whether the a low recurrence score accurately anticipated whether a patient would be free of distant recurrence more than 10 years after surgery, and whether there were a statistically significant relation between these two. By these criteria, in the setting of predicting 10-year recurrence in a population of node-negative, estrogen-receptor positive women, the recurrence score was accurate. In comparison to traditional clinical risk factors such as age and tumor size, the score demonstrated a higher degree of statistical significance (Paik et al., 2004). This predictor (commercialized as the Oncotype Dx assay by Genomic Health), is being validated in a Phase III trial (TAILORx Breast Cancer Trial) sponsored by the National Cancer Institute and guided by the Eastern Cooperative Oncology Group.

The trial will analyze whether patients within a low to intermediate risk range will benefit from addition of chemotherapy to hormonal therapy; and whether a particular threshold recurrence score can be used to anticipate whether chemotherapy would be beneficial or not. Patients below the risk range will receive hormonal therapy alone, while patients above this range will receive combined chemotherapy and hormonal therapy. Patients in the risk range will be randomized to hormonal therapy alone versus combined therapy.

As this trial accrues, Oncotype DX is finding use among medical oncologists on a day-to-day basis to inform the process of placing estrogen-receptor positive, node-negative patients on chemotherapy (Figure 35.4).

PATHWAY PREDICTION

Another promising application of genome-scale information is for pathway prediction. Early pilot studies indicated that overexpression of oncogenic activities in cell culture could evoke transcriptional profiles that both predicted and quantified oncogenic pathway activation in a fashion robust enough to be applied *in vivo* and appropriately identify the dysregulated activities evoking tumorigenesis in transgenic mouse models.



No. at Risk									
Low risk	338	328	313	298	276	258	231	170	38
Intermediate risk	149	139	128	116	104	96	80	66	16
High risk	181	154	137	119	105	91	63	13	

Figure 35.4 Likelihood of distant recurrence using the Oncotype DX assay to identify node-negative, estrogen-receptor positive patients to distinguish low versus high-risk patients.

This provides a facile approach to interrogating the potentially numerous growth activation, local invasion, metastatic, and oncogenic pathways that may be present in breast cancer. Analysis of the effects of chemotherapeutic drugs on lung cancer and breast cancer tissue culture lines suggest that this strategy identify metagenes that appropriately predict response to pharmacologic intervention (Huang et al., 2003a, b; Bild et al., 2006); Potti et al., 2006). This line of work will potentially provide clinicians tools based on tumor biology to decide whether patients for whom the decision to institute chemotherapy is equivocal by conventional clinico-pathologic criteria should undertake the toxicities of systemic therapy for maximal benefit.

THE REALITY OF CLINICAL GENOMICS

At this point in its development, the application of gene expression analysis to clinical decision-making in breast cancer is promising. The studies outlined above provide good evidence that genomics can potentially unmask the molecular particularities of a woman's breast cancer, but until the time-tested standard of a randomized, prospective trial actually validates this possibility, it remains only a promise.

There is little doubt that the massively multivariate data provided by genome-scale expression analysis of breast tumors potentially revolutionizes our understanding and our clinical approach to breast cancer. This depends on robust and reproducible measurement of the expression of thousands of genes. Yet at this point in time, it is arguable that microarray data alone are neither robust nor reproducible. The strong movement to using RT-PCR data on a handful of selected genes, as with the Oncotype DX product from Genomic Health, Inc. is a byproduct of skepticism about the consistency of genome-scale expression data derived

from microarrays. On the other hand, 5 years of experience with gene expression data demonstrate that their richness and complexity are concordant with the heterogeneity of cancer biology. Even early studies with a limited number of genes being studied–with first generation methodologies for interpreting expression data–show that there are pervasive expression patterns that unambiguously link a tumor to a single individual regardless of whether samples were taken before or after chemotherapy.

Many groups ultimately treat microarrays as a convenient screening tool for developing a restricted subset of genes that can be used as a predictor for a single outcome measure. Among the best known, the 70-gene predictor that is currently being tested in European clinical trials, and the 21-gene RT-PCR Oncotype DX assay that is currently commercially available to patients, divide patients into two to a few subgroups. Dividing patients into a "high risk" versus "low risk" groups, while representing a potential refinement in clinical decision-making, does not do justice to the potential of gene expression data.

The strongest challenge to such methodology is that further study of the dataset that produced the 70-gene predictor by an independent group reveals that there are many 70-gene predictors utilizing different genes that can predict risk with no less accuracy (Ein-Dor et al., 2005). If this is the case, why choose a particular group of 70 genes? It is tempting to take this analysis as evidence that the field need be reassessed and that predictors based on microarrays merely represent an arbitrary choice of what genes to include.

On the other hand, a new study in which data from 295 patients were analyzed utilizing previously published prognostic and predictive gene expression models showed that there was a remarkable degree of "concordance" – similar classification ability – between models even when the gene sets diverged significantly. The authors concluded that similar biological phenotypes were being tracked, even if constituent genes in a particular classification scheme might differ (Fan et al., 2006). Such a finding suggests a different interpretation: one may approach the data by extending the analysis beyond making prediction on a single outcome. Surely a group of patients can be divided into "high risk" versus "low risk", luminal versus basal, high probability for recurrence or low, and surely multiple gene expression-based models do equally well in making such divisions. So what does the presence of so many essentially "indistinguishable" predictors mean?

Where the answer lies is in the fact that the questions being asked may be extended further. If microarrays afford one a dataspace of thousands of patients and hundreds and thousands of genes, why stop at one division of the data? Does a single classification of "high risk" versus "low risk" do justice to the data? Why not apply every possible predictor to subdivide this complex dataspace as many ways as possible? In theory, by successively applying different predictors to a group of breast cancer patients, one can segregate the set down to small, homogeneous groups of patients with individualized prognosis. One might argue that "similar biological phenotypes" is an oversimplification. If one only wanted to divide a group of patients in two and stop, different gene expression models may perform similarly, but if one wanted to get at the biological particularities of particular patients, successive partitioning with different models (what our group would call metagenes) might carry us to that point.

Taking a step in this direction, one group found that a "core serum response" (CSR) profile derived from analysis of quiescent fibroblasts stimulated with serum appeared to be a useful way to study the behavior of tumors (Chang et al., 2004). By segregating a patient population into high risk and low risk first with the 70-gene predictor, and then applying the CSR. They found that patients who were high risk in addition to having an active CSR signature, were at much higher risk for metastatic disease than those with an inactive or "quiescent" profile. This method of successively partitioning genome-scale data maximizes one's opportunity for using all the useful data present in a dataset (Chang et al., 2005).

2009 UPDATE

At the time of this update, the two largest prospective, randomized trials for the application of gene expression signatures in breast cancer – the TAILORx (OncotypeDX), and MINDACT (Mammaprint) trials – continue to accrue patients. As noted in the main chapter, these trials seek to identify presumably low-risk patients with early stage, lymph node-negative disease who are at risk for recurrence. As such, the OncotypeDx and Mammaprint platforms, which both interrogate a small subset of genes, are prognostic tests. Their intent is to unveil the discordance between a clinically low-risk phenotype and a molecular phenotype indicative of aggressive biologic behavior and thereby optimize the risk-benefit equation for using cytotoxic chemotherapy. These trials represent a critically important first step in developing level I data for the day-to-day utility of gene expression indicators in clinical practice.

With the goal of tailoring a specific therapy to a specific breast cancer in mind, the decision point arrived at by the OncotypeDX or Mammaprint platforms does not specifically address how to tailor the treatment itself to the molecular characteristics of an individual's manifestation of breast cancer. This is particularly relevant with the growing armamentarium of pathway-directed therapies such as receptor-targeted biologics or Tyrosine Kinase Inhibitors.

One irony of breast cancer treatment is that, while modulation of the estrogen pathway and targeting of the ERBB2 (HER2) cell surface receptor represent landmark applications of growth pathway-directed therapies, large numbers of patients whose markers suggest that they should respond to these treatments possess disease that is unresponsive.

Evidence indicates that polymorphisms in loci related to the oxidative metabolism of Tamoxifen may explain why

estrogen receptor-positive breast cancers "escape" the inhibitory effects of Tamoxifen—a phenomenon long attributed to tumor resistance to estrogen pathway blockade. Patients with particular CYP2D6 polymorphisms demonstrate lower levels of Endoxifen, a potent metabolic product of Tamoxifen (Lim, et al., 2005; Tan, et al., 2008). As presented at the San Antonio Breast Cancer Symposium in 2008, retrospective subgroup analysis of the Austrian Breast and Colorectal Study Group-8 trial (a prospective trial evaluating treatment of post-menopausal women with estrogen receptor-positive breast cancers with 5 years of Tamoxifen, or 2 years of Tamoxifen followed by 3 years of Arimidex (Anastrozole), an aromatase inhibitor) showed that women treated with Tamoxifen alone, who were also identified to be "poor metabolizers" of Tamoxifen, had nearly four-fold increased risk for recurrence over "extensive metabolizers". Interestingly, poor metabolizers who were randomized to the switch to an aromatase inhibitor and who had not developed a recurrence while on Tamoxifen experienced a normalization of their relative risk for recurrence (Goetz, et al., 2008). This particular example is illustrative of metabolic polymorphisms rather than intrinsic characteristics of the tumor itself that can explain heterogeneity in response to treatment.

While the HER2-NEU marker is an important prognostic marker and identifies patients who potentially benefit from Herceptin (Trastuzumab), only a third of patients who demonstrate either ERBB2 amplification or high HER2-NEU expression at the cell surface respond to treatment. One explanation lies at the receptor level where there is evidence to suggest that some HER2-positive tumors may possess a truncation at the amino terminal that corresponds to impaired response to trastuzumab (Scaltriti et al., 2007). Another explanation lies in dysregulation of a downstream pathway. A functional genomic RNA interference (RNAi) screen for determinants of trastuzumab resistance has shown that PTEN is a dominant modifer of trastuzumab response. While PTEN mutations are only observed in a small minority of clinical samples, activating mutations of the upstream gene PIK3CA are more common and confer a similar resistance phenotype (Berns, et al. 2007).

Such data argue that tests that assess the functional status of cancer-related pathways may ultimately prove more useful in the clinical arena than individual mutations. In this vein, tests that only measure a restricted subset of genes cannot provide the depth of information needed to reconstruct the patterns of gene expression related to pathway activation. Genome-scale transcriptional profiling is one strategy to carry currently available transcriptome-based technologies from the realm of the prognostic to the predictive.

2009 UPDATE REFERENCES

Berns, K. et al. (2007). A functional genetic approach identifies the PI3K pathway as a major determinant of trastuzumab resistance in breast cancer. *Cancer Cell* 12(4), 395–402.

Goetz, M.P. et al. (2008). Cytochrome P450 2D6 and homeobox 13/interleukin-17B receptor: combining inherited and tumor gene markers for prediction of tamoxifen resistance. *Clin Cancer Res* 14(18), 5864–5868.

Lim, Y.C., Zeruesenay D., David, A.F. and Skaar, T.C. (2005). Endoxifen (4-hydroxy-N-desmethyl-tamoxifen) has anti-estrogenic effects in breast cancer cells with potency similar to 4-hydroxy-tamoxifen. *Cancer chemotherapy and pharmacology* 55(5), 471–478.

Llombart-Cussac, A. (2008). Improving decision-making in early breast cancer: who to treat and how? *Breast Cancer Res Treat*, 12(Supplement1), 15–24.

Olufunmilayo, O., Grushko, T., Nanda, R. and Huo, D. (2008). Advances in Breast Cancer: Pathways to Personalized Medicine. *Clin Can Res* (AACR) 14(24), 7988.

Riggins, R.B., Schrecengost, R.S., Guerrero, M.S. and Bouton, A.H. (2007). Pathways to tamoxifen resistance. *Cancer letters* 256(1), 1–24.

Scaltriti, M. et al. (2007). Expression of p95HER2, a truncated form of the HER2 receptor, and response to anti-HER2 therapies in breast cancer. *J Nat Can Inst* 99(8), 628–638.

Shen, R., Chinnaiyan, A.M. and Ghosh, D. (2008). Pathway analysis reveals functional convergence of gene expression profiles in breast cancer. *BMC Med Genomics* 1, 28.

Sotiriou, C. and Piccart, M.J. (2007). Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care? *Nat Rev Cancer* 7(7), 545–553.

Swanton, C. and Downward, J. (2008). Unraveling the complexity of endocrine resistance in breast cancer by functional genomics. *Cancer cell* 13(2), 83–85.

Tan, S.-H., Lee, S.-C., Goh, B.-C. and Wong, J. (2008). Pharmacogenetics in Breast Cancer Therapy. *Clin Can Res* 14(24), 8027.

REFERENCES

Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., Powell, J.I., Yang, L., Marti, G.E., Moore, T., Hudson, J., Jr, Lu, L., Lewis, D.B., Tibshirani, R., Sherlock, G., Chan, W.C., Greiner, T.C., Weisenburger, D.D., Armitage, J.O., Warnke, R., Levy, R., Wilson, W., Grever, M.R., Byrd, J.C., Botstein, D., Brown, P.O. and Staudt, L.M. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403(6769), 503–511.

Bezwoda, W.R., Esser, J.D., Dansey, R., Kessel, I. and Lange, M. (1991). The value of estrogen and progesterone receptor determinations in advanced breast cancer. Estrogen receptor level but not progesterone receptor level correlates with response to tamoxifen. *Cancer* 68(4), 867–872.

Bild, A.H., Yao, G., Chang, J.T., Wang, Q., Potti, A., Chasse, D., Joshi, M.B., Harpole, D., Lancaster, J.M., Berchuck, A., Olson, J.A., Jr, Marks, J.R., Dressman, H.K., West, M. and Nevins, J.R. (2006). Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 439(7074), 353–357.

Chang, H.Y., Sneddon, J.B., Alizadeh, A.A., Sood, R., West, R.B., Montgomery, K., Chi, J.T., van de Rijn, M., Botstein, D. and Brown, P.O. (2004). Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol* 2(2), E7.

Chang, H.Y., Nuyten, D.S., Sneddon, J.B., Hastie, T., Tibshirani, R., Sørlie, T., Dai, H., He, Y.D., van't Veer, L.J., Bartelink, H., van de Rijn, M., Brown, P.O. and van de Vijver, M.J. (2005). Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc Natl Acad Sci USA* 102(10), 3738–3743.

Cheng, S.H., Tsou, M.H., Liu, M.C., Jian, J.J., Cheng, J.C., Leu, S.Y., Hsieh, C.Y. and Huang, A.T. (2000). Unique features of breast cancer in Taiwan. *Breast Cancer Res Treat* 63(3), 213–223.

Chien, K.R. (2006). Herceptin and the heart – a molecular modifier of cardiac failure. *N Engl J Med* 354(8), 789–790.

Chu, K.C., Tarone, R.E., Kessler, L.G., Ries, L.A., Hankey, B.F., Miller, B.A. and Edwards, B.K. (1996). Recent trends in U.S. breast cancer incidence, survival, and mortality rates. *J Natl Cancer Inst* 88(21), 1571–1579.

Clemons, M. and Goss, P. (2001). Estrogen and the risk of breast cancer. *N Engl J Med* 344(4), 276–285.

Collaborative Group on Hormonal Factors in Breast Cancer (2001). Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *Lancet* 358(9291), 1389–1399.

Ein-Dor, L., Kela, I., Getz, G., Givol, D. and Domany, E. (2005). Outcome signature genes in breast cancer: is there a unique set? *Bioinformatics* 21(2), 171–178.

Fan, C., Oh, D.S., Wessels, L., Weigelt, B., Nuyten, D.S., Nobel, A. B., van't Veer, L.J. and Perou, C.M. (2006). Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* 355(6), 560–569.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M. A., Bloomfield, C.D. and Lander, E.S. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286(5439), 531–537.

Hartmann, L.C., Schaid, D.J., Woods, J.E., Crotty, T.P., Myers, J.L., Arnold, P.G., Petty, P.M., Sellers, T.A., Johnson, J.L., McDonnell, S. K., Frost, M.H. and Jenkins, R.B. (1999). Efficacy of bilateral prophylactic mastectomy in women with a family history of breast cancer. *N Engl J Med* 340(2), 77–84.

Huang, E., Cheng, S.H., Dressman, H., Pittman, J., Tsou, M.H., Horng, C.F., Bild, A., Iversen, E.S., Liao, M., Chen, C.M., West, M., Nevins, J.R. and Huang, A.T. (2003). Gene expression predictors of breast cancer outcomes. *Lancet* 361(9369), 1590–1596.

Huang, E., Ishida, S., Pittman, J., Dressman, H., Bild, A., Kloos, M., D'Amico, M., Pestell, R.G., West, M. and Nevins, J.R. (2003). Gene expression phenotypic models that predict the activity of oncogenic pathways. *Nat Genet* 34(2), 226–230.

Jemal, A., Siegel, R., Ward, E., Murray, T., Xu, J., Smigal, C. and Thun, M.J. (2006). Cancer statistics. *CA Cancer J Clin* 56(2), 106–130.

Manni, A., Arafah, B. and Pearson, O.H. (1980). Estrogen and progesterone receptors in the prediction of response of breast cancer to endocrine therapy. *Cancer* 46(12 Suppl), 2838–2841.

Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F. L., Walker, M.G., Watson, D., Park, T., Hiller, W., Fisher, E.R., Wickerham, D.L., Bryant, J. and Wolmark, N. (2004). A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 351(27), 2817–2826.

Peto, J. and Mack, T.M. (2000). High constant incidence in twins and other relatives of women with breast cancer. *Nat Genet* 26(4), 411–414.

Potti, A., Dressman, H.K., Bild, A., Riedel, R.F., Chan, G., Sayer, R., Cragun, J., Cottrill, H., Kelley, M.J., Petersen, R., Harpole, D., Marks, J., Berchuck, A., Ginsburg, G.S., Febbo, P., Lancaster, J. and Nevins, J.R. (2006). Genomic signatures to guide the use of chemotherapeutics. *Nat Med* 12(11), 1294–1300.

Pritchard, K.I., Shepherd, L.E., O'Malley, F.P., Andrulis, I.L., Tu, D., Bramwell, V.H. and Levine, M.N.National Cancer Institute of Canada Clinical Trials Group (2006). HER2 and responsiveness of breast cancer to adjuvant chemotherapy. *N Engl J Med* 354(20), 2103–2111.

Rebbeck, T.R., Friebel, T., Lynch, H.T., Neuhausen, S.L., van't Veer, L., Garber, J.E., Evans, G.R., Narod, S.A., Isaacs, C., Matloff, E., Daly, M.B., Olopade, O.I. and Weber, B.L. (2004). Bilateral prophylactic mastectomy reduces breast cancer risk in BRCA1 and BRCA2 mutation carriers: the PROSE Study Group. *J Clin Oncol* 22(6), 1055–1062.

Slamon, D.J., Clark, G.M., Wong, S.G., Levin, W.J., Ullrich, A. and McGuire, W.L. (1987). Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 235(4785), 177–182.

Sørlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Thorsen, T., Quist, H., Matese, J.C., Brown, P.O., Botstein, D., Eystein Lønning, P. and Børresen-Dale, A.L. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 98(19), 10869–10874.

Thor, A.D., Berry, D.A., Budman, D.R., Muss, H.B., Kute, T., Henderson, I.C., Barcos, M., Cirrincione, C., Edgerton, S., Allred, C., Norton, L. and Liu, E.T. (1998). erbB-2, p53, and efficacy of adjuvant therapy in lymph node-positive breast cancer. *J Natl Cancer Inst* 90(18), 1346–1360.

van de Vijver, M.J., He, Y.D., van't Veer, L.J., Dai, H., Hart, A.A., Voskuil, D.W., Schreiber, G.J., Peterse, J.L., Roberts, C., Marton, M.J., Parrish, M., Atsma, D., Witteveen, A., Glas, A., Delahaye, L., van der Velde, T., Bartelink, H., Rodenhuis, S., Rutgers, E.T., Friend, S.H. and Bernards, R. (2002). A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347(25), 1999–2009.

Venkitaraman, A.R. (2002). Cancer susceptibility and the functions of BRCA1 and BRCA2. *Cell* 108(2), 171–182.

CHAPTER

Colorectal Cancer

G.L. Wiesner, T.P. Slavin and J.S. Barnholtz-Sloan

INTRODUCTION

Over the next year, approximately 112,340 new cases of colorectal cancer (CRC) will be diagnosed in the United States and over one million cases across the globe, making CRC one of the most common cancers worldwide (ACS, 2007; Curado et al., 2007). Accounting for 10% of cancer-related deaths, CRC is the third most common cause of cancer mortality in both men and women in the United States (ACS, 2007). CRC occurs in all races and all peoples, but the rates vary among people with different racial and ethnic backgrounds. African Americans have an increased incidence and mortality of CRC compared to European Americans in the United States, while Hispanic Americans and American Indian/Alaskan Natives have the lowest incidence rates (Ries et al., 2007). Survival rates also differ among racial/ethnic groups and are highly dependent on stage of disease at diagnosis; when the cancer is diagnosed at an early stage the 5-year relative survival can be as high as 90%. However, only about a third of all CRC cases are diagnosed in the early stages (Jemal et al., 2007). Despite this dismal statistic, both CRC incidence and mortality rates have steadily decreased over the last two decades (Ries et al., 2007). The simultaneous decline in these rates is most likely a reflection of an increase in screening practices that interrupt the development of invasive cancer by removing benign colon polyps or early stage disease in screened individuals (Burt, 2000; Vogelaar et al., 2006; Winawer et al., 2003). Screening may also decrease the incidence of CRC by reducing the time that the colon epithelium is exposed to potential environmental risk factors.

Many studies have shown that there are both genetic and environmental causes for CRC, as reviewed (Giovannucci and Wu, 2006; Potter and Hunter, 2002). Several inherited genetic syndromes that involve a high risk for CRC have been discovered such as hereditary nonpolyposis colorectal cancer (HNPCC), familial adenomatous polyposis (FAP), and other rare inherited polyposis syndromes (Rustgi, 2007). However, these syndromes taken together only account for 5% or less of all CRC diagnoses (Figure 36.1; Burt and Neklason, 2005). Important for public health is the approximately 15–20% of CRC cases with a family history of CRC and/or polyps (Burt, 1996a; Sandhu et al., 2001). The family members in this group have an elevated lifetime risk for CRC, implying additional, yet undiscovered, genetic and/or common environmental susceptibility factors underlying colon neoplasia.

The current theory of CRC causation states that multiple sequential mutations in cellular regulatory genes of the colon epithelium occur before an invasive cancer can develop (Figure 36.2; Fearon and Vogelstein, 1990; Vogelstein et al., 1988). This multi-step model of cancer development is also influenced by environmental factors, such as diet or smoking, and modifier genes, which are presumed to additionally alter key cellular functions. It is notable that much of the theoretical framework for understanding the root causes of carcinogenesis have stemmed from genetic and genomic research in CRC. A wider appreciation

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

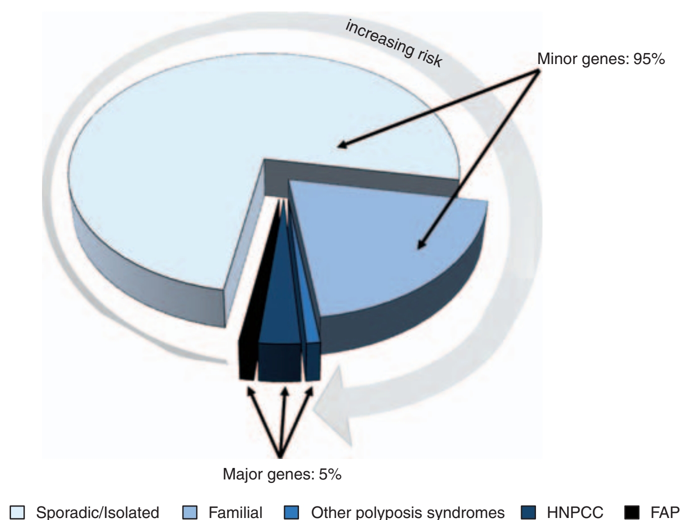


Figure 36.1 Genetic predisposition to CRC. The risk for CRC increases along a continuum; risk is lowest with a sporadic/isolated family history (~75% of cases), moderate with clear positive family history of disease (~20% of cases), and highest in known monogenic inherited CRC syndromes (~5% of cases).

of the multi-step phenomena, the pathways involved, and the global epigenetic alterations in cancer were gained using a genomic viewpoint of disease causation. The increasing availability and affordability of genetic/genomic tools will only improve the diagnosis, screening, treatment, and monitoring of patients in the future. With broader application of these technologies, genomics will be the mainstay of personalized medical care, rather than a private application for the unique patient or family.

This chapter will review the current understanding of genetic predisposition, somatic genomic alterations and environmental influences in the development of CRC. The application of genomic technologies for the patients with CRC or at risk for CRC will also be discussed. Lastly, the emerging genomic-based technologies that promise to personalize and improve healthcare will be reviewed.

GENOMIC MODEL OF CRC

Most malignant tumors of the colon arise from an adenomatous polyp, a benign growth of the colon that is easily identified and removed using colonoscopic screening and biopsy methods. Seminal studies of colonic karyotypes demonstrated the adenoma to carcinoma sequence, in which specific chromosomal and/or genetic alterations correlated with progressive growth of colon tumors from benign adenoma to overt cancer (Figure 36.2; Fearon and Vogelstein, 1990; Vogelstein et al., 1988). In fact, two separate progression pathways for CRC development have been discovered and validated: (1) the chromosomal instability (CIN) pathway and (2) the mismatch repair pathway (MMR) pathway. The multi-step process of carcinogenesis is complex, involving as many as six discrete molecular clonal alterations in the adenomatous tissue over a period of 20 or more years before transforming into an invasive carcinoma (Knudson, 2001). The earliest step in the CIN pathway is a mutation in the APC gene, a key factor in the WNT/β-catenin pathway, which is associated with the formation of small, microscopic clusters of proliferating cells called an aberrant crypt which then develops into a benign adenoma. About 15% of CRC have underlying errors in the mismatch repair (MMR) mechanism, identified by microsatellite instability (MSI), loss of MMR proteins and/or mutations in various DNA repair genes and growth factors (Figure 36.2; Ionov et al., 1993; Thibodeau et al., 1993). The newly described CpG island methylator phenotype (CIMP) may constitute an additional third CRC progression pathway in which epigenetic alterations in promoter regions of regulatory genes that cause the development of polyps and cancer (Toyota et al., 1999).

The key initiating event in the CIN pathway is a mutation in the adenomatous polyposis coli (APC; OMIM 611731) gene, which normally forms a stabilizing complex for β-catenin in the WNT signaling pathway (Albuquerque et al., 2002; Powell et al., 1992). Typically described as a tumor suppressor gene with a "gatekeeper" function, pathologic mutations in APC interrupt the normal APC–β-catenin interaction, which in turn, allows for unregulated nuclear signaling by β-catenin to downstream messengers (Chung, 2000; Kinzler and Vogelstein, 1996). Even though germline mutations in the APC gene cause a small fraction of CRC, approximately 85% of sporadic CRC harbor APC mutations, either by point mutations or overt 5q chromosome deletions (Powell et al., 1992; Rajagopalan et al., 2003; Vogelstein et al., 1988). Non-random gains or losses of chromosomal material is a consistent feature in the CIN pathway, with chromosome 18q and 17p frequently altered, presumably causing haploinsufficiency of gene targets like deleted in colorectal carcinoma (DCC; OMIM 120470) and TP53 (OMIM 191170), respectively, that then confers a selective advantage for specific clones (Rajagopalan et al., 2003).

The factors that initiate, promote, and favor metastasis are not completely understood, but most likely involve acquired mutations in specific gene targets, such as mutations in the K-ras oncogene and the TP53 tumor suppressor gene. Approximately 30–50% of large adenomas and CRC have an acquired mutation in the K-ras gene, a member of the tyrosine kinase super family that links extracellular signals to targeted genes in growth and differentiation pathways (Martinez et al., 1999; Vogelstein et al., 1988). Germline alterations in other Ras genes are often found in somatic tumors and are also associated with cancer development in rare inherited developmental disorders, indicating the pleiomorphic nature of this transcription factor (Schubbert et al., 2007). Similarly, germline mutations in the TP53 gene cause the inherited Li-Fraumeni syndrome (OMIM 151623), while acquired mutations occur in a wide variety of cancer types. For CRC, acquired TP53 mutations are found to be a late event in the adenoma to carcinoma sequence (Kinzler and Vogelstein,

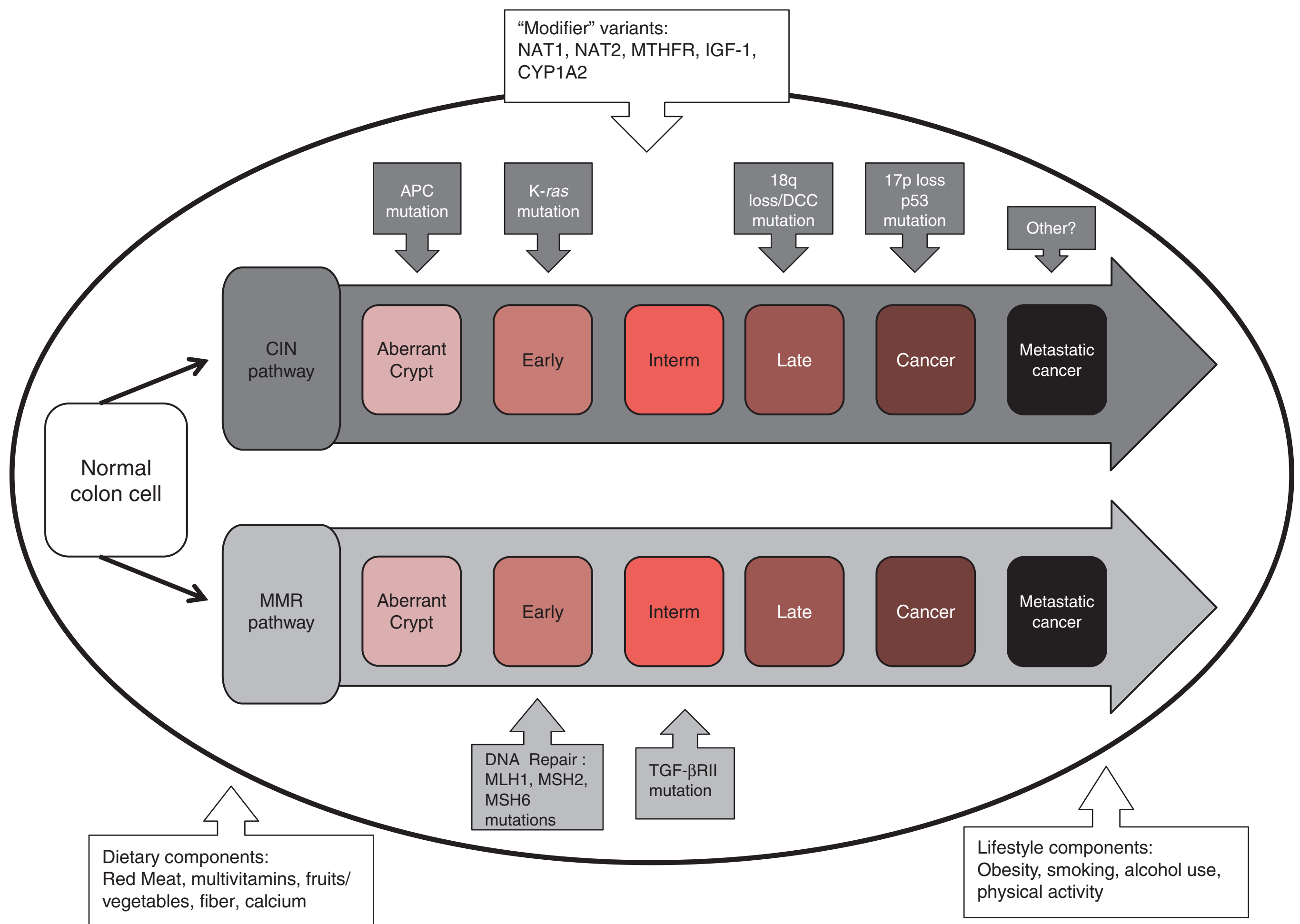


Figure 36.2 Multi-step progression of CRC development. There are many interacting factors, both genomic and environmental, that influence the transformation of a normal colonic epithelial cell into a metastatic cancer. The nuclear progression of the adenoma to carcinoma sequence is illustrated within the cell (i.e., depicted by the oval), denoting the multiple genetic alterations that occur in the CIN or MMR pathways. Outside the cell, are the many environmental dietary and lifestyle factors that promote the initiation and progression of the cancer. These factors may work through the constitutional genome of the individual that encode metabolizing proteins, such as NAT1/2 and MTHFR, that work to modify the course of cancer development. On the edge of the cell, are genes thought to be modifiers that could be involved with factors inside or outside the cell.

1996). As gene discovery in cancer advances, additional somatic mutations in other oncogenes, such as AKT1, will deepen the molecular understanding of CIN pathway progression in CRC (Carpten et al., 2007).

In contrast to CIN tumors, up to 15% of CRC has little, if any, evidence of aneuploidy. MMR-deficient tumors have a recognizable cellular phenotype with poorly differentiated mucinous histopathology with a peritumoral lymphocytic infiltration (Messerini et al., 1996). This type of CRC that evolves along the described MMR pathway (Figure 36.2) is thought to be caused by the loss of repair function of single basepair mismatches formed during S-phase in the epithelial colon cell, either through an inherited or acquired inactivation of one of the MMR genes. This inactivation renders the cell susceptible to further somatic mutations throughout the genome, increasing the cellular mutation rate from 100 to 1000 fold in mono- and di-nucleotide sequences (de la Chapelle, 2004; Gryfe, 2006). Molecular analysis of these sequences can reveal a unique profile known as MSI that correlates with cellular loss of specific proteins encoded by the MMR genes, hMSH2, hMLH1, hMSH6, hPMS1, and hPMS2 (OMIM 609309, 120436, 600678, 600258, and 600259, respectively) (Ionov et al., 1993; Thibodeau et al., 1993). The increased mutation rate in mono- or di-nucleotide tracts could explain the high frequency of acquired mutations in TGFβRII (OMIM 190182) and other genes that appear to be targeted in MSI tumors, as well as the more rapid rate of adenoma

to carcinoma progression compared to the CIN pathway CRC (Boland et al., 2007). TGFβRII (OMIM 190182) is an extracellular receptor and a member of the serine/threonine protein family containing a mono-tract of A_{10} that is somatically mutated in MMR deficient tumors, causing a loss of the natural inhibitory signal to the TGFβ/SMAD pathway and proliferation of the aberrant clone (Boland et al., 2007; Ionov et al., 1993; Markowitz et al., 1995; Samowitz et al., 2005).

The CIN and MMR pathways appear to have little overlap, with the CIN tumors exhibiting distal location in the bowel, non-random chromosomal aneuploidy, and mutations in APC, K-ras and TP53 genes and the MSI tumors primarily exhibiting a proximal location, mucinous histopathology, MSI, loss of MMR proteins, and acquired mutations in TGFβRII. While the initial observation of tumor instability was in tumors from patients with germline mutations in the MMR genes, it is now clear that the majority of MSI tumors are sporadic and may be due to widespread epigenetic inactivation of CpG islands in the promoter regions of specific genes, such as hMLH1 and others. Researchers have proposed that the CIMP is an additional third CRC pathway, with older age of onset and tumors that exhibit a serrated histopathology, MSI, and increased mutations in specific oncogenes, K-ras and BRAF (OMIM 164757) (Deng et al., 2004; Minoo et al., 2007; Samowitz et al., 2005). Serrated adenomas are a newly recognized type of polyp with cancerous potential and hyperplastic and adenomatous features. When advanced, these lesions can develop areas of dysplasia with somatic CpG island methylation and mutations in the BRAF oncogene (OMIM 164757). BRAF mutations can be used to differentiate a tumor into CIMP or CIN, as a correlative study of 126 colon tumors using six CIMP-related markers, six tumor suppressor genes, MSI, and loss of heterozygosity (LOH), demonstrated that the BRAF V600E mutation is highly predictive of the CIMP phenotype (Minoo et al., 2007). Promoter methylation of MLH1 and other genes, such as $p16^{INK4\alpha}$, $p14^{ARF}$ (OMIM 600160) is also now recognized as an important mechanism of colon carcinogenesis and may cause wider dysregulation in the cancer development process, providing further molecular evidence of heterogeneity in CRC carcinogenic pathways (Goel et al., 2007; Minoo et al., 2007). Nonetheless, it is not clear if the CIMP phenotype is a discrete and separate pathway from the MMR deficiency pathway and further work will need to clarify whether specific genes underlie these genomic mechanisms (O'Brien et al., 2006; Ogino et al., 2007; Xu et al., 2004).

The search for the genes in the CIN, MMR and CIMP pathways is ongoing using the current genomic tools of comparative genomic hybridization (CGH) and gene expression microarrays on cancer tissues (Vendrell et al., 2005). Ried and colleagues examined tissue from 14 normal colons, 12 adenomas, 14 high grade adenomas, and 16 CRCs using florescent microarray probes and found increases of chromosome 20q, 13q, 8q, 7p, 1q, and 5p and losses on chromosomes 4, 8p, 18q, and 17p (Ried et al., 1996). Advances in technology have improved the resolution of CGH studies for copy number variations. One recent study examined 42 primary CRC and 37 tumor cell lines, using an oligonucleotide array of 22,500 elements mapping to 16,097 loci. Fifty minimal common regions or MCRs, defined by the occurrence of overlapping events in two or more samples, were identified. Of the 28 amplified and the 22 deleted copy number MCRs, 5 regions held previously known genes, 15 recurring MCR regions were more focal and contained less than 12 genes, and 10 regions contained a total of 65 identifiable genes linked to cancer. Importantly, this method identified 21 new candidate genes that had not previously been linked to CRC (Martin et al., 2007). Correlating gains and losses of chromosomal material with cancer transformation is important, but whether these copy number variants alter the expression pattern and activity of known cancer susceptibility genes is not known. For CRC, this issue has begun to be explored by analyzing colon biospecimens with expression arrays. For example, Nosho and colleagues examined 550 known susceptibility genes for at least a threefold elevation or threefold diminution of activity in 36 adenomas and 14 early invasive cancers and found 13 upregulated and 19 downregulated genes (Nosho et al., 2005). This study confirms the progression of colon adenomas to carcinomas in CRC, but also suggests that the expression of a relatively small number of genes is altered in colon neoplasias.

In order to identify a set of susceptibility genes associated with acquired cancer development, an alternative approach of high-throughput sequencing of over 13,000 candidate genes was used to systematically categorize the genetic alterations in 11 breast and 11 colon tumors. Individual tumors had on average a large number (~90) of different mutated genes even though a small subset of 189 genes was consistently mutated in all of the tumors (Sjoblom et al., 2006). Using a pathway analysis of the global sequencing data, the team developed a "landscape" of breast and colon tumor non-random mutational events and showed that, while some genes were consistently mutated in all tumors examined, the majority of genetic defects in cancer were not frequently mutated. Nonetheless, taken together, these mutations caused defects in specific pathways that provided an overall fitness advantage to that specific clone (Wood et al., 2007). In the future, there will be more reliance on genomic tools in the discovery and delineation of cancer pathways in CRC, with future directions focused on developing and improving personalized tests and treatments.

PREDISPOSITION FOR CRC

Predisposition for CRC is a continuum of increasing risk and is dependent on several risk factors, including age, gender, nationality, existing medical conditions (i.e., previous polyps or inflammatory bowel disease) and family history. From a healthcare perspective, CRC is one of the most common tumors of adulthood, and categorizing the cancer into inherited monogenic, familial and sporadic/isolated forms is useful to determine the level of risk for disease (Figure 36.1). Sporadic CRC is a disease of aging, as the risk is low in younger individuals and increases later in life. The lifetime risk for CRC development in

the United States is ~6% compared to 80–100% in individuals who harbor mutations in monogenic familial syndromes (Burt, 1996b; Doxey et al., 2005; Lynch and de la Chapelle, 1999; Ries et al., 2007). Up to 20% of CRC cases will have a family history of the disease, which confers a moderate two- to three-fold elevation in risk compared to the general population. Although the risk increases for these individuals with earlier onset of disease and with more than one family member affected, there is less precision in determining the level of risk (Fletcher et al., 2007; Sandhu et al., 2001; St John et al., 1993; Winawer et al., 1996). Genetic researchers use these categories to identify mid- to high-risk kindreds for gene discovery and other genomic studies (Wiesner et al., 2000, 2003). Further, these major categories also form the basis for clinical care for individuals at risk for CRC, determining the level of screening, and identification of those patients and specific family members who should consider genetic testing (Lynch et al., 2007).

Inherited Monogenic Syndromes

The characterization of the monogenic CRC syndromes (Table 36.1) has provided an essential understanding of the genomic natural history of cancer. Altogether, these syndromes have been recognized for over 100 years, even though they account for only 5% of all CRC cases (see Figure 36.1; Burt and Neklason, 2005). Primarily divided into polyposis and nonpolyposis syndromes, most monogenic forms are caused by highly penetrant germline alleles in genes that encode proteins in the CIN or MSI pathways (Figure 36.2). Following the multi-hit cancer paradigm, cancer development arises in individuals harboring a germline mutation after acquired alterations of specific regulatory genes occur in the colon epithelium. Clinically, most of these syndromes are inherited as autosomal dominant traits with variable expressivity and incomplete penetrance, with an extremely high risk for individuals and families who carry the mutated allele(s). However, because of the rarity of pathologic alleles in the general population, these mutations do not confer a high attributable risk for CRC in the general population. Thus, recognition of these monogenic syndromes is essential for proper risk management, which begins with genetic evaluation and testing (Doxey et al., 2005; Lagerstedt et al., 2007; Rustgi, 2007).

Familial Adenomatous Polyposis

Familial adenomatous polyposis (FAP; OMIM 175100) was the first recognized form of monogenic inherited CRC accounting for about 1% of all CRC (Rustgi, 2007). Inherited as an autosomal dominant condition, individuals typically develop multiple adenomatous polyps in adolescence, which, if left untreated, will ultimately progress to invasive CRC (Jo and Chung, 2005; Kwak and Chung, 2007). Gastrointestinal manifestations include the development of hundreds to thousands of adenomatous polyps, invasive CRC, gastric cancer, and peri-ampullary duodenal cancer. If prophylactic surgery is not performed, CRC occurs in nearly 100% of mutation carriers, with the average age at onset of initial CRC by 40 years or less (Jo and Chung, 2005; Kwak and Chung, 2007). There are several extra-gastrointestinal manifestations of FAP, some that are useful in diagnosis of the condition, like congenital hyperplasia of the pigment epithelium (CHRPE), and some, such as osteomas and desmoid tumors, that also contribute to the morbidity and mortality of the disorder.

FAP is caused by germline mutations in the APC gene, which has been shown to be a key mutational target in sporadic CIN CRC (Figure 36.2). Karyotypic analysis of two patients with polyposis and developmental delay allowed researchers to identify the location of the gene on the long arm of chromosome 5, illustrating the historical impact of early genomic technology in delineating the causes of CRC (Bodmer et al., 1987; Herrera et al., 1986). After the APC gene was identified using positional cloning, mutational analyses of families with FAP have shown that germline mutations occur throughout the gene, with most mutations causing premature stop codons and more severe disease (Groden et al., 1991). Further genetic analysis of families has also demonstrated several allelic variants of FAP, such as Gardner syndrome, Turcot syndrome, and the attenuated form of FAP (AFAP; OMIM 175100). AFAP is characterized by later onset of CRC, a lesser number of polyps termed oligopolyposis, and a lack of extra-gastrointestinal manifestations. Patients with AFAP have mutations clustering in the 5′ or 3′ segment of the APC gene, while patients with Gardner syndrome have mutations primarily between codons 767 and 1513 (Bisgaard and Bulow, 2006). Patients with profuse and extensive polyposis have been identified with APC mutations near the β-catenin binding site (Nieuwenhuis and Vasen, 2007). Further, some APC alleles, such as the I1307K allele found in up to 28% of patients with Ashkenazi Jewish descent and a family history of disease, do not manifest a high rate of polyp formation, again emphasizing the need to fully explore the genomic mechanism(s) of polyp and cancer formation (Laken et al., 1997, 1999).

Genotype–phenotype correlation studies have also shown that up to 30% of patients with polyposis have no identifiable mutations in the APC gene (Moisio et al., 2002). Using a pathway analysis approach, a possible autosomal recessive form of oligopolyposis called Mut Y Homologue gene (MYH)-associated polyposis (MAP; OMIM 604933) was recognized in 2002, with the description of a sibship with multiple adenomatous polyps and CRC but without detectable mutations in the APC gene. Candidate gene analysis of the excision repairs genes in this kindred identified mutations in the MYH gene located on chromosome 1p (Al-Tassan et al., 2002). Defects in the MYH gene leads to the accumulation of G:C to T:A transversions in genes with susceptible GAA sequences. In fact, APC is an example of a gene that is highly predisposed to these type of transversions (Al-Tassan et al., 2002; Cheadle and Sampson, 2003; Lipton and Tomlinson, 2004). MAP has been identified in 7–8% of APC negative polyposis families and is associated with early-onset colon neoplasia, with the average polyp onset of 46 years and CRC of 49.7 years (Doxey et al., 2005; Jo and Chung, 2005). Interestingly, a recent population-based case control study suggests that the risk for polyps and cancer may not be strictly recessive, as both monoallelic and biallelic mutation carriers

TABLE 36.1 Monogenic colorectal cancer syndromes

Syndrome	Gene	OMIM	Gene map	Major syndrome features[a]
Polyposis				
Adenomatous polyposis				
FAP	APC	175100	5q21–22	>100 Colorectal adenomatous polyps or < 100 polyps with family history
Attenuated FAP	APC	175100	5q21–22	<100 Colorectal adenomatous polyps; oligopolyposis
Gardner	APC	175100	5q21–22	Polyposis, osteoma, desmoid, CHRPE
Turcot	APC	175100	5q21–22	Polyposis, glioblastoma, meduloblastoma
MYH-associated	MYH	604933	1p32–34	Autosomal recessive adenomatous oligopolyposis and pilomatricomas
Hamartomatous polyposis				
Juvenile polyposis	SMAD4	174900	18q21.1	Juvenile polyps, pulmonary arterio-venous malformations, digital clubbing
	BMPR1A	174900	10q22.3	
Cowden	PTEN	158350	10q23.31	Papillomatous papules, acral keratoses, facial trichemommas, mucosal lesions breast/uterine/thyroid/brain/renal cell cancers, fibroids
Bannayan-Ruvalcaba Riley	PTEN	153480	10q23.31	Lipomas, glans penis macules, macrocephaly, pectus excavatum, scoliosis, hyperextensibility, mental retardation
Peutz-Jeghers	STK11 / LKB	175200	19p13.3	Mucosal hyperpigmentation, thyroid cancer, sex cord/gynecomastic tumors, ovarian cysts/cancer, breast/bladder/lung cancers
Basal cell nevus	PTCH	109400	9q22.3	Coarse facies, basal cell cancer, jaw keratocysts, macrocephaly, forehead bossing, facial milia, medulloblastoma, cardiac and ovarian fibromas
Multiple endocrine neoplasia 2B	RET	162300	10q11.2	Mucosal neuromas, enlarged lips, medullary thyroid cancer, pheochromocytoma, "marfanoid" body habitus
Mixed polyposis				
Hereditary mixed polyposis[b]	HMPS1	610069	15q15	Typical juvenile polyps, colonic adenomas, and colorectal carcinomas
Non-polyposis				
HNPCC (Lynch)	MSH2	120435/ 609309	2p21–22	Early-onset MSI CRC diagnosed younger than 50 years; presence of synchronous, metachronous CRC, gastrointestinal cancer, uterine, ovarian cancer. Refer to Bethesda Criteria[c]
	MLH1	120436	3p21.3	As above
	MSH6	600678	2p16	As above, with a higher rate of uterine cancer
	MLH3	604395	14q24.3	No kindreds have been identified with germline mutations to date
	PMS1/2	600258/ 600259	2q31/ 7p22	
Turcot	MLH1 or PMS2	276300		Early-onset CRC and brain cancers, primarily glioblastoma
Muir-Torre	MSH2 or MLH1	609309		Early-onset CRC and sebaceous skin tumor
Familial CRC[b]	CRCS1	608812	9q22–23	Early-onset MSS CRC
	CRCS2	611469	8q24	

[a] Clinical manifestations from Online Mendelian Inheritance in Man (OMIM) (http://www.ncbi.nlm.nih.gov/sites/entrez?db=OMIM; Accessed December 2007).

[b] Linkage analysis identified chromosomal location; specific gene(s) not identified.

[c] Bethesda criteria for HNPCC include CRC in an individual at 50 years or less without a family history or kindreds meeting Amsterdam I or II guidelines (see Box 36.1, Umar et al., 2004 and Lindor et al., 2006).

of the most common MYH Y165C or G382D variants had a twofold increased risk compared to controls (Croitoru et al., 2004). A later population-based study found a similar result for the MYH Y165C variant, but not for other identified variants, suggesting that MYH variants are low-penetrant alleles (Balaguer et al., 2007).

Hereditary Nonpolyposis Colorectal Cancer

Hereditary nonpolyposis colorectal cancer (HNPCC) or Lynch syndrome comprises 1–3% of CRC and is characterized by early-onset proximal CRC exhibiting MSI (de la Chapelle, 2005; Kwak and Chung, 2007). Inherited as an autosomal dominant trait, HNPCC individuals face a 50–70% lifetime risk of developing CRC, in addition to other malignancies (Hampel et al., 2005). Other HNPCC-associated cancers include tumors of the gastrointestinal tract, hepatobiliary system, urinary collecting system, and female reproductive system. Women with HNPCC have a lifetime risk for endometrial cancer as high as 40–60% and up to a ~12% lifetime risk of developing ovarian cancer (Hampel et al., 2006; Kwak and Chung, 2007). Clinical variants of HNPCC also exist (Table 36.1), including Muir-Torre syndrome with skin keratoacanthomas and sebaceous neoplasms, and the Turcot syndrome with brain glioblastomas (Kwak and Chung, 2007; Rowley, 2005; Rustgi, 2007). That these syndromes overlap with similar phenotypes caused by APC/CIN pathway mutations reinforces the concept of genetic and genomic heterogeneity in CRC.

HNPCC is caused by germline mutations in one of five MMR genes, hMSH2, hMLH1, hMSH6, hPMS1, or hPMS2 (Table 36.1). Several guidelines have been developed, including the Amsterdam I, Amsterdam II, and the Bethesda criteria, to assist in the recognition of kindreds with HNPCC (see Box 36.1; Rodriguez-Bigas et al., 1997; Vasen, 2000; Vasen et al., 1991). Initially proposed as eligibility criteria for gene discovery studies, the Amsterdam I guidelines are the most stringent and have been described simply as the 3-2-1 rules, requiring the occurrence of three individuals with CRC in two generations, one of whom developed CRC less than 50 years and is an immediate relative of the other two (Vasen et al., 1991). Amsterdam II family phenotype allows the substitution of other extracolonic tumors, such as endometrial cancer, renal cancer or small bowel cancer in assessing families (Vasen et al., 1999). Because families have been identified with MMR gene mutations who do not meet these criteria, the less stringent Bethesda criteria were developed to clinically identify individuals and families who should consider germline testing (Umar et al., 2004).

Other Major CRC-Associated Genetic Syndromes

The inherited hamartomatous syndromes comprise an additional set of CRC polyposis syndromes distinct from the adenomatous polyposis or nonpolyposis syndromes. The hamartomatous syndromes are characterized by a malformed overgrowth of mesodermal, endodermal, or ectodermal cellular colonic elements (Schreibman, 2005). These rare syndromes carry a large risk of

BOX 36.1 Criteria used for HNPCC testing

Amsterdam criteria I:

1. One member diagnosed with CRC before age 50 years.
2. Two affected generations.
3. Three affected relatives, one of them a first-degree relative of the other two.
4. FAP should be excluded.
5. Tumors should be verified by pathological examination.

Amsterdam criteria II:

1. There should be at least three relatives with a Lynch syndrome-associated cancer (CRC or cancer of the endometrium, small bowel, ureter, or renal pelvis).
2. One should be a first-degree relative of the other two.
3. At least two successive generations should be affected.
4. At least one should be diagnosed before age 50 years.
5. FAP should be excluded in the CRC cases.
6. Tumors should be verified by pathological examination.

Revised Bethesda Guidelines for Colorectal Tumor MSI Testing:

1. CRC diagnosed in an individual younger than 50 years.
2. Presence of synchronous, metachronous colorectal, or other Lynch syndrome-associated tumors (i.e., endometrial, stomach, ovarian, pancreas, ureter and renal pelvis, biliary tract, and brain tumors; sebaceous gland adenomas and keratoacanthomas; and carcinoma of the small bowel), in an individual regardless of age.
3. CRC with MSI-high pathologic associated features diagnosed in an individual younger than 60 years.
4. CRC or Lynch syndrome-associated tumor diagnosed in at least one first-degree relative younger than 50 years.
5. CRC or Lynch syndrome-associated tumor diagnosed at any age in two first- or second-degree relatives.

Reviewed in:

Umar, A., Boland, C.R., Terdiman, J.P., et al. (2004). Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst* 96(4), 261–268.

Lindor, N.M., Petersen, G.M., et al. (2006). Recommendations for the care of individuals with an inherited predisposition to Lynch syndrome: A systematic review. *JAMA* 296(12), 1507–1517.

Genetics of Colorectal Cancer (PDQ®) (http://www.cancer.gov/cancertopics/pdq/genetics/colorectal/healthprofessional)

developing CRC as well as other types of cancer (Table 36.1). Together, they occur at approximately 1/10th the frequency of the adenomatous syndromes and therefore account for less than 1% of CRC (Schreibman, 2005). The approximate incidence of each syndrome listed is 1 in 200,000 births, with the exception of juvenile polyposis syndrome, which is seen 1 in 100,000 births (Schreibman, 2005). While a comprehensive review is beyond the scope of this chapter, each of these syndromes have been associated with mutations in specific pathways, such as the PTEN pathway (OMIM 601728) in Cowden and Bannayan-Ruvalcabe-Riley syndromes and the LKB1 pathway (OMIM 602216) in Peutz-Jeghers syndrome, which again highlights the importance of a genomic approach in delineating the mechanisms of CRC development (Rustgi, 2007).

Familial CRC

Familial CRC accounts for about 20% of all CRC cases and is loosely defined as a positive history of at least one family member with CRC but not meeting the criteria of any of the cancer syndromes discussed previously (Figure 36.1 and Table 36.1). However, this group is heterogeneous and includes families with multiple affected members as well as families with a single affected individual. For example, the label syndrome X has been used to describe kindreds with early-onset CRC and microsatellite stable (MSS) tumors, suggesting that there are as yet unidentified single gene defects that account for familial clustering of CRC (Lindor et al., 2005; Lipkin and Afrasiabi, 2007). Genome linkage studies scans support this notion and have identified multiple regions of interest, including chromosome 9q21.3 (CRCS1, OMIM 608812), chromosome 8q24 (CRCS2, OMIM 611469), chromosome 15 (HCRA/CRAC1, OMIM 610069), and a variety of other loci (Daley et al., 2008; Jaeger et al., 2003; Haiman et al., 2007a; Tomlinson et al., 2007; Wiesner et al., 2003). The analysis of the HCRA/CRAC1 region is instructive of the improvements in genomic technology in gene discovery, as kindreds with oligopolyposis were first linked and verified to this region using microsatellite markers (Daley et al., 2008; Jaeger et al., 2003). The region was next examined with single nucleotide polymorphism (SNP) variant analysis in CRC cases and controls and found statistical evidence for variants near GREM1 and SCG5 with a P value of 4.44×10^{-14} (Jaeger et al., 2008).

Rather than a discrete autosomal dominant or recessive condition, most familial forms of CRC are complex genetic traits involving the interaction of many variables. Risk assessment is based on epidemiologic studies that show a two- to four-fold increase in risk for family members, depending on the number of close relatives affected with CRC and the age at onset of disease. Thus, for a person who has a single parent with CRC, the lifetime risk doubles that of the general population (Sandhu et al., 2001). However, the risk is fourfold the general population risk for individuals with three close relatives with CRC or if the family member developed cancer at an early age (Boardman et al., 2007; St John et al., 1993). Importantly, the National Polyp Study demonstrated that the risk for CRC also increases with a family history of adenomatous polyps, indicating that there are familial risk factors in the initiation of colon neoplasias (Winawer et al., 1996). Causes are likely to include undiscovered low penetrance or recessive alleles as well as undefined gene–gene interactions and gene–environmental interactions, among others. Future genome-wide association studies may identify links that can be then further studied.

Sporadic or Isolated CRC

Approximately 75% of all CRC occurs in individuals with no known genetic predisposing factors for disease and about 80% are caused by somatic defects in the CIN pathway as shown in Figure 36.2 (Gryfe, 2006). This set is often designated as the "general population" and includes a wide spectrum of individuals who may or may not have a combination of modifiable environmental factors and/or somatic genetic changes that have been shown to modify one's risk for CRC or affect prognosis and response to treatment. Epidemiologists have also studied many environmental risk factors to better understand risk of CRC, beyond reporting of a family history of CRC. The factors that have been shown to be associated with an increased risk of CRC include obesity (insulin resistance), limited physical inactivity, smoking, heavy alcohol use, a diet high in red or processed meat and low in fruits and vegetables (folate), fiber, and calcium/dairy foods, non-multivitamin use, non-steroidal anti-inflammatory drug use (NSAIDs), and non-post-menopausal hormone replacement therapy (Giovannucci and Wu, 2006; Potter and Hunter, 2002). In addition, other underlying colon inflammatory diseases, such as Crohns disease and ulcerative colitis, have been shown to have a significant increased risk of CRC development, although the lifetime estimates vary widely from 7% to 30% (Eaden et al., 2001; Ekbom et al., 1990). The overall risk of CRC development is similar for each of these inflammatory diseases (Gillen et al., 1994; Greenstein et al., 1981). After excluding inherited monogenic and familial forms of CRC, much of the remaining variation in genetic risk for CRC is most likely explained by combinations of common low-risk variants. This "common disease – common variant" hypothesis is believed to hold true for most human complex diseases, including cancer. Several researchers have suggested that studying SNPs for association with CRC risk should be a powerful research strategy and several candidate gene and whole genome linkage scans have been performed seeking to identify these putative common variants (Botstein and Risch, 2003).

Historically, genetic risk factor studies have focused on single variants in single genes involved in metabolism of heterocyclic amines from dietary factors and folate metabolism pathways (Potter and Hunter, 2002). These candidate genes have been shown to be associated with risk of CRC when considered together with certain environmental exposures (Potter, 1999). Polymorphisms in three relevant enzymes for metabolism of heterocyclic amines and polycyclic aromatic hydrocarbons, *N*-acetyltransferases (NAT1; OMIM 108345 and NAT2: OMIM

243400) and cytochrome-P1A2 (CYP1A2; OMIM 124606), have been studied extensively for their role in risk of CRC associated with meat consumption, cigarette smoking and alcohol use with inconsistent results (Brockton et al., 2000; Chen et al., 1998; de Jong et al., 2002; Potter et al., 1999; Vineis and McMichael, 1996; Ye and Parry, 2002).

One of the possible nutrients responsible for the association between vegetables and multivitamin intake and reduced CRC risk is folic acid (Giovannucci, 2002a). Therefore, the association between CRC risk, folate intake and variants in key folate metabolism genes, such as methylenetetrahydrofolate reduc-tase (MTHFR; OMIM 607093) and thymidylate synthase (TS; OMIM 188350), has been studied repeatedly (Chen et al., 1998; Curtin et al., 2004; Hubner et al., 2007; Slattery et al., 1999; Ulrich et al., 2002). The data from the studies of MTHFR polymorphisms show that individuals with the TT genotype for the polymorphism 677C → T are "hyper-responders" to folate and to alcohol. If the diet was high in folate and low in alcohol, then the patients were at a significantly lower risk of CRC compared to those individuals with the CC or CT genotype. Additionally, a growing body of evidence suggests that individuals with higher circulating levels of IGF-1 have an increased risk of CRC; the association between another IGF candidate, IGF-binding protein 3 (IGFBP3), and CRC risk is less consistent (Giovannucci, 2001; Giovannucci et al., 2000; Kaaks et al., 2000; Ma et al., 1999). These data suggest that common variants exist in these genes, altering predisposition to CRC in the general population. However, genomic technology does not yet have the capacity to test for all such variants simultaneously and to determine the exact level of statistical risk for an individual. Thus translating these variants into clinical risk algorithms must wait for the complete identification of variants and improved study designs to determine the attributable risk for each predisposing or protective factor.

With the invention of genotype microarrays and other high-throughput genotyping techniques, the capacity to genotype more than 500,000 SNPs on a single individual at one time is now available. Hence, many genome-wide association (GWA) studies of various complex diseases are utilizing this technology to assess genetic risk on a genome-wide global scale. CRC is no exception, where several case-control studies have now identified a region on the long arm of chromosome 8 (8q24) that may confer increased genetic risk for CRC (Broderick et al., 2007; Poynter et al., 2007; Tomlinson et al., 2007; Zanke et al., 2007). This region has also been implicated in prostate cancer susceptibility, supporting the notion that there is variability in the disease phenotype from defects in cancer pathways that overlap in tumor development (Gudmundsson et al., 2007; Haiman et al., 2007b; Yeager et al., 2007). There are multiple different interesting CRC candidate genes in this region that warrant further study. In addition, GWA studies have demonstrated that SNPs in the SMAD7 gene located on chromosome 18, which is part of the TGFβ/WNT signaling pathway, is associated with CRC (Broderick et al., 2007). Key to understanding common variants and risk for CRC will be the correlation of copy number variants and gene expression variants on sporadic CRC within the context of specific cellular pathways. As whole-genome approaches become more accessible and more affordable and bioinformatics techniques advance, large-scale, genome-wide studies of gene–gene and gene–environment interactions will be able to better delineate factors associated with CRC risk.

RISK ASSESSMENT, EVALUATION, AND GENETIC TESTING

Personalized medicine will rely on improved genomic and proteomic tests for diagnosis treatment, including new drugs and therapies. The practical application of genomic information in healthcare today is primarily in risk assessment and direct genetic testing. As knowledge of genetic susceptibility to cancers increases, researchers, healthcare providers, and the public are finding it increasingly difficult to incorporate recommendations for healthcare based on genomic discoveries (Acheson et al., 2000; Burke, 2005). Several studies have shown that physician knowledge about genetic and genomic concepts is limited, even though a family history of cancer is among the most commonly encountered genetic issue encountered in primary care (Acheson et al., 2000; Carroll et al., 2003; Giardiello et al., 1997). New tools are needed to assist clinicians to recognize at-risk individuals and families with hereditary cancer susceptibility, to tailor cancer screening and prevention and to screen for eligibility to participate in cancer genetic research (Acheson and Wiesner, 2004; Acheson et al., 2006). One such tool under development will categorize families by level of risk for cancer, including CRC, and will generate a report for patients and physicians with screening and genetic testing recommendations (Acheson et al., 2006). Computerized programs have been developed for breast cancer and are used to assist the clinician in evaluating the individual's risk and in considering genetic testing (Berry et al., 2002; Evans et al., 2004; Kelly and Sweet, 2007; Lindor et al., 2007). Risk algorithms similar to BRCAPRO called MMRPRO are based on known risk factors for colon cancer of family history and other risk factors (Chen et al., 2006).

Genetic testing for FAP and HNPCC are conditions in which genomic technology has been successfully translated into clinical care. Various guidelines for genetic testing for colon cancer susceptibility have been suggested by professional organizations and experts in the field, with the consistent recommendation that genetic counseling and evaluation should be provided to patients considering genetic testing (ASCO, 2003; Levin et al., 2006; Lynch et al., 2007; NCCN, 2007). Genetic evaluation includes a detailed medical history including the cancer and pre-neoplastic diagnoses, the construction of a three to four generation family tree, physical examination, risk assessment, and consideration of genetic testing, including a full discussion of possible genetic discrimination. Genetic testing is generally warranted if the personal or family history is suggestive of an inherited monogenic syndrome, the testing can be adequately interpreted and/or the results can be used to guide medical and

surgical management for the patient and other family members (ASCO, 2003). One of the most difficult issues in genetic testing for cancer is in the interpretation of results, as adequate studies to determine the clinical utility of the test are often not available, due to the rarity of the individual disorders and the lack of power in the statistical analyses needed to determine sensitivity and specificity of the tests (Burke et al., 2002). Further, most tests based on medical sequencing can also identify variants with indeterminate clinical significance and, thus, may not provide a diagnosis or further clarification of risk for the individual or family. Clinicians must also recognize that locus and allelic genetic heterogeneity will alter the final interpretation, as a negative result may be a reflection of incomplete testing.

Risk assessment of the family medical tree is the first step in deciding what test to offer a patient of family who may be at risk for colorectal neoplasia (Figure 36.3). The family tree should be examined for evidence of autosomal dominant inheritance and any occurrences of early-onset CRC or polyposis to determine whether the family member represents a polyposis syndrome or nonpolyposis syndrome. Verifying the pathology of the reported CRC and colon polyps and other cancers in the family is often recommended (Box 36.1 and Table 36.1). If the individual has multiple adenomatous polyps and a suggestive family history, genetic testing for APC mutations is warranted, However, the clinician should recall that up to 20–30% of FAP patients are affected due to spontaneous *de novo* APC mutations in the germline, without a family history of the disorder in either parent (Burt and Neklason, 2005). Thus, genetic testing of the APC gene is important for any patient with multiple adenomatous polyps at an early age regardless of family history (NCCN, 2007). Additionally, FAP is one disorder in which the testing of children may be warranted because screening would commence by age 10, if a pathologic mutation is identified (ASCO, 2003).

The evaluation of HNPCC begins with analysis of the family tree using the Amsterdam I, Amsterdam II, and Bethesda criteria (see Box 36.1). HNPCC testing is first done with tumor tissue screening for loss of immunohistochemical (IHC) staining

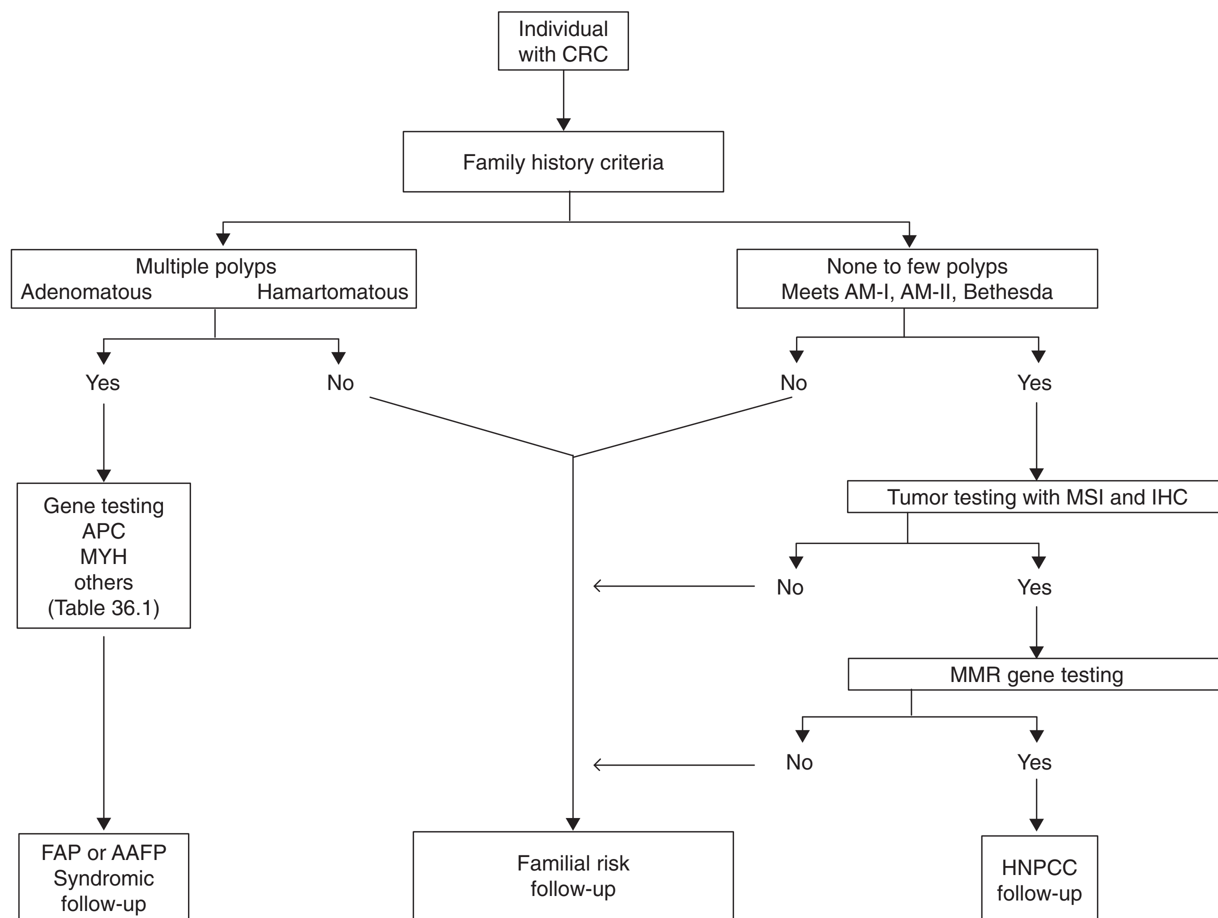


Figure 36.3 Clinical decision tree for the systematic evaluation of suspected inherited CRC syndromes. Evaluation begins with a detailed family history of CRC and/or polyps for each patient and the progresses to either the appropriate long term follow-up and observation or genetic testing depending on the family history and polyp status of the patient.

for proteins encoded by hMSH2, hMLH1, hMSH6, hPMS1, or hPMS2, and DNA polymerase chain reaction assay for MSI (NCCN, 2007). This strategy is based on the observation that a high level of MSI (MSI-H) in HNPCC tumors are also accompanied by loss of the MMR protein that corresponds to the gene mutation. The Bethesda guidelines recommend a standard set of mono- and di-nucleotide tracks that are particularly sensitive to MMR deficiency, with MSI-H defined as instability in 30% or more markers (Rodriguez-Bigas et al., 1997; Umar et al., 2004). As acquired BRAF mutations are highly associated with sporadic MSI-H tumors, testing tumor DNA for BRAF V600E mutation will help to categorize the patient (Minoo et al., 2007). If the screening tests are indicative of an inherited MMR deficiency, direct sequencing of MMR genes is next performed looking for germline mutational error in the family (Hendriks et al., 2006; Kwak and Chung, 2007).

SCREENING AND SURVEILLANCE

Many studies have proven that screening for CRC with fecal occult stool testing and colonoscopy methods has reduced the incidence, morbidity, and mortality of the disease (Burt, 1996b; USPSTF, 2002; Winawer et al., 2003). Despite this, less than 35% of adults follow screening guidelines, and there is great interest in genomic screening tools that may improve compliance and or clinical utility for identifying CRC in average risk and high-risk populations (CDC, 1996). Standard recommendations for average risk individuals have been published by the American Cancer Society (ACS), and other professional organizations, such as the National Comprehensive Cancer Network (NCCN) (ACS, 1999; NCCN, 2007; Winawer et al., 2006). In general these recommendations suggest colonoscopic screening starting at age 50, and that occult blood testing of the stool with left sided endoscopic screening or double contrast barium enema be performed if colonoscopy is not available (Davila et al., 2005; Smith et al., 2001). However, if the individual has a modest family history and no other risk factors, colonoscopy screening should commence at least a decade earlier at age 40 (Fletcher et al., 2007).

Specific guidelines have been developed for FAP and HNPCC with screening and surveillance recommendations starting at a younger age for specific tumor manifestations of the condition (Lindor et al., 2006; NCCN, 2007). In general, at risk family members for FAP and HNPCC begin a screening and surveillance program at 10 and 20 years, respectively (Lindor et al., 2006; Rustgi, 2007). While a review of these guidelines is beyond the scope of this chapter, the importance of identification and screening cannot be overstated. One study demonstrated a 70% reduction in the standardized mortality rate in 140 HNPCC families enrolled in a large scale surveillance program in the Netherlands (de Jong et al., 2006).

Newer non-invasive genetic technologies, such as stool and serum testing, have been developed to reduce the need for invasive colonoscopic screening and/or increase the sensitivity of the examination (Hundt et al., 2007; Levin, 2006). Several studies have shown that fecal material retains the cellular remnants of the colonic epithelium and that the nuclear DNA from these cells remains intact. Molecular analysis of genes known to be altered in CRC progression were first examined in patients with CRC finding that three genes, TP53, BAT26, and K-Ras, accurately identified such patients (Dong et al., 2001). Multiplex tests for specific genes and methylation signatures are now available commercially; one early study examined 21 gene mutations in 80 cases compared to 212 controls and identified 63% of CRC cases and 57% of advanced adenomas, with a calculated specificity of 94% in individuals without colon neoplasia or with small adenomas (Tagore et al., 2003). Aberrant methylation of three genes in stool DNA found nearly 94% sensitivity but only 77% specificity in a small sample of 52 patients with CRC, 35 patients with benign colon disease (21 with adenomas), and 24 patients with negative colonoscopies (Huang et al., 2007).

The identification of potential serum biomarkers for CRC diagnosis has also received some attention, given the ease of collecting serum as compared to other invasive diagnostic techniques for CRC. Newer approaches, mostly mass spectrometry (MS) based surface-enhanced laser desorption/ionizing time-of flight (SELDI-TOF) or matrix-assisted laser desorption/ionizing time-of flight (MALDI-TOF) methodologies, can produce high-throughput proteomic "signatures" from a variety of biospecimens. The ability of the measured protein intensities to distinguish cancer from non-cancerous biospecimens is assessed with multivariate statistical analyses. A recent proteomic study of CRC by Ward and colleagues (2006) identified a combination of four proteins that were significantly different between CRC cases and controls and then used this signature to classify serum samples from 62 CRC cases and 31 controls with a 95% sensitivity and 91% specificity. However, further validation of these proteomic techniques in sera of CRC cases compared to controls is necessary in order to advance CRC personalized medicine.

PROGNOSIS AND TREATMENT

Prognosis and treatment preferences after CRC diagnosis are highly dependent on stage of disease. The early stages of CRC (stage II or less) are treated primarily with surgery with or without chemotherapy and can have a 5-year survival >90% (Ries et al., 2007). However, the 5-year survival for late-stage disease is <10% and treatment can include a combination of surgery, chemotherapy, and/or radiotherapy. Additional significant independent predictors of survival are age at diagnosis, race/ethnicity, and family history of CRC. Whether a colon tumor develops along the CIN or MSI pathway may also have therapeutic implications. Sporadic MSI tumors have a significantly improved prognosis compared to MSS tumors (Bettstetter et al., 2007; Lim et al., 2004; Popat et al., 2005; Samowitz et al., 2001). Although the results are inconsistent, patients with stage II or III

MSI tumors may not benefit from 5-Flouorouracil (5-FU) adjuvant chemotherapy, a standard agent used in CRC (Barratt et al., 2002; Jover et al., 2006; Ribic et al., 2003). In contrast, CIN CRCs have been observed to be associated with poor prognosis, but not as consistently as the improved prognosis finding for MSI tumors (Bardi et al., 2004; Risques et al., 2003; Vendrell et al., 2005). Resolution of these inconsistencies in clinical outcome may need to wait until genomic methodologies mature more fully.

As most sporadic colorectal tumors are MSS, progression is linked to abnormalities of chromosome number, allelic imbalance or LOH, which are independently associated with poor survival (Bardi et al., 2004; Lanza et al., 1998; Risques et al., 2003; Vendrell et al., 2005). Chromosomes 5q, 17p, and 18q show the most frequent LOH in sporadic CRC, therefore, studies of alterations in specific genes in these chromosomal locations have been studied for an association with prognosis with conflicting results (Munro et al., 2005; Popat and Houlston, 2005). A systematic review of TP53 abnormalities and CRC outcome from 241 studies involving ~19,000 individuals showed that the presence of a TP53 mutation conferred a 30% decrease in survival compared to those with no mutation (Munro et al., 2005). Another review of genetic changes and outcome focused on another common genetic event in CRC, 18q loss and/or loss of DCC expression, showing that individuals with loss of DCC expression have twice the risk of dying compared to individuals without these genetic changes (Popat and Houlston, 2005). MSI status, p53 mutation and DCC expression are commonly identified in CRC; however it is still not well understood whether the association seen with clinical outcome is indicative of an underlying biological pathway or a specific defective gene (Gryfe, 2006).

Promising research in breast cancer has recently shown the utility of gene-specific expression profiling and clinical outcome, and this technology has been rapidly accepted by clinicians to guide chemotherapy (Morris and Carey, 2007). Similar utility of whole-genome gene expression profiling in predicting prognosis after CRC diagnosis has also been sought (Barrier et al., 2007; Eschrich et al., 2005; Wang et al., 2004). Two studies, both using the Affymetrix Human U133A gene expression chips, have identified a discrete set of expressed genes that predicted survival with 78% accuracy (Barrier et al., 2007; Wang et al., 2004). However, there was little overlap in the genes included in these two predictive gene sets, and only Wang and colleagues validated their findings for prediction of recurrence. A third study, using gene expression information from spotted cDNA arrays, found an optimal gene set of 43 genes with 93% sensitivity and 84% specificity in predicting CRC outcome (Eschrich et al., 2005). Gene expression profiling has also been shown to be able to discriminate different histological types of CRC, different stages of CRC and CRC from control colon tissue (Joyce and Pintzas, 2007). Again, there was little to no overlap in expressed genes with the previously described gene sets, highlighting the significant challenges faced in developing this technology. Differences in microarray platforms, data analysis techniques and, in particular, the inclusion/exclusion criteria for study subjects could have contributed the observed differences.

Although gene expression profiling has yielded some interesting prognostic candidates, many other types of genetic changes could also be significantly associated with prognosis. Genome-wide analyses of copy number changes and methylation status have been used in other cancers to predict clinical outcomes and could also prove useful in CRC outcomes analysis. Additionally, the role of environmental risk factors in prognosis is poorly understood and needs further study particularly studying gene–environment interactions and how these might affect prognosis.

PHARMACOGENETICS/GENOMICS OF CHEMOPREVENTION AND CHEMOTHERAPY

Pharmacogenetics/genomics seeks to discover the genotype or genotypes that predict overall drug response in terms of clinical outcomes and toxicity and to tailor this information for patients who have been categorized to have good drug response and low toxicity to a particular treatment (Marsh, 2007). These discoveries would also help in identifying patients at risk for adverse clinical outcomes or toxicities and as a result treat them with a different agent. Agents have been developed as chemopreventative agents for individuals at risk for CRC as well as treatment therapies for patients diagnosed with CRC. The multiple mutational events in both sporadic and inherited forms of CRC (Table 36.1 and Figure 36.2) have been explored for predictive and prognostic potential with a few showing promising results (Allen and Johnston, 2005).

Some of the factors that are associated with CRC, such as diet, are modifiable and have therefore been targets for prevention and intervention studies or have been used as potential chemopreventative agents (Boursi and Arber, 2007; Giovannucci, 2002b). The most well-established agent in CRC chemoprevention is NSAID use, in the form of regular over-the-counter aspirin (Markowitz, 2007; Raju and Cruz-Correa, 2006). The molecular basis of NSAID chemoprevention is mostly due to the inhibition of cyclooxygenase (COX) enzymes, which have two isoforms COX-1 and COX-2, but NSAIDs may also inhibit CRC through non-COX mediated pathways (Brown and DuBois, 2005; Hawk and Levin, 2005). The strongest evidence for a protective effect of NSAID use on risk of CRC comes from two double-blind randomized trials, showing a 20–45% reduced risk of CRC depending on aspirin dose (Baron et al., 2003; Sandler et al., 2003). Sulindac, a non-steroidal agent, has also been used in chemoprevention trials for high-risk patients with FAP, with modest success (Keller et al., 1999). Other chemopreventative agents that show promise are calcium carbonate, selenium and hormone replacement therapy; however, the molecular mechanisms of these agents are not well understood and evidence from randomized trials of these agents is inconsistent (Gryfe, 2006).

Systemic combination chemotherapy is now the mainstay of treatment for patients with advanced CRC, and includes the combination of 5-FU with other agents, such as irinotecan or oxaliplatin (Kelly and Goldberg, 2005). From a genomic standpoint, several genes have been identified that play a role in either predicting response or side effects to specific drugs used for individuals with CRC. Many of these genes are specific drug targets with known variants or genes and/or enzymes involved in drug metabolism pathways. Table 36.2 lists the drugs commonly used to treat CRC and their associated genetic target(s) (Kruzelock and Short, 2007; O'Dwyer et al., 2007).

The most well understood of these chemotherapy agents and its biological markers for drug response are 5-FU or capecitabine, which is metabolized to become 5-FU (Parker and Cheng, 1990). 5-FU is given intravenously, while capecitabine is dosed orally (Diasio, 2001). The four key enzymes that can influence the toxicity and/or efficacy of 5-FU (and capecitabine) are: thymidylate synthetase (TS), dihydropyrimidine dehydrogenase (DPD), thymidine phosphorylase (TP), and methylenetetrahydrofolate reductase (MTHFR). In general, low levels of TS, DPD, or TP have been shown to confer better drug response and hence better survival (Kruzelock and Short, 2007). Interestingly, when intratumoral gene expression levels of the combination of TS, DPD, and TP were analyzed, 92% of responders to 5-FU had very low levels of all three enzymes compared to non-responders, who had at least one enzyme expressed at above normal threshold levels (Salonga et al., 2000). Two specific SNP variants in MTHFR (677C → T and 1298A → C) are functionally linked to reduced activity of this enzyme; harboring the T allele of the 677 variant may be associated with increased response rates to 5-FU (Cohen et al., 2003; Etienne et al., 2004; Weisberg et al., 1998).

Newer chemotherapeutic medications for advanced CRC, such as the topoisomerase inhibitor, irinotecan, also appear to have functional variants that are associated with the development of serious side effects (Table 36.2). The UGT1A1*28 7/7 variant was significantly associated with the development of grade III/IV neutropenia in a study of 141 patients with advanced CRC treated with the combination chemotherapy FOLFIRI. In addition, patients who harbored both the TS 3′-UTR 6 + /6 and XRCC3-241 C/C genotypes had an increased risk of disease progression compared to similar staged non-carrier patients (Ruzzo et al., 2007).

Most of the current knowledge in CRC predicting response to treatment is based on gene expression transcriptional analysis (i.e., at the mRNA level only). However, it is the protein that plays the crucial active role in all cells in the body. Quantification of gene expression from mRNA analysis is not always directly correlated with measured protein level as there can be translational and post-translational modifications that determine the amount of cellular protein. Therefore, proteomic profiling has great potential to better understanding the relationship between protein levels and response to treatment (Nishizuka et al., 2003). Ma and colleagues used a well-described panel of 60 human cancer cell lines (NCI-60), which includes 7 CRC cell lines, to predict response to a wide array of anticancer drugs using reverse-phase protein lysate microarrays (Ma et al., 2006; Paweletz et al., 2001). They were able to

TABLE 36.2 Drugs commonly used to treat CRC and their associated genetic target(s)

Drug	Genetic targets			
5-FU and Capecitabine	Thymidylate synthetase (TS)	Dihydropyrimidine dehydrogenase (DPD)	Thymidine phosphorylase (TP)	Methylenetetrahydrofolate reductase (MTHFR)
Irinotecan	Uridine diphosphateglucuronosyltranferase 1A (UGT1A)	Cytochrome P450s (specifically the CYP3A family)	Carboxylesterases (CES)	Adenosine-triphosphate binding cassettes (ABC transporters)
Oxaliplatin	Excision repair cross-complementing 1 and 2 (ERCC1 and ERCC2)	Glutathione-S-transferase-P1 (GSTP1)		
Bevacizumab	Vascular endothelial growth factor (VEGF)			
Cetuximab	Epidermal growth factor receptor (EGFR)			
Panitumumab	Epidermal growth factor receptor (EGFR)	Transforming Growth Factor alpha (TGFα)		

identify eight protein markers that predicted response to 5-FU (CDH1, CDH2, KRT8, ERBB2, MSN, MVP, MAP2K1, and MGMT), where all but one of these was known to be involved in colon cancer development. Additionally the CRC cell lines clustered into two separate groups, which also differed from the other cell lines using these eight protein markers, showing the utility of this set to not only predict chemosensitivity to 5-FU, but to distinguish the CRC cell lines from other cancer types. Proteomic analysis shows tremendous promise not only as a diagnostic tool but also as a clinical outcome prediction tool.

While these studies hold the promise of profiling patients to tailor the selection of drug treatment, there is a lack of consensus among cancer specialists as to the use of pharmacogenomic profiling for patients with CRC. The American Society of Clinical Oncology guidelines currently support genetic testing for variants in TS, DPD, and TP genes for prognostic information, but not for selection of chemotherapeutic agents, such as 5-FU (Locker et al., 2006). Validation of additional predictive markers of response will require randomized clinical trials by tumor markers status, which would require large sample sizes since the prevalence of some of these markers is low in the population and the individual markers may have small effects individually but larger overall effects in combination with other markers (Sargent et al., 2005).

2009 UPDATE

Several studies have examined the clinical utility of whole-genome technologies for the early detection of colon cancer. Assays have been developed to detect somatic gene mutations, epigenetic modifications of specific genes, and gene-specific expression in fecal cells and serum; however, it is not clear which method, if any, will ultimately improve standard clinical screening. Echoing an earlier prospective study (Imperiale et al., 2004), a large multisite study of 4482 average risk patients demonstrated that fecal DNA analysis is superior to standard fecal occult blood testing in identifying clinically relevant advanced adenomas or colon cancers (Ahlquist et al., 2008). This study also compared the first generation 23 marker DNA assay to a smaller 3 marker DNA assay, which includes hypermethylation of the vimentin gene and showed that the smaller test had a twofold increased in sensitivity (40% versus 20%) (Ahlquist et al., 2008). Hypermethylation of the vimentin gene with rigorous assessment of DNA integrity in the stool sample may be important elements of the fecal assay, as a study of 82 colon cancer cases and 363 controls demonstrated a sensitivity and specificity of 83% and 82%, respectively, for colon cancer when only these two markers were used (Chen et al., 2005; Itzkowitz et al., 2008). The search for noninvasive, early detection methods will continue, as more sensitive and specific methods for analyzing DNA from fecal and other tissues are currently under development (Diehl et al., 2008; Koga et al., 2008).

Analysis of microRNAs (miRNA), a family of short non-coding small RNAs, has also emerged as an area that may hold significant diagnostic biomarker and therapeutic target potential in cancer research (Zhang et al., 2007). miRNA microarray expression profiling of paired CRC tissue and normal colon tissue from 84 CRC patients was recently tested for an association with survival (Schetter et al., 2008). This study found that 37 miRNAs were differentially expressed in the CRC tissue and that one of these, miRNA-21, was associated with a 2.5-fold increased risk of death in both the initial set of 84 CRC tissue/normal pairs and a validation set of 113 separate individuals with CRC, independent of other covariates, such as stage of disease, age at diagnosis, gender, race, and tumor location. Interestingly, miRNA-21 is expressed at high levels in most cancerous solid tumors and has been shown to act an as antiapoptotic factor (Volinia et al., 2006). Profiling of miRNAs in colon cancer tissue may also provide prognostic information for patients. Seventeen of 315 miRNAs profiled in a set of 49 stage II colon cancers and 10 normal mucosal tissues were found to correlate with metastasis in microsatellite stable cancers with a sensitivity of 77% and specificity of 83% (Schepeler et al., 2008). Differences between primary tumor and metastatic colon cancers may exist, as differential expression of 80 genes were identified using oligonucleotide arrays of 12 sets of normal, colon cancer, and metastasis from a single individual (Koh et al., 2008). Importantly, these types of studies may pave the way for chemotherapy decision strategies much like current breast cancer care based on individual tumor expression profiles (Munoz et al., 2008).

The search for genetic factors that increase colon neoplasia continues with whole-genome SNP association studies (Broderick et al., 2007; Joyce and Pintzas, 2007; Tomlinson et al., 2007; Zanke et al., 2007). Tenesa and colleagues identified a cancer susceptibility locus at 11q24 and confirmed two previously identified loci, 8q24 and 18q21, with a genome-wide SNP analysis, first analyzing 1012 early-onset colon cancer cases and controls and then testing the strongest associations in nearly 14,000 cases and controls (Tenesa et al., 2008). This population was then combined with other collections in a metaanalysis based on over 20,000 cases and 20,000 controls, which identified four previously unreported loci at 14q22.2 near the BMP4 gene, at 16q22.1 near the CDH1 gene, at 19q13.1 near the RHPN2 gene, and at 20p12.3 (Houlston et al., 2008). Further study will be required to verify these putative susceptibility loci, as linkage has not consistently been verified in different study populations, such as the previous linkage to 3q22 and chromosome 9q (Kemp et al., 2006; Papaemmanuil et al., 2008; Wiesner et al., 2003). Comparison

between genome-wide association studies can be difficult due to a variety of differences between studies, such as eligibility criteria, sample collection and storage, incomplete correlative demographic and clinical information, nucleic acid extraction methods and quantification, diversity of microarrays tools used and hybridization techniques, and small sample sizes. Future studies will most likely focus on larger populations with standardized phenotypic criteria.

2009 UPDATE REFERENCES

Ahlquist, D.A., Sargent, D.J. et al. (2008). Stool DNA and occult blood testing for screen detection of colorectal neoplasia. *Ann Intern Med* 149(7), 441–450, W81.

Broderick, P., Carvajal-Carmona, L. et al. (2007). A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat Genet* 39(11), 1315–1317.

Chen, W.D., Han, Z.J. et al. (2005). Detection in fecal DNA of colon cancer-specific methylation of the nonexpressed vimentin gene. *J Natl Cancer Inst* 97(15), 1124–1132.

Diehl, F., Schmidt, K. et al. (2008). Analysis of mutations in DNA isolated from plasma and stool of colorectal cancer patients. *Gastroenterology* 135(2), 489–498.

Houlston, R.S., Webb, E. et al. (2008). Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat Genet* 40(12), 1426–1435.

Imperiale, T.F., Ransohoff, D.F. et al. (2004). Fecal DNA versus fecal occult blood for colorectal-cancer screening in an average-risk population. *N Engl J Med* 351(26), 2704–2714.

Itzkowitz, S., Brand, R. et al. (2008). A simplified, noninvasive stool DNA test for colorectal cancer detection. *Am J Gastroenterol* 103(11), 2862–2870.

Joyce, T. and Pintzas, A. (2007). Microarray analysis to reveal genes involved in colon carcinogenesis. *Expert Opin Pharmacother* 8(7), 895–900.

Kemp, Z., Carvajal-Carmona, L. et al. (2006). Evidence for a colorectal cancer susceptibility locus on chromosome 3q21-q24 from a high-density SNP genome-wide linkage scan. *Hum Mol Genet* 15(19), 2903–2910.

Koga, Y., Yasunaga, M. et al. (2008). Detection of colorectal cancer cells from feces using quantitative real-time RT-PCR for colorectal cancer diagnosis. *Cancer Sci* 99(10), 1977–1983.

Koh, K.H., Rhee, H. et al. (2008). Differential gene expression profiles of metastases in paired primary and metastatic colorectal carcinomas. *Oncology* 75(1–2), 92–101.

Munoz, M., Estevez, L.G. et al. (2008). Evaluation of international treatment guidelines and prognostic tests for the treatment of early breast cancer. *Cancer Treat Rev* 34(8), 701–709.

Papaemmanuil, E., Carvajal-Carmona, L. et al. (2008). Deciphering the genetics of hereditary non-syndromic colorectal cancer. *Eur J Hum Genet* 16(12), 1477–1486.

Schepeler, T., Reinert, J.T. et al. (2008). Diagnostic and prognostic microRNAs in stage II colon cancer. *Cancer Res* 68(15), 6416–6424.

Schetter, A.J., Leung, S.Y. et al. (2008). MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma. *JAMA* 299(4), 425–436.

Tenesa, A., Farrington, S.M. et al. (2008). Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet* 40(5), 631–637.

Tomlinson, I., Webb, E. et al. (2007). A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet* 39(8), 984–988.

Volinia, S., Calin, G.A. et al. (2006). A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc Natl Acad Sci USA* 103(7), 2257–2261.

Wiesner, G.L., Daley, D. et al. (2003). A subset of familial colorectal neoplasia kindreds linked to chromosome 9q22.2-31.2. *Proc Natl Acad Sci USA* 100(22), 12961–12965.

Zanke, B.W., Greenwood, C.M. et al. (2007). Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet* 39(8), 989–994.

Zhang, W., Dahlberg, J.E. et al. (2007). MicroRNAs in tumorigenesis: A primer. *Am J Pathol* 171(3), 728–738.

REFERENCES

Acheson, L.S. and Wiesner, G.L. (2004). Current and future applications of genetics in primary care medicine. *Prim Care* 31(3), 449–460, vii.

Acheson, L.S., Wiesner, G.L. et al. (2000). Family history-taking in community family practice: Implications for genetic screening. *Genet Med* 2(3), 180–185.

Acheson, L.S., Zyzanski, S.J. et al. (2006). Validation of a self-administered, computerized tool for collecting and displaying the family history of cancer. *J Clin Oncol* 24(34), 5395–5402.

Al-Tassan, N., Chmiel, N.H. et al. (2002). Inherited variants of MYH associated with somatic G:C → T:A mutations in colorectal tumors. *Nat Genet* 30(2), 227–232.

Albuquerque, C., Breukel, C. et al. (2002). The 'just-right' signaling model: APC somatic mutations are selected based on a specific level of activation of the beta-catenin signaling cascade. *Hum Mol Genet* 11(13), 1549–1560.

Allen, W.L. and Johnston, P.G. (2005). Role of genomic markers in colorectal cancer treatment. *J Clin Oncol* 23(20), 4545–4552.

American Cancer Society (1999). Colorectal cancer screening. ACS guidelines. *CA Cancer J Clin* 49(1), 4.

American Cancer Society (2007). Cancer Facts and Figures, 2007 (http://www.cancer.org/docroot/STT/content/STT_1x_Cancer_Facts__Figures_2007.asp), American Cancer Society, Inc. Atlanta, GA.

American Society of Clinical Oncology (2003). American Society of Clinical Oncology policy statement update: Genetic testing for cancer susceptibility. *J Clin Oncol* 21(12), 2397–2406.

Balaguer, F., Castellvi-Bel, S. et al. (2007). Identification of MYH mutation carriers in colorectal cancer: A multicenter, case-control, population-based study. *Clin Gastroenterol Hepatol* 5(3), 379–387.

Bardi, G., Fenger, C. et al. (2004). Tumor karyotype predicts clinical outcome in colorectal cancer patients. *J Clin Oncol* 22(13), 2623–2634.

Baron, J.A., Cole, B.F. et al. (2003). A randomized trial of aspirin to prevent colorectal adenomas. *N Engl J Med* 348(10), 891–899.

Barratt, P.L., Seymour, M.T. et al. (2002). DNA markers predicting benefit from adjuvant fluorouracil in patients with colon cancer: A molecular study. *Lancet* 360(9343), 1381–1391.

Barrier, A., Roser, F. et al. (2007). Prognosis of stage II colon cancer by non-neoplastic mucosa gene expression profiling. *Oncogene* 26(18), 2642–2648.

Berry, D.A., Iversen, E.S., Jr. et al. (2002). BRCAPRO validation, sensitivity of genetic testing of BRCA1/BRCA2, and prevalence of other breast cancer susceptibility genes. *J Clin Oncol* 20(11), 2701–2712.

Bettstetter, M., Dechant, S. et al. (2007). Distinction of hereditary nonpolyposis colorectal cancer and sporadic microsatellite-unstable colorectal cancer through quantification of MLH1 methylation by real-time PCR. *Clin Cancer Res* 13(11), 3221–3228.

Bisgaard, M.L. and Bulow, S. (2006). Familial adenomatous polyposis (FAP): Genotype correlation to FAP phenotype with osteomas and sebaceous cysts. *Am J Med Genet A* 140(3), 200–204.

Boardman, L.A., Morlan, B.W. et al. (2007). Colorectal cancer risks in relatives of young-onset cases: Is risk the same across all first-degree relatives?. *Clin Gastroenterol Hepatol* 5(10), 1195–1198.

Bodmer, W.F., Bailey, C.J. et al. (1987). Localization of the gene for familial adenomatous polyposis on chromosome 5. *Nature* 328(6131), 614–616.

Boland, C.R., Koi, M. et al. (2007). The biochemical basis of microsatellite instability and abnormal immunohistochemistry and clinical behavior in Lynch Syndrome: From bench to bedside. *Fam Cancer*.

Botstein, D. and Risch, N. (2003). Discovering genotypes underlying human phenotypes: Past successes for mendelian disease, future approaches for complex disease. *Nat Genet* 33(Suppl), 228–237.

Boursi, B. and Arber, N. (2007). Current and future clinical strategies in colon cancer prevention and the emerging role of chemoprevention. *Curr Pharm Des* 13(22), 2274–2282.

Brockton, N., Little, J. et al. (2000). N-acetyltransferase polymorphisms and colorectal cancer: A HuGE review. *Am J Epidemiol* 151(9), 846–861.

Broderick, P., Carvajal-Carmona, L. et al. (2007). A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat Genet* 39(11), 1315–1317.

Brown, J.R. and DuBois, R.N. (2005). COX-2: A molecular target for colorectal cancer prevention. *J Clin Oncol* 23(12), 2840–2855.

Burke, W. (2005). Taking family history seriously. *Ann Intern Med* 143(5), 388–389.

Burke, W., Atkins, D. et al. (2002). Genetic test evaluation: information needs of clinicians, policy makers, and the public. *Am J Epidemiol* 156(4), 311–318.

Burt, R. and Neklason, D.W. (2005). Genetic testing for inherited colon cancer. *Gastroenterology* 128(6), 696–716.

Burt, R.W. (1996a). Familial risk and colon cancer. *Int J Cancer* 69(1), 44–46.

Burt, R.W. (1996b). Familial risk and colorectal cancer. *Gastroenterol Clin North Am* 25(4), 793–803.

Burt, R.W. (2000). Colon cancer screening. *Gastroenterology* 119(3), 837–853.

Carpten, J.D., Faber, A.L. et al. (2007). A transforming mutation in the pleckstrin homology domain of AKT1 in cancer. *Nature* 448(7152), 439–444.

Carroll, J.C., Brown, J.B. et al. (2003). Genetic susceptibility to cancer. Family physicians' experience. *Can Fam Physician* 49, 45–52.

Center for Disease Control and Prevention (1996). Screening for colorectal cancer – United States, 1992–1993, and new guidelines. *MMWR Morb Mortal Wkly Rep* 45(5), 107–110.

Cheadle, J.P. and Sampson, J.R. (2003). Exposing the MYtH about base excision repair and human inherited disease. *Hum Mol Genet* 12(Spec No 2), R159–R165.

Chen, J., Giovannucci, E. et al. (1998). A prospective study of methylenetetrahydrofolate reductase and methionine synthase gene polymorphisms, and risk of colorectal adenoma. *Carcinogenesis* 19(12), 2129–2132.

Chen, J., Stampfer, M.J. et al. (1998). A prospective study of N-acetyltransferase genotype, red meat intake, and risk of colorectal cancer. *Cancer Res* 58(15), 3307–3311.

Chen, S., Wang, W. et al. (2006). Prediction of germline mutations and cancer risk in the Lynch syndrome. *JAMA* 296(12), 1479–1487.

Chung, D.C. (2000). The genetic basis of colorectal cancer: insights into critical pathways of tumorigenesis. *Gastroenterology* 119(3), 854–865.

Cohen, V., Panet-Raymond, V. et al. (2003). Methylenetetrahydrofolate reductase polymorphism in advanced colorectal cancer: A novel genomic predictor of clinical response to fluoropyrimidine-based chemotherapy. *Clin Cancer Res* 9(5), 1611–1615.

Croitoru, M.E., Cleary, S.P. et al. (2004). Association between biallelic and monoallelic germline MYH gene mutations and colorectal cancer risk. *J Natl Cancer Inst* 96(21), 1631–1634.

Curado, M.P., Edwards, B. et al. (2007). *Cancer Incidence in Five Continents, Vol. IX (No. 160)*. IARC Scientific Publications, Lyon, France.

Curtin, K., Bigler, J. et al. (2004). MTHFR C677T and A1298C polymorphisms: Diet, estrogen, and risk of colon cancer. *Cancer Epidemiol Biomarkers Prev* 13(2), 285–292.

Daley, D., Lewis, S., et al. (2008). Identification of susceptibility genes for cancer in a genome-wide scan; Results from the colon neoplasia sibling study. *Am J Hum Genet* 82(3), 723–736.

Davila, R.E., Rajan, E. et al. (2005). ASGE guideline: The role of endoscopy in the diagnosis, staging, and management of colorectal cancer. *Gastrointest Endosc* 61(1), 1–7.

de Jong, A.E., Hendriks, Y.M. et al. (2006). Decrease in mortality in Lynch syndrome families because of surveillance. *Gastroenterology* 130(3), 665–671.

de Jong, M.M., Nolte, I.M. et al. (2002). Low-penetrance genes and their involvement in colorectal cancer susceptibility. *Cancer Epidemiol Biomarkers Prev* 11(11), 1332–1352.

de la Chapelle, A. (2004). Genetic predisposition to colorectal cancer. *Nat Rev Cancer* 4(10), 769–780.

de la Chapelle (2005). The incidence of Lynch Syndrome. *Fam Cancer* 4, 233–237.

Deng, G., Bell, I. et al. (2004). BRAF mutation is frequently present in sporadic colorectal cancer with methylated hMLH1, but not in hereditary nonpolyposis colorectal cancer. *Clin Cancer Res* 10(1 Pt 1), 191–195.

Diasio, R.B. (2001). Clinical implications of dihydropyrimidine dehydrogenase on 5-FU pharmacology. *Oncology (Williston Park)* 15(1 Suppl 2), 21–26, discussion 27.

Dong, S.M., Traverso, G. et al. (2001). Detecting colorectal cancer in stool with the use of multiple genetic targets. *J Natl Cancer Inst* 93(11), 858–865.

Doxey, B.W., Kuwada, S.K. et al. (2005). Inherited polyposis syndromes: Molecular mechanisms, clinicopathology, and genetic testing. *Clin Gastroenterol Hepatol* 3(7), 633–641.

Eaden, J.A., Abrams, K.R. et al. (2001). The risk of colorectal cancer in ulcerative colitis: A meta-analysis. *Gut* 48(4), 526–535.

Ekbom, A., Helmick, C. et al. (1990). Increased risk of large-bowel cancer in Crohn's disease with colonic involvement. *Lancet* 336(8711), 357–359.

Eschrich, S., Yang, I. et al. (2005). Molecular staging for survival prediction of colorectal cancer patients. *J Clin Oncol* 23(15), 3526–3535.

Etienne, M.C., Formento, J.L. et al. (2004). Methylenetetrahydrofolate reductase gene polymorphisms and response to fluorouracil-based treatment in advanced colorectal cancer patients. *Pharmacogenetics* 14(12), 785–792.

Evans, D.G., Eccles, D.M. et al. (2004). A new scoring system for the chances of identifying a BRCA1/2 mutation outperforms existing models including BRCAPRO. *J Med Genet* 41(6), 474–480.

Fearon, E.R. and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell* 61(5), 759–767.

Fletcher, R.H., Lobb, R. et al. (2007). Screening patients with a family history of colorectal cancer. *J Gen Intern Med* 22(4), 508–513.

Giardiello, F.M., Brensinger, J.D. et al. (1997). The use and interpretation of commercial APC gene testing for familial adenomatous polyposis. *N Engl J Med* 336(12), 823–827.

Gillen, C.D., Walmsley, R.S. et al. (1994). Ulcerative colitis and Crohn's disease: A comparison of the colorectal cancer risk in extensive colitis. *Gut* 35(11), 1590–1592.

Giovannucci, E. (2001). Insulin, insulin-like growth factors and colon cancer: A review of the evidence. *J Nutr* 131(11 Suppl), 3109S–3120S.

Giovannucci, E. (2002a). Epidemiologic studies of folate and colorectal neoplasia: A review. *J Nutr* 132(8 Suppl), 2350S–2355S.

Giovannucci, E. (2002b). Modifiable risk factors for colon cancer. *Gastroenterol Clin North Am* 31(4), 925–943.

Giovannucci, E., Wu, K. (2006). Cancers of the colon and rectum. In *Cancer Epidemiology and Prevention, 3rd edition*, Oxford University Press, New York, pp. 809–829.

Giovannucci, E., Pollak, M.N. et al. (2000). A prospective study of plasma insulin-like growth factor-1 and binding protein-3 and risk of colorectal neoplasia in women. *Cancer Epidemiol Biomarkers Prev* 9(4), 345–349.

Goel, A., Nagasaka, T. et al. (2007). The CpG island methylator phenotype and chromosomal instability are inversely correlated in sporadic colorectal cancer. *Gastroenterology* 132(1), 127–138.

Greenstein, A.J., Sachar, D.B. et al. (1981). A comparison of cancer risk in Crohn's disease and ulcerative colitis. *Cancer* 48(12), 2742–2745.

Groden, J., Thliveris, A. et al. (1991). Identification and characterization of the familial adenomatous polyposis coli gene. *Cell* 66(3), 589–600.

Gryfe, R. (2006). Clinical implications of our advancing knowledge of colorectal cancer genetics: Inherited syndromes, prognosis, prevention, screening and therapeutics. *Surg Clin North Am* 86(4), 787–817.

Gudmundsson, J., Sulem, P. et al. (2007). Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet* 39(5), 631–637.

Haiman, C.A., Patterson, N. et al. (2007a). Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet* 39(5), 638–644.

Haiman, C.A., Le Marchand, L. et al. (2007b). A common genetic risk factor for colorectal and prostate cancer. *Nat Genet* 39(8), 954–956.

Hampel, H., Stephens, J.A. et al. (2005). Cancer risk in hereditary nonpolyposis colorectal cancer syndrome: Later age of onset. *Gastroenterology* 129(2), 415–421.

Hampel, H., Frankel, W. et al. (2006). Screening for Lynch syndrome (hereditary nonpolyposis colorectal cancer) among endometrial cancer patients. *Cancer Res* 66(15), 7810–7817.

Hawk, E.T. and Levin, B. (2005). Colorectal cancer prevention. *J Clin Oncol* 23(2), 378–391.

Hendriks, Y.M., de Jong, A.E. et al. (2006). Diagnostic approach and management of Lynch syndrome (hereditary nonpolyposis colorectal carcinoma): A guide for clinicians. *CA Cancer J Clin* 56(4), 213–225.

Herrera, L., Kakati, S. et al. (1986). Gardner syndrome in a man with an interstitial deletion of 5q. *Am J Med Genet* 25(3), 473–476.

Huang, Z.H., Li, L.H. et al. (2007). Detection of aberrant methylation in fecal DNA as a molecular screening tool for colorectal cancer and precancerous lesions. *World J Gastroenterol* 13(6), 950–954.

Hubner, R.A., Liu, J.F. et al. (2007). Thymidylate synthase polymorphisms, folate and B-vitamin intake, and risk of colorectal adenoma. *Br J Cancer* 97(10), 1449–1456.

Hundt, S., Haug, U. et al. (2007). Blood markers for early detection of colorectal cancer: A systematic review. *Cancer Epidemiol Biomarkers Prev* 16(10), 1935–1953.

Ionov, Y., Peinado, M.A. et al. (1993). Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* 363(6429), 558–561.

Jaeger, E.E., Woodford-Richens, K.L. et al. (2003). An ancestral Ashkenazi haplotype at the HMPS/CRAC1 locus on 15q13-q14 is associated with hereditary mixed polyposis syndrome. *Am J Hum Genet* 72(5), 1261–1267.

Jaeger, E., Webb, E. et al. (2008). Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. *Nat Genet* 40(1), 26–28.

Jemal, A., Siegel, R. et al. (2007). Cancer statistics, 2007. *CA Cancer J Clin* 57(1), 43–66.

Jo, W.S. and Chung, D.C. (2005). Genetics of hereditary colorectal cancer. *Semin Oncol* 32(1), 11–23.

Jover, R., Zapater, P. et al. (2006). Mismatch repair status in the prediction of benefit from adjuvant fluorouracil chemotherapy in colorectal cancer. *Gut* 55(6), 848–855.

Joyce, T. and Pintzas, A. (2007). Microarray analysis to reveal genes involved in colon carcinogenesis. *Expert Opin Pharmacother* 8(7), 895–900.

Kaaks, R., Toniolo, P. et al. (2000). Serum C-peptide, insulin-like growth factor (IGF)-I, IGF-binding proteins, and colorectal cancer risk in women. *J Natl Cancer Inst* 92(19), 1592–1600.

Keller, J.J., Offerhaus, G.J. et al. (1999). Rectal epithelial apoptosis in familial adenomatous polyposis patients treated with sulindac. *Gut* 45(6), 822–828.

Kelly, H. and Goldberg, R.M. (2005). Systemic therapy for metastatic colorectal cancer: current options, current evidence. *J Clin Oncol* 23(20), 4553–4560.

Kelly, K.M. and Sweet, K. (2007). In search of a familial cancer risk assessment tool. *Clin Genet* 71(1), 76–83.

Kinzler, K.W. and Vogelstein, B. (1996). Lessons from hereditary colo-rectal cancer. *Cell* 87(2), 159–170.

Knudson, A.G. (2001). Two genetic hits (more or less) to cancer. *Nat Rev Cancer* 1(2), 157–162.

Kruzelock, R.P. and Short, W. (2007). Colorectal cancer therapeutics and the challenges of applied pharmacogenomics. *Curr Probl Cancer* 31(5), 315–366.

Kwak, E.L. and Chung, D.C. (2007). Hereditary colorectal cancer syndromes: An overview. *Clin Colorectal Cancer* 6(5), 340–344.

Lagerstedt Robinson, K., Liu, T. et al. (2007). Lynch syndrome (hereditary nonpolyposis colorectal cancer) diagnostics. *J Natl Cancer Inst* 99(4), 291–299.

Laken, S.J., Petersen, G.M. et al. (1997). Familial colorectal cancer in Ashkenazim due to a hypermutable tract in APC. *Nat Genet* 17(1), 79–83.

Laken, S.J., Papadopoulos, N. et al. (1999). Analysis of masked mutations in familial adenomatous polyposis. *Proc Natl Acad Sci USA* 96(5), 2322–2326.

Lanza, G., Gafa, R. et al. (1998). Prognostic significance of DNA ploidy in patients with stage II and stage III colon carcinoma: A prospective flow cytometric study. *Cancer* 82(1), 49–59.

Levin, B. (2006). Molecular screening testing for colorectal cancer. *Clin Cancer Res* 12(17), 5014–5017.

Levin, B., Barthel, J.S. et al. (2006). Colorectal cancer screening clinical practice guidelines. *J Natl Compr Canc Netw* 4(4), 384–420.

Lim, S.B., Jeong, S.Y. et al. (2004). Prognostic significance of microsatellite instability in sporadic colorectal cancer. *Int J Colorectal Dis* 19(6), 533–537.

Lindor, N.M., Petersen, G.M. et al. (2006). Recommendations for the care of individuals with an inherited predisposition to Lynch syndrome: A systematic review. *JAMA* 296(12), 1507–1517.

Lindor, N.M., Lindor, R.A. et al. (2007). Predicting BRCA1 and BRCA2 gene mutation carriers: Comparison of LAMBDA, BRCAPRO, Myriad II, and modified Couch models. *Fam Cancer* 6(4), 473–482.

Lindor, N.M., Rabe, K. et al. (2005). Lower cancer incidence in Amsterdam-I criteria families without mismatch repair deficiency: Familial colorectal cancer type X. *JAMA* 293(16), 1979–1985.

Lipkin, S.M. and Afrasiabi, K. (2007). Familial colorectal cancer syndrome X. *Semin Oncol* 34(5), 425–427.

Lipton, L. and Tomlinson, I. (2004). The multiple colorectal adenoma phenotype and MYH, a base excision repair gene. *Clin Gastroenterol Hepatol* 2(8), 633–638.

Locker, G.Y., Hamilton, S. et al. (2006). ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. *J Clin Oncol* 24(33), 5313–5327.

Lynch, H.T., Boland, C.R. et al. (2007). Who should be sent for genetic testing in hereditary colorectal cancer syndromes? *J Clin Oncol* 25(23), 3534–3542.

Lynch, H.T. and de la Chapelle, A. (1999). Genetic susceptibility to non-polyposis colorectal cancer. *J Med Genet* 36(11), 801–818.

Ma, J., Pollak, M.N. et al. (1999). Prospective study of colorectal cancer risk in men and plasma levels of insulin-like growth factor (IGF)-I and IGF-binding protein-3. *J Natl Cancer Inst* 91(7), 620–625.

Ma, Y., Ding, Z. et al. (2006). Predicting cancer drug response by proteomic profiling. *Clin Cancer Res* 12(15), 4583–4589.

Markowitz, S.D. (2007). Aspirin and colon cancer – targeting prevention? *N Engl J Med* 356(21), 2195–2198.

Markowitz, S., Wang, J. et al. (1995). Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. *Science* 268(5215), 1336–1338.

Marsh, S. (2007). Pharmacogenomics. *Ann Oncol* 18(Suppl 9), ix24–ix28.

Martin, E.S., Tonon, G. et al. (2007). Common and distinct genomic events in sporadic colorectal cancer and diverse cancer types. *Cancer Res* 67(22), 10736–10743.

Martinez, M.E., Maltzman, T. et al. (1999). Risk factors for Ki-ras protooncogene mutation in sporadic colorectal adenomas. *Cancer Res* 59(20), 5181–5185.

Messerini, L., Mori, S. et al. (1996). Pathologic features of hereditary non-polyposis colorectal cancer. *Tumori* 82(2), 114–116.

Minoo, P., Moyer, M.P. et al. (2007). Role of BRAF-V600E in the serrated pathway of colorectal tumourigenesis. *J Pathol* 212(2), 124–133.

Moisio, A.L., Jarvinen, H. et al. (2002). Genetic and clinical characterisation of familial adenomatous polyposis: A population based study. *Gut* 50(6), 845–850.

Morris, S.R. and Carey, L.A. (2007). Gene expression profiling in breast cancer. *Curr Opin Oncol* 19(6), 547–551.

Munro, A.J., Lain, S. et al. (2005). P53 abnormalities and outcomes in colorectal cancer: A systematic review. *Br J Cancer* 92(3), 434–444.

National Comprehensive Cancer Network. (2007). NCCN clinical practice guidelines in oncology: Colorectal cancer screening V.1.2007. Retrieved December 17, 2007, from http://www.nccn.org.

Nieuwenhuis, M.H. and Vasen, H.F. (2007). Correlations between mutation site in APC and phenotype of familial adenomatous polyposis (FAP): A review of the literature. *Crit Rev Oncol Hematol* 61(2), 153–161.

Nishizuka, S., Charboneau, L. et al. (2003). Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays. *Proc Natl Acad Sci USA* 100(24), 14229–14234.

Nosho, K., Yamamoto, H. et al. (2005). Gene expression profiling of colorectal adenomas and early invasive carcinomas by cDNA array analysis. *Br J Cancer* 92(7), 1193–1200.

O'Brien, M.J., Yang, S. et al. (2006). Comparison of microsatellite instability, CpG island methylation phenotype, BRAF and KRAS status in serrated polyps and traditional adenomas indicates separate pathways to distinct colorectal carcinoma end points. *Am J Surg Pathol* 30(12), 1491–1501.

O'Dwyer, P.J., Eckhardt, S.G. et al. (2007). Priorities in colorectal cancer research: Recommendations from the Gastrointestinal Scientific Leadership Council of the Coalition of Cancer Cooperative Groups. *J Clin Oncol* 25(16), 2313–2321.

Ogino, S., Kawasaki, T. et al. (2007). TGFBR2 mutation is correlated with CpG island methylator phenotype in microsatellite instability-high colorectal cancer. *Hum Pathol* 38(4), 614–620.

Parker, W.B. and Cheng, Y.C. (1990). Metabolism and mechanism of action of 5-fluorouracil. *Pharmacol Ther* 48(3), 381–395.

Paweletz, C.P., Charboneau, L. et al. (2001). Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. *Oncogene* 20(16), 1981–1989.

Popat, S. and Houlston, R.S. (2005). A systematic review and meta-analysis of the relationship between chromosome 18q genotype, DCC status and colorectal cancer prognosis. *Eur J Cancer* 41(14), 2060–2070.

Popat, S., Hubner, R. et al. (2005). Systematic review of microsatellite instability and colorectal cancer prognosis. *J Clin Oncol* 23(3), 609–618.

Potter, J.D. (1999). Colorectal cancer: molecules and populations. *J Natl Cancer Inst* 91(11), 916–932.

Potter, J.D. and Hunter, D. (2002). Colorectal cancer. In *Textbook of Cancer Epidemiology* (H.-O. Adami, D. Hunter and D. Trichopoulos, eds), Oxford University Press, Inc., New York, pp. 188–211.

Potter, J.D., Bigler, J. et al. (1999). Colorectal adenomatous and hyperplastic polyps: Smoking and N-acetyltransferase 2 polymorphisms. *Cancer Epidemiol Biomarkers Prev* 8(1), 69–75.

Powell, S.M., Zilz, N. et al. (1992). APC mutations occur early during colorectal tumorigenesis. *Nature* 359(6392), 235–237.

Poynter, J.N., Figueiredo, J.C. et al. (2007). Variants on 9p24 and 8q24 are associated with risk of colorectal cancer: Results from the Colon Cancer Family Registry. *Cancer Res* 67(23), 11128–11132.

Rajagopalan, H., Nowak, M.A. et al. (2003). The significance of unstable chromosomes in colorectal cancer. *Nat Rev Cancer* 3(9), 695–701.

Raju, R. and Cruz-Correa, M. (2006). Chemoprevention of colorectal cancer. *Dis Colon Rectum* 49(1), 113–124. discussion 124–5.

Ribic, C.M., Sargent, D.J. et al. (2003). Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer. *N Engl J Med* 349(3), 247–257.

Ried, T., Knutzen, R. et al. (1996). Comparative genomic hybridization reveals a specific pattern of chromosomal gains and losses during the genesis of colorectal tumors. *Genes Chromosomes Cancer* 15(4), 234–245.

Ries, L.A.G., Melbert, D., et al. (2007). SEER Cancer Statistics Review, 1975–2004, http://www.seer.cancer.gov/csr/1975_2004/, based on November 2006 SEER data submission, posted to the SEER web site, 2007. National Cancer Institute, Bethesda, MD.

Risques, R.A., Moreno, V. et al. (2003). Genetic pathways and genome-wide determinants of clinical outcome in colorectal cancer. *Cancer Res* 63(21), 7206–7214.

Rodriguez-Bigas, M.A., Boland, C.R. et al. (1997). A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome: Meeting highlights and Bethesda guidelines. *J Natl Cancer Inst* 89(23), 1758–1762.

Rowley, P.T. (2005). Inherited susceptibility to colorectal cancer. *Annu Rev Med* 56, 539–554.

Rustgi, A.K. (2007). The genetics of hereditary colon cancer. *Genes Dev* 21(20), 2525–2538.

Ruzzo, A., Graziano, F. et al. (2007). Pharmacogenetic profiling in patients with advanced colorectal cancer treated with first-line FOLFIRI chemotherapy. *Pharmacogenomics J*. Epub. doi: 10.1038.

Salonga, D., Danenberg, K.D. et al. (2000). Colorectal tumors responding to 5-fluorouracil have low gene expression levels of dihydropyrimidine dehydrogenase, thymidylate synthase, and thymidine phosphorylase. *Clin Cancer Res* 6(4), 1322–1327.

Samowitz, W.S., Curtin, K. et al. (2001). Microsatellite instability in sporadic colon cancer is associated with an improved prognosis at the population level. *Cancer Epidemiol Biomarkers Prev* 10(9), 917–923.

Samowitz, W.S., Albertsen, H. et al. (2005). Evaluation of a large, population-based sample supports a CpG island methylator phenotype in colon cancer. *Gastroenterology* 129(3), 837–845.

Sandhu, M.S., Luben, R. et al. (2001). Prevalence and family history of colorectal cancer: Implications for screening. *J Med Screen* 8(2), 69–72.

Sandler, R.S., Halabi, S. et al. (2003). A randomized trial of aspirin to prevent colorectal adenomas in patients with previous colorectal cancer. *N Engl J Med* 348(10), 883–890.

Sargent, D.J., Conley, B.A. et al. (2005). Clinical trial designs for predictive marker validation in cancer treatment trials. *J Clin Oncol* 23(9), 2020–2027.

Schreibman, I.R., Baker, M., Amos, C. and McGarrity T.J. (2005). The hamartomatous polyposis syndrome. *Am J Gastroenterology* 100, 476–490.

Schubbert, S., Shannon, K. et al. (2007). Hyperactive Ras in developmental disorders and cancer. *Nat Rev Cancer* 7(4), 295–308.

Sjoblom, T., Jones, S. et al. (2006). The consensus coding sequences of human breast and colorectal cancers. *Science* 314(5797), 268–274.

Slattery, M.L., Potter, J.D. et al. (1999). Methylenetetrahydrofolate reductase, diet, and risk of colon cancer. *Cancer Epidemiol Biomarkers Prev* 8(6), 513–518.

Smith, R.A., von Eschenbach, A.C. et al. (2001). American Cancer Society guidelines for the early detection of cancer: update of early detection guidelines for prostate, colorectal, and endometrial cancers Also: Update 2001 – testing for early lung cancer detection. *CA Cancer J Clin* 51(1), 38–75, quiz 77–80.

St John, D.J., McDermott, F.T. et al. (1993). Cancer risk in relatives of patients with common colorectal cancer. *Ann Intern Med* 118(10), 785–790.

Tagore, K.S., Lawson, M.J. et al. (2003). Sensitivity and specificity of a stool DNA multitarget assay panel for the detection of advanced colorectal neoplasia. *Clin Colorectal Cancer* 3(1), 47–53.

Thibodeau, S.N., Bren, G. et al. (1993). Microsatellite instability in cancer of the proximal colon. *Science* 260(5109), 816–819.

Tomlinson, I., Webb, E. et al. (2007). A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet* 39(8), 984–988.

Toyota, M., Ahuja, N. et al. (1999). CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci USA* 96(15), 8681–8686.

Ulrich, C.M., Bigler, J. et al. (2002). Thymidylate synthase promoter polymorphism, interaction with folate intake, and risk of colorectal adenomas. *Cancer Res* 62(12), 3361–3364.

Umar, A., Boland, C.R. et al. (2004). Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst* 96(4), 261–268.

United States Preventive Services Task Force (2002). Screening for colorectal cancer: recommendations and rationale, Agency for Healthcare Research and Quality; Rockville, MD, 2008.

Vasen, H.F. (2000). Clinical diagnosis and management of hereditary colorectal cancer syndromes. *J Clin Oncol* 18(21 Suppl), 81S–92S.

Vasen, H.F., Mecklin, J.P. et al. (1991). The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). *Dis Colon Rectum* 34(5), 424–425.

Vasen, H.F., Watson, P. et al. (1999). New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. *Gastroenterology* 116(6), 1453–1456.

Vendrell, E., Morales, C. et al. (2005). Genomic determinants of prognosis in colorectal cancer. *Cancer Lett* 221(1), 1–9.

Vineis, P. and McMichael, A. (1996). Interplay between heterocyclic amines in cooked meat and metabolic phenotype in the etiology of colon cancer. *Cancer Causes Control* 7(4), 479–486.

Vogelaar, I., van Ballegooijen, M. et al. (2006). How much can current interventions reduce colorectal cancer mortality in the US?

Mortality projections for scenarios of risk-factor modification, screening, and treatment. *Cancer* 107(7), 1624–1633.

Vogelstein, B., Fearon, E.R. et al. (1988). Genetic alterations during colorectal-tumor development. *N Engl J Med* 319(9), 525–532.

Wang, Y., Jatkoe, T. et al. (2004). Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer. *J Clin Oncol* 22(9), 1564–1571.

Ward, D.G., Suggett, N. et al. (2006). Identification of serum biomarkers for colon cancer by proteomic analysis. *Br J Cancer* 94(12), 1898–1905.

Weisberg, I., Tran, P. et al. (1998). A second genetic polymorphism in methylenetetrahydrofolate reductase (MTHFR) associated with decreased enzyme activity. *Mol Genet Metab* 64(3), 169–172.

Wiesner, G.L., Lewis, S. et al. (2000). Human studies research: Ethics, family and pedigree studies. In *Encoclopedia of biotechnology: Ethical, legal and policy issues* (T.H. Murray and M.J. Mehlma, eds), Wiely and Sons, New York, pp. 595–611. 2.

Wiesner, G.L., Daley, D. et al. (2003). A subset of familial colorectal neoplasia kindreds linked to chromosome 9q22.2-31.2. *Proc Natl Acad Sci USA* 100(22), 12961–12965.

Winawer, S., Fletcher, R. et al. (2003). Colorectal cancer screening and surveillance: Clinical guidelines and rationale-update based on new evidence. *Gastroenterology* 124(2), 544–560.

Winawer, S.J., Zauber, A.G. et al. (1996). Risk of colorectal cancer in the families of patients with adenomatous polyps National Polyp Study Workgroup. *N Engl J Med* 334(2), 82–87.

Winawer, S.J., Zauber, A.G. et al. (2006). Guidelines for colonoscopy surveillance after polypectomy: A consensus update by the US Multi-Society Task Force on Colorectal Cancer and the American Cancer Society. *CA Cancer J Clin* 56(3), 143–159, quiz 184–5.

Wood, L.D., Parsons, D.W. et al. (2007). The genomic landscapes of human breast and colorectal cancers. *Science* 318(5853), 1108–1113.

Xu, X.L., Yu, J. et al. (2004). Methylation profile of the promoter CpG islands of 31 genes that may contribute to colorectal carcinogenesis. *World J Gastroenterol* 10(23), 3441–3454.

Ye, Z. and Parry, J.M. (2002). Meta-analysis of 20 case-control studies on the N-acetyltransferase 2 acetylation status and colorectal cancer risk. *Med Sci Monit* 8(8), CR558–CR565.

Yeager, M., Orr, N. et al. (2007). Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet* 39(5), 645–649.

Zanke, B.W., Greenwood, C.M. et al. (2007). Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet* 39(8), 989–994.

CHAPTER

Genomic Evaluation and Management of Prostate Cancer

Phillip G. Febbo and Philip W. Kantoff

INTRODUCTION

Prostate cancer is the second leading cause of cancer death in men living in Western countries. The management of men diagnosed with prostate cancer remains predicated on a combination of clinical and pathological characteristics of the individual and the cancer. Molecular features of prostate cancer remain largely of academic interest and are not used in the delivery of care. However, the dramatic response of metastatic prostate cancer to castration demonstrates the profound impact of successful molecularly targeted therapy, and increasing insight into the molecular pathogenesis of prostate cancer undoubtedly holds further therapeutic promise (Balakumaran and Febbo, 2006; Nelson et al., 2003).

The genetic and epigenetic events that result in the transformation of prostate epithelial cells are becoming better understood, and genomic tools have fueled many of the recent discoveries. The synchronous development of multiple, high-quality models of prostate cancer, together with sophisticated genomic technologies, have accelerated our molecular understanding of prostate cancer over the past decade and have been instrumental in the identification of key molecular events driving prostate cancer initiation, progression, and metastasis.

The molecular story of prostate cancer, however, remains far from complete. A comprehensive molecular characterization of the approximately 220,000 men diagnosed each year in the United States of America remains untenable. As the sequence of the human genome and technologies for global assessment of DNA, RNA, and protein patterns are now available, it is hoped that the many molecular insights established over the past decade will be effectively woven together into a molecular map of prostate cancer to characterize and define therapy for each individual diagnosed.

This chapter will highlight recent prostate cancer discoveries empowered by our nascent ability to perform global analysis on DNA variation, RNA expression, and protein expression.

GENETIC PREDISPOSITION AND ALTERATIONS IN PROSTATE CANCER

Highly Penetrant Prostate Cancer Genes

Despite significant efforts, there have been limited gains in the identification of germline mutations accounting for families with a significantly increased risk of prostate cancer. Likely due to a combination of disease heterogeneity and high disease prevalence, few highly penetrant prostate cancer-causing genes have been identified to date. Genes located in regions in linkage disequilibrium with prostate cancer that have confirmed mutations in at least some families at increased risk include ELAC2, MSR1, and RNAseL (reviewed in [Schaid, 2004]), but there are no genes that are the equivalent to those such as BRCA1 and breast cancer (see Chapter 35).

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

In 2002, germline mutations within the ribonuclease L gene (RNAseL) gene were found to account for some families with a high risk of prostate cancer (Carpten et al., 2002). Subsequent work has found that a variant of RNAseL, the non-synonymous change Arg462Gln, confers decreased enzymatic activity and is associated with an increased risk of prostate cancer (Casey et al., 2002). While the mechanism(s) behind the association between prostate cancer risk and RNAseL remains unclear, a screen for viral DNA in prostate cancers identified a novel gamma-retrovirus in prostate cancers occurring primarily in individuals with a variant of RNAseL associated with decreased activity (Urisman et al., 2006). The virus, now referred to as xenotropic murine leukemia virus-related virus (XMRV), has been isolated and shown to infect human cells (Dong et al., 2007). Further studies are currently underway to determine if infection with the virus is associated with an increased risk for prostate cancer. This intriguing finding suggests a provocative mechanism connecting the RNaseL variant with an increased risk of prostate cancer through compromised immunity and resulting viral infection.

Germline Variation and Prostate Cancer Risk

Variation of many germline polymorphisms have been associated with an individual's risk of developing prostate cancer (Pomerantz et al., 2007). Most early studies focused on genetic polymorphisms, oligonucleotide repeat (di-, tri- etc.), or single nucleotide repeat polymorphisms, with potential functional impact on proteins within the androgen metabolism pathways (reviewed in [Singh et al., 2005]). However, as the paradigm of genomic assessment has shifted from single polymorphism to allele structure (haplotypes), more comprehensive analyses are now possible.

In 2006, analysis of germline polymorphisms and their association with prostate cancer identified a region on chromosome 8 in significant linkage disequilibrium with the disease (Amundadottir et al., 2006). Although the association was identified initially in an Icelandic population, two variants were found to have a strong association with prostate cancer risk across multiple populations, with the largest population attributable risk observed in African-American men (Amundadottir et al., 2006). Using an alternative approach of admixture mapping in order to identify prostate cancer risk alleles in African-American men, an independent group also identified this locus (Freedman et al., 2006). Subsequently, multiple studies have confirmed the association of this region with risk of prostate cancer and have determined that there are two independent loci that impact risk of prostate cancer located within 8q24 (Gudmundsson et al., 2007; Haiman et al., 2007a, b; Schumacher et al., 2007; Wang et al., 2007; Yeager et al., 2007). While the proximity of these polymorphisms to the MYC oncogene (~150 KB) suggests that they may act through modulation of MYC expression, the specific markers are within non-transcribed regions of the genome and the mechanism underlying the association of these loci with prostate cancer risk remains unknown.

With the advent of technologies to rapidly assay hundreds of thousands of SNPs (see Chapter 4), genome-wide association studies are becoming more prevalent and are likely to result in additional insights. For example, using a 100K SNP array from Affymetrix, investigators looked for loci with strong association with prostate cancer for incident cases within men participating in the Framingham heart study (Murabito et al., 2007). While no loci met statistical criteria for genome-wide significance, association with prostate cancer was found for two SNPs within the MSR1 gene (discussed above as being associated with a familial risk of prostate cancer). Thus, in time, more granular technologies and assessment of large cohorts with different racial and ethnic constitution will likely result in additional insights as to the genetic influence on the development of prostate cancer.

Mitochondrial DNA

Another compelling finding is the association of mutations within mitochondrial genes and prostate cancer risk (Petros et al., 2005). When the cytochrome oxidase subunit I gene (COI) was sequenced in patients with prostate cancer, 11–12% were found to have mutations that altered conserved amino acids compared to <2% individuals without cancer (Petros et al., 2005). Both germline and somatic mitochondrial mutations have been found to contribute to the risk of prostate cancer, underscoring their potential importance (Gomez-Zaera et al., 2006; Petros et al., 2005). A plausible mechanistic explanation involves the role these mutations may play to increase reactive oxygen species and oxidative stress. Increased oxidative stress is a commonly associated mechanism with prostate cancer development and may explain the observed associations between prostate cancer onset and NKX3.1 loss (Ouyang et al., 2005), as well as GST-Pi methylation (Parsons et al., 2001) in prostate cancer. Thus, changes in the mitochondrial genome may further highlight the importance of oxidative stress on the development and progression of prostate cancer.

Epigenetic Changes in Prostate Cancer

Along with genetic changes in the DNA sequence, epigenetic modifications (see Chapters 6) have been associated with the development and progression of prostate cancer. Methylation of GST-Pi is one of the earliest and most consistent findings in prostate cancer (Lee et al., 1994). Subsequently, many additional genes with CpG-rich promoters have been found to be preferentially methylated with the development of prostate cancer. A few relatively recent examples include 14-3-3sigma (Pulukuri and Rao, 2006), P4501A1 (Okino et al., 2006), TIMP-2 (Pulukuri et al., 2007), and RASFF1A (Kawamoto et al., 2007). Clearly, epigenetic regulation plays a role in the development and progression of prostate cancer, but the relative importance of the individual genes thus far identified remains largely unknown.

Investigations using technology platforms with greater coverage of the human genome are currently ongoing and are likely to continue to identify specific genomic regions altered with

prostate cancer progression (reviewed in [Nelson et al., 2007]). Already patterns of post-translational acetylation and dimethylation of histones H3 and H4 have been shown to identify two groups of low-grade prostate cancers with different risks of recurrence independent of tumor stage, preoperative prostate-specific antigen levels, and capsular invasion (Seligson et al., 2005). Importantly, the availability of high-throughput methods such as DNAse hypersensitivity that assess high-order chromatin structure (Crawford et al., 2006) represents yet another outstanding opportunity for discovery.

Somatic DNA Alterations

Genomic screens for DNA-based changes have recapitulated previous findings from targeted analyses. In one study, a microarray genotyping single nucleotide polymorphisms (SNPs) identified loss of heterozygosity in regions on chromosomes 8p (22%), 8q (24%), 16q (20%), 13q (18%), 10q (12%), and 4q (12%) (Lieberfarb et al., 2003). With the exceptions of 8q and 4q, each of these regions had been previously implicated in prostate cancer (Nelson et al., 2003). Similarly, in a recent study using array comparative genomic hybridization (aCGH), gains were seen at 1q, 7, 8q, 16p and 17q and losses at 2q, 4p/q, 6q, 8p, 13q, 16q, 17p, and 18q (Saramaki et al., 2006). With the increased resolution offered by the 16,000-feature microarray used for this study, smaller regions of gains or loss could be discerned.

Genomic analyses of somatic changes in prostate cancer have highlighted the role of specific oncogenes and pathways in prostate cancer (Table 37.1). A few are consistently associated with prostate cancer development and progression and deserve specific mention. Importantly, the detailed knowledge of these pathways in prostate cancer and specific details with respect to their role in prostate cancer provide important molecular paradigms with which to consider the role of novel molecular alterations in prostate cancer as they are identified.

PTEN/PI3K/mTOR

PTEN antagonizes the growth-promoting effects of the family of proteins constituting the dimeric complex with PI3K activity by removing the phosphate moiety from the 3′position of phosphotidylinositol 3,4,5-triphosphate (PIP3) (Li et al., 1997).

TABLE 37.1 Common chromosomal alterations in prostate cancer

Gene	Chromosomal location	Change observed	Function	Reference
ETS family members	21,7,others	Translocation downstream of AR regulated genes (primarily TMPRSS2)	Transcription factors	(Tomlins, 2006)
GSTPi	11	Methylation of promoter	Responds to oxidative stress	(Nelson, 1994)
PTEN	10q23	Loss	Tumor suppressor gene involved in regulation of the phosphatidylinositol 3-kinase pathway. Loss of the PTEN enzyme can lead to rapid cell growth and proliferation.	Hughes (2005)
NKX3.1	8p21	Loss/Inactivation	Tumor suppressor gene that is prostate specific and regulates epithelial growth and differentiation.	Hughes (2005); Shand (2006)
c-myc	8q	Gene amplification	Encodes a transcription factor that regulates the expression of multiple genes leading to increased cell proliferation.	Hughes (2005); Jenkins (1997)
Rb	13q	Loss (LOH, mutation)	Tumor suppressor gene that prevents replication of damage DNA through the cell cycle.	Hughes (2005); Phillips (1994)
p53	17p	Loss (LOH, mutation)	Tumor suppressor gene that prevents entry of damaged DNA into the cell cycle and promotes apoptosis.	Grignon (1997); Hughes (2005)
AR	X	Amplification, mutation in hormone refractory disease	Transcription factor	Koivisto (1995)

Homozygous loss of PTEN is common in advanced prostate cancer (Sansal and Sellers, 2004), and loss of one allele is likely to be an early event (Bello-DeOcampo and Tindall, 2003). Inhibition of PI3K by targeted molecules devastates prostate cancer cells with PTEN loss largely through pro-apoptotic effects (Lin et al., 1999).

It is now clear that PTEN has gene dosage effects in prostate cancer that are profound. The two transgenic mice thus far developed with prostate-specific Pten knockout (Trotman et al., 2003; Wang et al., 2003) have provided remarkable insight into PTEN biology. Although complete inactivation of PTEN is observed in a significant number of cases of advanced prostate cancer, only one allele is lost in many patients at presentation. Trotman and coworkers created a Pten hypomorphic allele to generate a series of Pten transgenic mice that had progressively decreasing levels of the PTEN protein expressed in the prostate. The incidence of prostate cancer, latency and progression correlated with Pten dose in the prostate, thus providing definitive evidence that PTEN copy number is important in prostate cancer and that the hemizygous state of PTEN may play a key role in the initiation of prostate cancer (Trotman et al., 2003).

The impact of this work has been extended with the recent demonstration of cooperativity between PTEN and p53 (Chen et al., 2005c). PTEN protects p53 protein from Mdm2-mediated degradation (Mayo et al., 2002), and p53 enhances PTEN transcription (Stambolic et al., 2001). In transgenic mice, conditional inactivation of p53 fails to produce a prostate tumor phenotype in mice, and in the same strain complete PTEN inactivation in the prostate triggers tumor formation with a long latent period. However, combined inactivation of Pten and p53 leads to formation of invasive prostate cancer as early as 2 weeks after puberty and is often lethal by 7 months of age (Chen et al., 2005c). This work demonstrated that acute PTEN inactivation induces growth arrest through the p53-dependent cellular senescence pathways both *in vitro* and *in vivo*, which can be fully rescued by combined loss of p53. This could potentially explain observations in human prostate cancer where evidence of cellular senescence is seen in early stages (by strong beta-gal staining in regions of prostatic intraepithelial neoplasia (PIN) but not in frank tumor) where cells presumably have lost p53. p53 appears to be an important checkpoint that restricts Pten-deficient tumors, and lysates from pten-null mice prostates showed a 10-fold induction of p53 (Chen et al., 2005c).

Along with major genetic modifiers such as p53, germline genetic variation is also likely to impact the phenotype of PTEN loss in prostate epithelium. This possibility is suggested by the somewhat disparate phenotypes of two prostate-specific, complete Pten knockout mice. The Pten-null C57Bl/6 × 129/Balb/c F2 mice developed by Wang and coworkers using a cre-lox knockout approach (Wang et al., 2003) exhibit a latent period for PIN formation of 8–10 months for heterozygotes mice and as short as 1.5 months for the homozygous mutant mice, which further supports the dose-dependent effects of PTEN. In addition, homozygous mutants develop invasive adenocarcinoma that progresses to metastasis and does not respond to castration. However, in a pure C57Bl/6 background, PTEN loss results in invasive cancer but not metastasis (Trotman et al., 2003).

MYC

c-Myc is a transcriptional factor that is required for expressing many genes involved in cell-cycle transition events and proliferation (Adhikary and Eilers, 2005). Overexpression of MYC will result in increased apoptosis unless there are mutations in MYC itself (Hemann et al., 2005) and/or deficits in homeostatic mechanisms that facilitate escape from p53 and p19ARF mediated apoptosis (Dang et al., 2005). Frequent amplification of chromosome 8q24 and/or the MYC oncogene has been demonstrated in androgen-independent prostate cancer with greater frequency of amplification noted as prostate cancer progresses (Edwards et al., 2003).

p53 and Rb

While distinct in their biology, p53 and Rb have both been implicated in late-stage prostate cancer and have a demonstrated causative role in transgenic mouse models. Loss of heterozygosity at the Rb loci is frequent in early prostate cancer (60% in [Phillips et al., 1994]). Rb loss results in dysregulation of the E2F transcription factors which activate genes critical for cell-cycle progression and proliferation. p53 is best characterized as causing cell-cycle arrest and apoptosis in the setting of DNA damage. p53 mutations have been associated with prostate cancer. In addition, both p53 and Rb immunohistochemistry have prognostic importance in prostate cancer (Theodorescu et al., 1997).

Fascinating recent work has further demonstrated significant interaction between Rb loss and p53 signaling (Hill et al., 2005). While the supportive role of the prostatic stroma in cancer development and progression has been recognized for some time (Cunha et al., 2002), somatic genetic changes have inconsistently been found and there was little direct support for processes whereby genetic or epigenetic events preferentially occur in stromal adjacent to malignant epithelial cells. In transgenic mice with Rb loss, there is paracrine selective pressure on the stromal fibroblasts to escape from p53 growth arrest. As a result, investigators found p53 mutations occurring within the stromal cells (Hill et al., 2005). While this specific model provides only a single example of somatic genetic changes in the stroma driven by epithelial malignancy, it likely represents a common phenomenon that helps explain the stable phenotypic differences observed between stromal cells adjacent to either benign epithelial glands or malignant epithelial glands (Cunha et al., 2003).

PROSTATE CANCER DETECTION

The discovery of circulating prostate-specific antigen (PSA) and its subsequent use as a screening test has resulted in earlier detection, but PSA screening has yet to demonstrate a survival benefit. Importantly, while PSA can diagnose prostate cancer earlier than symptoms, physical exam, or other blood chemistries, it

remains limited as a predictive marker and provides little insight as to the specific biology of an individual's prostate cancer. The goal of many genomic projects in prostate cancer is to improve upon the specificity of serum PSA and to discover and develop biomarkers that provide additional predictive or biological insight (Figure 37.1).

Urine Biomarkers

Prostate cancer cells slough into the prostatic glands and ducts and can be found in the urine. As high-grade prostate intraepithelial neoplasia often is associated with prostate cancer, the identification and characterization of prostate cells within the urine may provide an opportunity. In addition, degraded proteins derived from the serum and passing into the urine may also provide diagnostic, prognostic, or predictive information (M'Koma et al., 2007). Single molecular markers demonstrate correlation between biopsy results and urine protein levels. For example, there is good correlation between the presence of α-methylmalonyl co-A racemase (AMACR) detected in the urine and biopsy results in men undergoing diagnostic evaluation for prostate cancer (Rogers et al., 2004). When PCR-based promoter methylation assays were used to compare cells collected in the urine to prostate biopsies, correlation was very high for three genes (94% for GST-pi and APC, 82% for EDNRB) (Rogers et al., 2006). Additional urine biomarkers of potential clinical utility currently under study include PCA3 (de Kok et al., 2002; Tinzl et al., 2004) and the TMPRSS2/ERG translocations (Hessels et al., 2007). Thus, molecular events detected in cells or prostate-specific proteins present in the urine seem to reflect changes within the prostate and may serve as biomarkers to provide a means for earlier diagnosis.

Broader proteomic approaches have demonstrated some early success in correlating protein patterns with prostate pathology. Using a mass spectrometry approach (MALDI-TOF-MS), investigators have identified calgranulin B in the urine of men with prostate cancer following prostatic massage (Rehman et al., 2004). Alternatively, by adsorbing urine proteins onto a reverse-phase resin and subsequently directly spotting them onto a specific matrix (α-cyano-4-hydroxycinnamic acid), another group demonstrated peaks that could be used to distinguish between individuals with benign prostatic hypertrophy (BPH), PIN, and prostate cancer with approximately 75% specificity and 70% sensitivity (M'Koma et al., 2007). These early analyses require validation, but suggest that urine holds some promise as a source for informative biomarkers. The relative utility of urine-based analysis compared to serum-based analysis remains to be fully explored.

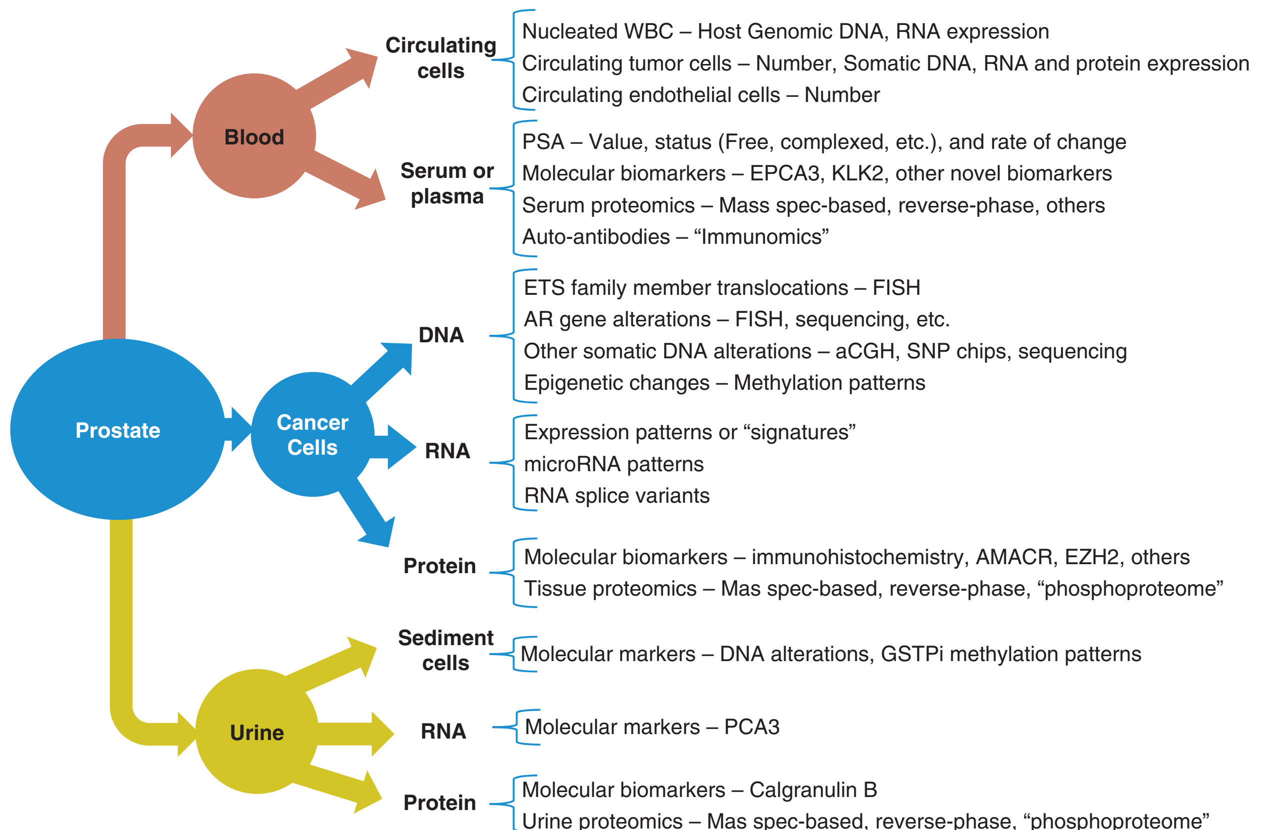


Figure 37.1 Biomarkers in prostate cancer.

Serum Biomarkers

There has been great interest in improving our ability to detect and anticipate the behavior of prostate cancer beyond that currently enabled by serum PSA. PSA kinetics (e.g., PSA velocity or PSA doubling time) has been recently broadly recognized to improve the prognostic accuracy of PSA (D'Amico et al., 2005; Freedland et al., 2005; Partin et al., 1994), but still provide relatively limited insight into the molecular biology of an individual's cancer. To this end, multiple groups have performed proteomic analysis of serum samples in an attempt to identify more sensitive and/or specific markers. Comprehensive analysis of circulating serum proteins in patients at risk or diagnosed with prostate cancer have used two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), MS-based technologies (e.g., matrix assisted laser desorption ionization time-of-flight (MALDI-TOF) or surface enhanced laser desorption ionization time-of-flight (SELDI-TOF)), and reverse-phase protein arrays (Ornstein and Tyson, 2006).

There have been early suggestions that proteomics can distinguish patients with prostate cancer from patients with a normal PSA (Petricoin et al., 2002) and can distinguish patients with advanced prostate cancer from patients with advanced breast or bladder cancer (Villanueva et al., 2006). Significant effort is now focused on determining the best technological platforms and clinical applications for use in the care of men diagnosed with prostate cancer. However, blood collection and processing remains a significant challenge to open-ended proteomic technologies such as MALDI-TOF or SELDI-TOF, as small changes in storage or protein isolation can significantly impact proteomic profiles (Gelmann and Semmes, 2004). When sample processing is standardized, differential patterns have been identified that correlate with the presence of prostate cancer (Villanueva et al., 2006) and the presence of bone metastasis (Li et al., 2005), but these findings require more extensive testing before their true prognostic or predictive value is known.

An alternative approach, referred to as reverse-phase proteomics, has also demonstrated promise (Wang et al., 2005). In a specific example, prostate cancer cDNA library is used to create a bacteriophage library and single clones expressing prostate cancer antigens are spotted onto glass microarrays. Patient's serum is labeled and hybridized to the arrays in the presence of an alternatively labeled "normal" serum. This technique determines differential presence of auto-antibodies directed against prostate cancer antigens. Investigators found that this approach had greater sensitivity and specificity than PSA in men with serum PSA measured between 4 and 10 ng/dl (Wang et al., 2005). These reverse-phase proteomic approaches may be less impacted by sample processing and be deployable clinically.

Currently, many associations between proteomic analysis and prostate cancer remain limited to single investigator teams, and no single marker or proteomic-based model has been incorporated into clinical practice. However, with improved standardization of processing and platforms, serum proteomics have great promise to improve prostate cancer care. Early studies have already provided proof of concept by identifying serum proteins that may improve upon PSA for prostate cancer diagnosis and treatment and prospective evaluation and clinical implementation of the best markers is likely to occur over the next few years.

Circulating Tumor Cells

As tumors progress, malignant cells can be found within the circulation. Surrogate measures of circulating tumor cells (CTC), such as RT-PCR for PSA from whole blood (Seiden et al., 1994), have been associated with prostate cancer risk (Ghossein et al., 1999). The development of methods to isolate CTC from blood has created further opportunity to determine if the CTC number and nature provides an emerging opportunity for biomarker development. Patients with prostate cancer seem to have a relatively heavy burden of CTC, and increased numbers of CTC are associated with more advanced disease (Moreno et al., 2001) and survival (Moreno et al., 2005). Decreases in CTC numbers in response to therapy have recently been found to be associated with survival in men treated with castration-refractory prostate cancer (Moreno et al., 2007). However, it remains to be seen if the biology of an individual's CTC significantly reflects that of their metastatic disease.

Given the relatively small numbers (a median of 5 CTC/7.5 ml of blood in patients with advanced prostate cancer [Moreno et al., 2005]), the isolation and analysis of CTC has proved to be a significant challenge. Thus far, analysis has been largely restricted to specific proteins or genes (PSA, PSMA, etc.) (Chen et al., 2005a). However, investigators have demonstrated the ability to detect molecular markers such as EGFR expression, DNA ploidy, and/or androgen receptor (AR) amplification in samples with >5 CTC/7.5 ml of blood (Shaffer et al., 2007). As cells are more efficiently isolated and methods of DNA or RNA amplification improved, it is possible that genomic analysis of this compartment may provide surrogate prognostic or predictive biomarkers.

Imaging Prostate Cancer

Imaging of the prostate is performed routinely, but thus far has demonstrated limited utility in defining intraprostatic disease burden, and no imaging modality has provided strong predictive correlates with prostate cancer behavior. Imaging modalities that have been in routine use and tested for their ability to define either intraprostatic tumor burden or histopathological correlates include transrectal ultrasound, pelvic CT, pelvic MR, pelvic MR with endorectal coils, Prostascint scans, and positron-emission tomography scans. A few imaging modalities are successful in suggesting extraprostatic disease (CT, pelvic MR +/− endorectal coils) or enlarged regional lymph nodes (MR +/− endorectal coil). In a recent study, MR evidence of non-organ confined disease (either extracapsular, seminal vesicle, or lymph node involvement) contributed to the prognostic nomograms (Kattan nomogram) that are frequently used in clinical practice (Wang et al., 2006). However, no modality has been proven to provide sufficient clinically useful information about intraprostatic

tumor burden, pathology, or biology so as to warrant widespread adoption.

MR spectroscopy is an emerging technology that has yet to be fully evaluated for its potential to provide prognostic and/or predictive information about cancer (Sorensen, 2006). Preliminary studies focused on prostate cancer have shown considerable promise (Hricak, 2005). MR spectroscopy of prostates after removal have demonstrated spectroscopic correlates with pathology (Cheng et al., 2005) and *in vivo* studies are underway. MR spectroscopy was found to contribute to the anatomical information provided by MR in the study by Wang and coworkers although the contribution was not found to be statistically significant (Wang et al., 2006). While widespread use of MR spectroscopy has been somewhat limited because of concerns regarding technical reproducibility, emerging work has started to identify spectroscopic approaches that are more readily performed across institutions. In addition, while there has been relatively poor correlation between MR imaging and intraprostatic tumor volume or pathology, few studies have addressed the correlation of MR spectroscopy and intratumor volume and/or pathology. Finally, there have been no studies bringing together MR spectroscopy and genomic analyses of prostate cancer to determine if there are spectroscopic findings that identify prostate cancer tumors with poor prognostic or specific predictive genomic signatures.

GENOMIC CHANGES ASSOCIATED WITH PROSTATE CANCER BEHAVIOR

The progression of the prostate epithelial cells from benign to malignant and from local to metastatic is associated with genetic, epigenetic, RNA, and protein alterations. Significant effort has lead to our understanding how specific molecular events impact the development and progression of prostate cancer. More recently, global approaches of genomic interrogation have been applied to discover and define the most important molecular events in prostate cancer.

Prostate Cancer Initiation

Early expression analyses of prostate cancer focused on global differences between benign tissue (either "normal" prostate glands or BPH) and "cancer." Multiple groups have demonstrated significant differential gene expression between benign and malignant localized prostate cancer specimens using microarrays (Dhanasekaran et al., 2001; Lapointe et al., 2004; Luo et al., 2001; Singh et al., 2002; Welsh et al., 2001). The difference in global gene expression between tissue samples comprised of benign epithelium and those containing malignant prostate cells were profound and several genes (e.g., AMACR and hepsin) were common to multiple groups. Importantly, microarray analysis demonstrated that AMACR had consistently increased expression in prostate cancer (Luo et al., 2001). AMACR expression was subsequently validated by immunohistochemistry (Luo et al., 2002; Rubin et al., 2002) and is currently used clinically in some situations when the diagnosis of malignancy is unclear.

Perhaps one of the most significant findings resulting from microarray analysis is the identification of recurrent chromosomal rearrangements associated with prostate cancer (Tomlins et al., 2005). By looking at data from gene expression microarrays, this group identified outlier genes whose expression was uncharacteristically high in individual cases of prostate cancer compared to other cancers. This approach identified the frequent aberrant expression of two transcription factors belonging to the ETS family, *ERG* and *ETV1*. When the genomic location and organization of these genes were further interrogated, they were found to be fused to the 5′ end of the prostate-specific, androgen-regulated gene *TMPRSS2,* with one or the other fusion present in a very high proportion of localized prostate cancer specimens (23/29) (Tomlins et al., 2005), suggesting that fusion of an ETS family member with TMPRSS2 may be a common initiating event in prostate cancer. Multiple groups have likewise identified these translocations in a high percentage of prostate cancer patients (Hermans et al., 2006; Iljin et al., 2006; Tu et al., 2007). In addition, there have been early reports that the presence of ETS translocations are associated with more aggressive pathology (Mosquera et al., 2007) and higher prostate cancer specific mortality (Demichelis et al., 2007). The translocation between *TMPRSS2* and an ETS family member appears to be an important early event in prostate cancer and, while it may have less impact on late-stage disease (Hermans et al., 2006), targeting these translocations is likely to be a clinically important approach.

Biochemical Relapse Following Surgery

As an imperfect but feasible measure of progression, many groups have analyzed the difference in RNA expression patterns between prostate cancers from men who were cured (or remained biochemically free of disease for at least 4 years) versus men who had biochemical recurrence (Glinsky et al., 2004; Henshall et al., 2003; Lapointe et al., 2004; Singh et al., 2002). Interestingly, while each independent group could consistently find expression patterns anticipating biochemical recurrence using multi-gene predictive models, there was little overlap in the genes used in each group's specific model. This can be interpreted in many ways, but is likely due to the consistent presence of expression structure associated with biochemical relapse that is subtle (i.e., low signal to noise ratio) yet deep (many genes). The consistency of such an outcome signature suggests that genetic events in the primary tumors do determine a propensity to spread; however, the biology driving such expression patterns remains far from clear, and recurrence signatures may not be unique to prostate cancer.

The pre-existence of genetic or epigenetic events that determine aggressive behavior is also supported by DNA-based analysis. Using aCGH, investigators have found a higher frequency of DNA copy number changes associated with recurrent disease and specifically identified loss of 8p being associated with advanced stage and gain of 11q13 with biochemical recurrence

(Paris et al., 2004). The specific regions found to be amplified contained the *MEN1* gene, and subsequent analysis confirmed a difference in RNA expression.

Integrative Analysis

While preliminary genomic analysis of prostate cancer was largely restricted to single molecular compartments (i.e., DNA, RNA, etc.), there is an increasing utilization of integrative analysis that uses more than one molecular compartment to investigate tumor biology. An integrative approach using extensive genomic and proteomic analysis of prostate cancer found that, compared to benign prostate, 64 proteins were altered in localized prostate cancers, and an additional 156 proteins were detected in metastatic cancer. Only 48–64% concordance was observed between the RNA and proteomic analyses but the genes that were correlated with outcome by either both methods served as good predictors of clinical outcome in prostate cancer as well as other tumors (Varambally et al., 2005). Such integrative analyses are likely typical of future genomic studies and will help to determine molecular sub-types of prostate cancer in order to provide a more comprehensive understanding of prostate cancer biology and improve prostate cancer treatment.

Another approach, while not involving more than one molecular compartment, is to use differential gene expression as a phenotype and to assess the similarity between gene expression as prostate cancer progresses with a broad range of expression phenotypes derived experimentally. In a recent paper, gene expression was assessed in prostate samples ranging from benign to malignant disease including a number of metastatic samples (Tomlins et al., 2007). Differential gene expression between stages of progression was used to assess similarity with gene expression changes associated with other expression-based phenotypes. For example, this process – termed the molecular concept map (MCM) – found that genes associated with proliferation in cell lines had a coordinated increase in expression as prostate cancer transitioned from local to metastatic cancer. While somewhat limited due to their analysis of only 44 individuals were, this analytic method has started to identify the biology underlying prostate cancer progression and metastasis and suggests further analysis of larger datasets will be fruitful.

Correlates with Pathology

There are also gene expression signatures associated with prostate cancer pathology (Singh et al., 2002). Genes found to have increased expression in tumors of increasing pathological grade (as measured using the Gleason grading system) were found to be consistent across independent sets of prostate cancer specimens (Singh et al., 2002). Interestingly, a large number of these genes were downstream targets of TGF-beta signaling and were among the top genes correlated with Gleason score (Febbo and Sellers, 2003). Subsequent work has further supported strong gene expression associations with Gleason grade (Halvorsen et al., 2005) and found that gene expression patterns can be used to distinguish Gleason grade 3 from Gleason grade 4/5 with 76% accuracy (True et al., 2006).

These analyses suggest that a biology associated with aggressive behavior can be found in the localized tumors and likely predates the spread of cancer. While genes such as *MUC1* (Lapointe et al., 2004), monoamine oxidase A (True et al., 2006) and many others have been associated with Gleason score, the specific biology driving prostate cancer aggressive disease is not fully understood.

Changes Associated with Progression to Metastatic Prostate Cancer

An alternative approach to identifying molecular changes associated with aggressive prostate cancer is to compare local to metastatic disease. By measuring gene expression in hormone-refractory prostate cancer and comparing it to local disease, RNA and protein expression of *enhance of zeste homologue 2* (EZH2) was found to be elevated (Varambally et al., 2002). Increased expression of EZH2 in localized tumors was subsequently associated with recurrence.

Interestingly, the prognostic importance of genes differentially expressed between local and metastatic tumors may be independent of tumor type. When local tumors from a variety of solid cancers were compared to a variety of metastatic samples, a 17-gene signature of metastasis was derived. When this signature was applied to prostate, breast, and meduloblastoma, the signature could anticipate disease recurrence greater than expected by chance alone (Ramaswamy et al., 2003). However, when the prognostic importance of single genes such as *EZH2* are assessed in independent datasets and in the context of multivariable models, they seldom hold greater prognostic value than established clinical and pathological features (Rhodes et al., 2003).

Literature Mining

In another example of integrative analysis, the interpretation of differential gene expression between different stages of prostate cancer can be aided by high-throughput annotation programs, such as those that perform text-based literature mining. In a recent example, genes found to have differential expression between local and metastatic prostate cancer were used as a set to find pathways that cooccurred in the corpus of medical literature more frequently than expected by chance alone. This approach identified FOSB as being associated with the genes differentially expressed in metastatic prostate cancer, and subsequent immunohistochemistry found increased nuclear FOSB staining in metastatic prostate cancer compared to local prostate cancer (Febbo et al., 2007). Additional annotation approaches using Gene Ontology or Pathway membership are being used with increasing frequency to help interpret genomic findings.

Prostate Cancer Response to Treatment

Genomics to Inform Clinical Trials

Genomics have already been incorporated into clinical trials so as to investigate the biological response to novel treatments in prostate cancer and identify potential mechanisms of resistance. In the setting of a phase II neoadjuvant Docetaxel (Taxotere)

trial for high-risk localized prostate cancer, microarray-based gene expression was measured in tumors following docetaxel (Taxotere) therapy and compared to untreated prostate tumors matched for Gleason grade (Febbo et al., 2005). While there were no genes with large differences in expression (>5 fold between treated and untreated tumors), gene set enrichment analysis (GSEA) (Subramanian et al., 2005) identified altered expression of genes involved in androgen metabolism. This finding further underscores the potential importance of androgen signaling in prostate cancer and has helped to inform the development a large multi-center phase III trial (CALBG [cancer and leukemia group B] 90203) that compares neoadjuvant docetaxel (Taxotere) and androgen ablation followed by surgery to surgery alone for men with similar high-risk prostate cancer (Eastham et al., 2003).

In another example, patterns of gene expression were determined in localized prostate cancer before and after neoadjuvant Imatinib Mesylate (Gleevec) given to men with intermediate to high-risk prostate cancer prior to prostatectomy (Febbo et al., 2006). Microarray analysis following laser capture microdissection and RNA amplification was used to determine gene expression changes associated with therapy from nine patients. The study revealed large gene expression differences, and the gene most differentially expressed, *MPK1*, was validated by immunohistochemistry. GSEA comparing Imatinib Mesylate (Gleevec)-treated prostate cancer with untreated or pre-treatment biopsies highlighted a potential impact of treatment on apoptosis of cells associated with tumor microvasculature. Additional studies are now underway that look to improve the impact of targeted therapy on microvasculature by either using agents that target multiple receptor tyrosine kinases on the tumor vasculature (Neoadjuvant Sunitinib [Sutent], Duke University Medical Center, Daniel George, P.I., and Neoadjuvant Sorafenib (Nexavar), Fred Hutchinson Cancer Research Center, Evan Yu, P.I., which targets both PDGFR and VEGF) or combining targeted therapy with chemotherapy (Neoadjuvant Bevacizumab [Avastin] and Docetaxel [Taxotere], Dana Farber Cancer Institute, William Oh, P.I.). Further studies using laser capture microdissection and RNA amplification with targeted agents have the potential to lead to the discovery of potential mechanisms of targeted therapy in cancer.

Guiding Cytotoxic or Targeted Therapy

In 2004, two pivotal trials, TAX 327 and SWOG 99-16, demonstrated a survival benefit for docetaxel in metastatic prostate cancer (Petrylak et al., 2004; Tannock et al., 2004). While this represents a significant advance, fully 40% of patients treated with docetaxel progress by the sixth cycle of docetaxel, and improved therapeutic options are required. Multiple phase III trials testing docetaxel-doublets are underway aimed at improving the outcome for men treated with docetaxel, but an alternative approach remains improved selection of patients who receive the greatest benefit from docetaxel. As men who normalize their PSA in response to docetaxel have median life expectancies that reach almost 3 years, additional therapies may only add toxicity. Recently, *in vitro* data from the NCI-60 panel of cell lines with Affymetrix microarray data and chemotherapy sensitivity data have found that chemotherapy sensitivity signatures can be developed that predict clinical response in patients with breast, lung, and ovarian cancer (Potti et al., 2006). Studies are now underway to determine if these genomic signatures can predict response to docetaxel in prostate cancer.

Along with chemotherapy sensitivity genomic signatures, pathway activity signatures have been published that may help identify patients most likely to respond to targeted therapy (Bild et al., 2006). Important questions remain regarding the most effective means of developing predictive pathway signatures to help guide therapy, but multiple studies are currently underway that use genomics to enrich for responsive patients. At Duke University, a single arm Phase II study of RAD001 (Everolimus – Novartis) in men with HRPC is underway in which CT-guided biopsies are performed prior to and following treatment with RAD001 (Everolimus) to evaluate the molecular, genetic, and genomic effects of RAD001 in these tumor specimens and to determine if response to RAD001 can be predicted based upon the molecular state an individual's tumor.

GENOMIC CHANGES ASSOCIATED WITH HORMONE-REFRACTORY PROSTATE CANCER

AR Activity

Prostate cancer's dependence on androgens has been known for over a half a century since Huggins and Hodges demonstrated the dramatic palliative effects of orchiectomy for patients with metastatic prostate cancer (Huggins and Hodges, 1941). Classically, androgens act by binding to the intracellular AR, a ~110kD nuclear transcription factor with a central DNA binding domain, a ligand binding domain, and both ligand-dependent and ligand-independent transcriptional activation domains (reviewed in [Febbo and Brown, 2002]). The regulation of AR activity involves a complex dance of protein–protein and protein–ligand binding eventually resulting in transcription of a subset of genes containing androgen response elements (AREs). While this classic mechanism of action remains valid, complexities including non-transcriptional effects of androgens, AR co-regulators and transcriptional specificity, and non-autonomous AR effects have built upon the "classic" understanding of AR actions (Taplin and Balk, 2004).

AR Transcriptional Targets

It is clear that identification of critical downstream targets of the AR likely holds significant biological and therapeutic importance in prostate cancer. The precise identification of AR targets is complicated by the ubiquitous presence of the degenerate AR response element in the genome and the complex interaction of AR with regulatory proteins. Of the known AR chaperones, Hsp90, has, perhaps, the most significant impact on AR activity. Inhibition of Hsp90 function results in proteosomal degradation

of proteins that require this chaperone for stability, including AR (Solit et al., 2002). In prostate cancer models, inhibition of hsp90 has been shown to modify the effect of dihydrotestosterone and inhibit the growth of hormone-sensitive and resistant tumors (Harashima et al., 2005; Solit et al., 2003). Interestingly, when the gene expression patterns following androgen exposure in LNCaP cells, an androgen sensitive prostate cancer cell line, was used as a "signature" with which to identify small molecules with expression antagonistic expression profiles, inhibitors of hsp90 were among the top chemicals on the list (Hieronymus et al., 2006).

Recent studies have looked at the global expression changes induced by castration in men with prostate cancer so as to glean biological insight. Interestingly, many of the genes identified in cancer cell lines were found to decrease following chemical castration when intraprostatic androgen levels were sufficiently reduced (Mostaghel et al., 2007). However, often some AR signaling was maintained as determined by the continued expression of putative AR target-genes and serum androgen levels did not correlate strongly with intraprostatic hormone levels. Thus, proactive assessment of continued AR activity and more aggressive AR inhibition is likely to be a fruitful area of future clinical research.

AR Addiction

Androgen ablation is the first line of therapy in prostate cancer. The significant biochemical and/or symptomatic response by most patients to castration demonstrates oncogenic addiction to AR signaling. However, most patients develop resistance to the treatment and prostate cancer progresses to the "hormone-refractory" stage. Importantly, it is now clear that while prostate cancer is progressing in the setting of low circulating levels of testosterone, most cancers are still dependent on AR signaling (reviewed in [Scher and Sawyers, 2005]). In approximately one-third of patients with hormone-refractory prostate cancer, the *AR* gene is amplified (Bubendorf et al., 1999; Koivisto et al., 1995; Visakorpi et al., 1995). For patients treated with specific AR-inhibitors, a subset will develop mutations in the AR gene that enable the mutant AR to be activated by additional androgens as well as progesterone and estrogen (Taplin et al., 1995, 1999). Similarly, a prostate cancer cell line, MDA PCa, established from a patient who had undergone castration, had two mutations in the AR receptor that made it more responsive to a sub-family of circulating steroids not effected by castration (Krishnan et al., 2002; Zhao et al., 1999).

Recently, in hormone-resistant prostate cancer xenografts, increased *AR* RNA expression was the most consistent RNA change associated with hormone-refractory growth (Chen et al., 2004). Subsequently, the authors demonstrated that increased AR expression was both sufficient and necessary for "hormone-refractory" growth of the xenografts (Chen et al., 2004). This work is supported by analysis of metastatic, androgen-independent human tumors that have very high expression of the AR and upregulation of metabolic enzymes that increase bioactive androgens when compared to local, untreated tumors using microarrays (Stanbrough et al., 2006). Thus, it has become very clear that continued AR signaling is a critical mechanism of hormone-refractory prostate cancer. Ongoing investigations are focused on the mechanisms by which prostate epithelial cells can maintain AR signaling in the setting of low circulating androgens.

2009 UPDATE

Important progress continues to be made in the biological understanding of the development and progression of prostate cancer. The desire to leverage biological insight for therapeutic gain continues to be fueled by disappointing negative results from multiple promising approaches for the prevention (Vitamin E and Selenium) and treatment (DN-101, GVAX, satraplatin) of prostate cancer (Jakobsdottir et al., 2009; Lippman et al., 2009). In 2008–2009, the field continued to be focused on studies aimed at the identification of additional prognostic and predictive genomic biomarkers, the clinical and biological implications of chromosomal translocations involving ETS family members, the persistent role of the AR in castration-resistant prostate cancer (CRPC), and overcoming therapeutic resistance to docetaxel.

Genome-wide association studies continue to identify novel regions associated with prostate cancer risk including loci on chromosomes 3, 6, 7, 9, 10, 11, 19, 22, and X (Duggan et al., 2007; Eeles et al., 2008; Hsu et al., 2009; Sun et al., 2009; Thomas et al., 2008). Individually, the identified regions had relatively minor absolute impact on risk, and the ideal means to aggregate alleles and produce a composite score remain unclear. Indeed, even when additive models of aggregate risk are shown to have promising results using logistic regression and receiver operator characteristic (ROC) curves, these may not translate into a clinically useful test (Jakobsdottir et al., 2009).

Expression-based prognostic and predictive models have also continued to be developed in order to improve upon risk stratification and to identify men at high risk for recurrence following definitive local therapy. Over the past 2 years, this work has transitioned from analysis of frozen prostate tissue to paraffin-embedded prostate tissue. Incorporating genes differentially expressed between prostate cancers with a high (4 or 5) or low (3) Gleason grade, investigators developed a model with an area under the curve of 0.81 in ROC analysis (Cheville et al., 2008). In the largest study of its kind, investigators at the Mayo Clinic were able to distinguish patients

who only had biochemical recurrence following prostatectomy ($n \sim 200$) from those who progressed to clinically metastatic disease ($n \sim 200$) (Nakagawa et al., 2008). While these models still seldom outperform multivariate predictive nomograms that use only clinical and pathological features (i.e., the Kattan nomograms and/or Partin tables), they offer the promise of molecular insight so as to guide the most appropriate treatment to those in greatest need.

Finally, new molecular compartments have been shown to be prognostically important in prostate cancer. First, the metabolic profiling has associated sarcosine levels with aggressive prostate cancer (Sreekumar et al., 2009). Using mass spectrometry, investigators from Michigan investigated the association between approximately 1100 metabolites in over 200 samples representing a range of prostate cancer progression. They found that, similar to RNA-based analysis, the different stages of prostate cancer could be differentiated based upon metabolic profile and that sarcosine was markedly elevated in advanced prostate cancer. Inhibition of the metabolic pathway generating sarcosine reduced invasive behavior of prostate cancer. In addition, micro-RNA 101 has been found to be a target of chromosomal loss resulting in increased expression of EZH2 and increased invasive behavior of prostate cancer (Varambally et al., 2008). While the particular implications of these findings for the diagnosis and treatment of prostate cancer remain unclear, additional work in blood, urine, and semen is likely to provide additional prognostic information.

The identification of translocations in prostate cancer has been expanded beyond the originally described ones (TMPRSS2:ERG1 and TMPRSS2:ETV1) (Tomlins et al., 2005). Indeed, additional translocations resulting in the juxtaposition of ETS family members with constitutively active regions have been identified (Tomlins et al., 2007). Most recently, a transcriptome resequencing effort has identified additional translocations (Maher et al., 2009). While general statements regarding the prognostic impact of these translocations have not held up, their frequency in prostate cancer continues to suggest their fundamental role and high potential as therapeutic targets.

Finally, there continues to be great enthusiasm for targeting AR in what was previously called androgen-independent or hormone-refractory prostate cancer. This stage is now commonly referred to as CRPC due to a number of independent groups finding persistent dependence on AR activity despite castrate levels of circulating testosterone. *De novo*, local synthesis of androgens has been identified as a common mechanism (Mostaghel et al., 2007; Stanbrough et al., 2006). The selective pressure of a low androgen state also results in a variety of AR mutations including those previously described in the ligand-binding domain (Taplin et al., 1999) as well as novel mutations impacting protein–protein interactions (AR and E3 ubiquitin ligase) and protein–DNA interactions (Steinkamp et al., 2009). These and many other findings continue to underscore the AR as a critical therapeutic target and are generating great enthusiasm for the novel AR-targeting agents, aberaterone (Haidar et al., 2003) and MDV3100 (Tran et al., 2009). These agents are in mid- to late-stage clinical trials and are likely to represent future therapies for men with CRPC.

It has also become clear that while most prostate cancers continue to have an active AR, the extent of AR activity is heterogeneous within and between individuals. Using a gene expression-based signature for AR, investigators have demonstrated that while activity of the AR decreases with progression to CRPC, some individuals maintain levels as high as those in patients with treatment naïve prostate cancer (Mendiratta et al., 2009). While the novel AR-targeting therapies are likely to represent a significant advance for patients with persistent AR activity, there are few promising treatment approaches for individuals who are found to have tumors with little to no AR production (i.e., AR-null tumors). It is likely that these AR-null tumors will become more important as more effective AR-targeting therapies select for cells that have become truly androgen independent. Using tissue-based, serum-based, or circulating tumor cell-based assays to distinguish between patients afflicted with tumors with high or low AR activity is likely to impact care by providing the most appropriate therapy for each man with CRPC.

2009 UPDATE REFERENCES

Cheville, J. C., Karnes, R. J.,Therneau, T. M., Kosari, F., Munz, J.M, Tillmans, L., Basal, E., Rangel, L.J., Bergstralh, E., Kovtun, I.V. et al. (2008). Gene panel model predictive of outcome in men at high-risk of systemic progression and death from prostate cancer after radical retropubic prostatectomy. *J Clin Oncol* 26(24), 3930–3936.

Duggan, D., Zheng, S.L., Knowlton, M., Benitez, D., Dimitrov, L., Wiklund, F., Robbins, C., Isaacs, S.D., Cheng, Y., Li, G. et al. (2007). Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP. *J Natl Cancer Inst* 99(24), 1836–1844.

Eeles, R.A., Kote-Jarai, Z., Giles, G.G., Olama, A.A., Guy, M., Jugurnauth, S.K., Mulholland, S., Leongamornlert, D.A., Edwards, S.M., Morrison, J. et al. (2008). Multiple newly identified loci associated with prostate cancer susceptibility. *Nat Genet* 40(3), 316–321.

Haidar, S., Ehmer, P.B., Barassin, S., Batzi-Hartmann, C. and Hartmann, R.W. (2003). Effects of novel 17alpha-hydroxylase/

C17, 20-lyase (P450 17, CYP 17) inhibitors on androgen biosynthesis *in vitro* and *in vivo*. *J Steroid Biochem Mol Biol* 84(5), 555–562.

Hsu, F.C., Sun, J., Wiklund, F., Isaacs, S.D., Wiley, K.E., Purcell, L.D., Gao, Z., Stattin, P., Zhu, Y., Kim, S.T. et al. (2009). A novel prostate cancer susceptibility locus at 19q13. *Cancer Res* 69(7), 2720–2723.

Jakobsdottir, J., Gorin, M.B., Conley, Y.P., Ferrell, R.E. and Weeks, D.E. (2009). Interpretation of genetic association studies: Markers with replicated highly significant odds ratios may be poor classifiers. *PLoS Genet* 5(2), e1000337.

Lippman, S.M., Klein, E.A., Goodman, P.J., Lucia, M.S., Thompson, I.M., Ford, L.G., Parnes, H.L., Minasian, L.M., Gaziano, J.M., Hartline, J.A. et al. (2009). Effect of selenium and vitamin E on risk of prostate cancer and other cancers: The Selenium and Vitamin E Cancer Prevention Trial (SELECT). *JAMA* 301(1), 39–51.

Maher, C.A., Kumar-Sinha, C., Cao, X., Kalyana-Sundaram, S., Han, B., Jing, X., Sam, L., Barrette, T., Palanisamy, N. and Chinnaiyan, A. M. (2009). Transcriptome sequencing to detect gene fusions in cancer. *Nature* 458(7234), 97–101.

Mendiratta, P., Mostaghel, E., Guinney, J., Tewari, A.K., Porrello, A., Barry, W.T., Nelson, P.S. and Febbo, P.G. (2009). Genomic strategy for targeting therapy in castration-resistant prostate cancer. *J Clin Oncol* 27(12), 2022–2029.

Mostaghel, E.A., Page, S.T., Lin, D.W., Fazli, L., Coleman, I.M., True, L.D., Knudsen, B., Hess, D.L., Nelson, C.C., Matsumoto, A.M. et al. (2007). Intraprostatic androgens and androgen-regulated gene expression persist after testosterone suppression: Therapeutic implications for castration-resistant prostate cancer. *Cancer Res* 67(10), 5033–5041.

Nakagawa, T., Kollmeyer, T.M., Morlan, B.W., Anderson, S.K., Bergstralh, E.J., Davis, B.J., Asmann, Y.W., Klee, G.G., Ballman, K.V. and Jenkins, R.B. (2008). A tissue biomarker panel predicting systemic progression after PSA recurrence post-definitive prostate cancer therapy. *PLoS ONE* 3(5), e2318.

Sreekumar, A., Poisson, L.M., Rajendiran, T.M., Khan, A.P., Cao, Q., Yu, J., Laxman, B., Mehra, R., Lonigro, R.J., Li, Y. et al. (2009). Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* 457(7231), 910–914.

Stanbrough, M., Bubley, G.J., Ross, K., Golub, T.R., Rubin, M.A., Penning, T.M., Febbo, P.G. and Balk, S.P. (2006). Increased expression of genes converting adrenal androgens to testosterone in androgen-independent prostate cancer. *Cancer Res* 66(5), 2815–2825.

Steinkamp, M.P., O'Mahony, O.A., Brogley, M., Rehman, H., Lapensee, E.W., Dhanasekaran, S., Hofer, M.D., Kuefer, R., Chinnaiyan, A., Rubin, M. A. et al. (2009). Treatment-dependent androgen receptor mutations in prostate cancer exploit multiple mechanisms to evade therapy. *Cancer Res* 69(10), 4434–4442.

Sun, J., Zheng, S.L., Wiklund, F., Isaacs, S.D., Li, G., Wiley, K.E., Kim, S.T., Zhu, Y., Zhang, Z., Hsu, F.C. et al. (2009). Sequence variants at 22q13 are associated with prostate cancer risk. *Cancer Res* 69(1), 10–15.

Taplin, M.E., Bubley, G.J., Ko, Y.J., Small, E.J., Upton, M., Rajeshkumar, B. and Balk, S.P. (1999). Selection for androgen receptor mutations in prostate cancers treated with androgen antagonist. *Cancer Res* 59(11), 2511–2515.

Thomas, G.K., Jacobs, B., Yeager, M., Kraft, P., Wacholder, S., Orr, N., Yu, K., Chatterjee, N., Welch, R., Hutchinson, A. et al. (2008). Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet* 40(3), 310–315.

Tomlins, S.A., Laxman, B., Dhanasekaran, S.M., Helgeson, B.E., Cao, X., Morris, D.S., Menon, A., Jing, X., Cao, Q., Han, B. et al. (2007). Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature* 448(7153), 595–599.

Tomlins, S.A., Rhodes, D.R., Perner, S., Dhanasejkaran, S.M., Mehra, R., Sun, X.W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R. et al. (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* 310(5748), 644–648.

Tran, C., Ouk, S., Clegg, N.J., Chen, Y., Watson, P.A., Arora, V., Wongvipat, J., Smith-Jones, P.M., Yoo, D., Kwon, A. et al. Development of a second-generation antiandrogen for treatment of advanced prostate cancer. *Science* 2009 April 9; [Epub ahead of print].

Varambally, S., Cao, Q., Mani, R.S., Shankar, S., Wang, X., Ateeq, B., Laxman, B., Cao, X., Jing, X., Ramnarayanan, K. et al. (2008). Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer. *Science* 322(5908), 1695–1699.

REFERENCES

Adhikary, S. and Eilers, M. (2005). Transcriptional regulation and transformation by Myc proteins. *Nat Rev Mol Cell Biol* 6(8), 635–645.

Amundadottir, L.T., Sulem, P., Gudmundsson, J., Helgason, A., Baker, A., Agnarsson, B.A., Sigurdsson, A., Benediktsdottir, K.R., Cazier, J.B., Sainz, J. et al. (2006). A common variant associated with prostate cancer in European and African populations. *Nat Genet* 38(6), 652–658.

Balakumaran, B.S. and Febbo, P.G. (2006). New insights into prostate cancer biology. *Hematol Oncol Clin North Am* 20(4), 773–796.

Bello-DeOcampo, D. and Tindall, D.J. (2003). TGF-betal/Smad signaling in prostate cancer. *Curr Drug Targets* 4(3), 197–207.

Bild, A.H., Yao, G., Chang, J.T., Wang, Q., Potti, A., Chasse, D., Joshi, M.B., Harpole, D., Lancaster, J.M., Berchuck, A. et al. (2006). Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 439(7074), 353–357.

Bubendorf, L., Kononen, J., Koivisto, P., Schraml, P., Moch, H., Gasser, T.C., Willi, N., Mihatsch, M.J., Sauter, G. and Kallioniemi, O.P. (1999). Survey of gene amplifications during prostate cancer progression by high-throughout fluorescence in situ hybridization on tissue microarrays. *Cancer Res* 59(4), 803–806.

Carpten, J., Nupponen, N., Isaacs, S., Sood, R., Robbins, C., Xu, J., Faruque, M., Moses, T., Ewing, C., Gillanders, E. et al. (2002). Germline mutations in the ribonuclease L gene in families showing linkage with HPC1. *Nat Genet* 30(2), 181–184.

Casey, G., Neville, P.J., Plummer, S.J., Xiang, Y., Krumroy, L.M., Klein, E.A., Catalona, W.J., Nupponen, N., Carpten, J.D., Trent, J.M. et al.

(2002). RNASEL Arg462Gln variant is implicated in up to 13% of prostate cancer cases. *Nat Genet* 32(4), 581–583.

Chen, C.D., Welsbie, D.S., Tran, C., Baek, S.H., Chen, R., Vessella, R., Rosenfeld, M.G. and Sawyers, C.L. (2004). Molecular determinants of resistance to antiandrogen therapy. *Nat Med* 10(1), 33–39.

Chen, B.T., Loberg, R.D., Neeley, C.K., O'Hara, S.M., Gross, S., Doyle, G., Dunn, R.L., Kalikin, L.M. and Pienta, K.J. (2005a). Preliminary study of immunomagnetic quantification of circulating tumor cells in patients with advanced disease. *Urology* 65(3), 616–621.

Chen, W.W., Chan, D.C., Donald, C., Lilly, M.B. and Kraft, A.S. (2005b). Pim family kinases enhance tumor growth of prostate cancer cells. *Mol Cancer Res* 3(8), 443–451.

Chen, Z., Trotman, L.C., Shaffer, D., Lin, H.K., Dotan, Z.A., Niki, M., Koutcher, J.A., Scher, H.I., Ludwig, T., Gerald, W. et al. (2005c). Crucial role of p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis. *Nature* 436(7051), 725–730.

Cheng, L.L., Burns, M.A., Taylor, J.L., He, W., Halpern, E.F., McDougal, W.S. and Wu, C.L. (2005). Metabolic characterization of human prostate cancer with tissue magnetic resonance spectroscopy. *Cancer Res* 65(8), 3030–3034.

Crawford, G.E., Holt, I.E., Whittle, J., Webb, B.D., Tai, D., Davis, S., Margulies, E.H., Chen, Y., Bernat, J.A., Ginsburg, D. et al. (2006). Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). *Genome Res* 16(1), 123–131.

Cunha, G.R., Hayward, S.W. and Wang, Y.Z. (2002). Role of stroma in carcinogenesis of the prostate. *Differentiation* 70(9–10), 473–485.

Cunha, G.R., Hayward, S.W., Wang, Y.Z. and Ricke, W.A. (2003). Role of the stromal microenvironment in carcinogenesis of the prostate. *Int J Cancer* 107(1), 1–10.

D'Amico, A.V., Renshaw, A.A., Sussman, B. and Chen, M.H. (2005). Pretreatment PSA velocity and risk of death from prostate cancer following external beam radiation therapy. *JAMA* 294(4), 440–447.

Dang, C.V., O'Donnell, A. and Juopperi, K.T. (2005). The great MYC escape in tumorigenesis. *Cancer Cell* 8(3), 177–178.

de Kok, J.B., Verhaegh, G.W., Roelofs, R.W., Hessels, D., Kiemeney, L.A., Aalders, T.W., Swinkels, D.W. and Schalken, J.A. (2002). DD3(PCA3), a very sensitive and specific marker to detect prostate tumors. *Cancer Res* 62(9), 2695–2698.

Demichelis, F., Fall, K., Perner, S., Andren, O., Schmidt, F., Setlur, S.R., Hoshida, Y., Mosquera, J.M., Pawitan, Y., Lee, C. et al. (2007). TMPRSS2:ERG gene fusion associated with lethal prostate cancer in a watchful waiting cohort. *Oncogene* 26(31), 4596–4599.

Dhanasekaran, S.M., Barrette, T.R., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K.J., Rubin, M.A. and Chinnaiyan, A.M. (2001). Delineation of prognostic biomarkers in prostate cancer. *Nature* 412(6849), 822–826.

Dong, B., Kim, S., Hong, S., Das Gupta, J., Malathi, K., Klein, E.A., Ganem, D., Derisi, J.L., Chow, S.A. and Silverman, R.H. (2007). An infectious retrovirus susceptible to an IFN antiviral pathway from human prostate tumors. *Proc Natl Acad Sci USA* 104(5), 1655–1660.

Eastham, J.A., Kelly, W.K., Grossfeld, G.D. and Small, E.J. (2003). Cancer and Leukemia Group B (CALGB) 90203: A randomized phase 3 study of radical prostatectomy alone versus estramustine and docetaxel before radical prostatectomy for patients with high-risk localized disease. *Urology* 62(Suppl 1), 55–62.

Edwards, J., Krishna, N.S., Witton, C.J. and Bartlett, J.M. (2003). Gene amplifications associated with the development of hormone-resistant prostate cancer. *Clin Cancer Res* 9(14), 5271–5281.

Febbo, P.G., Brown, M.A. (2002). Nuclear receptors in normal prostatic growth and disease. In *Prostate Cancer: Principles and Practice* (P.W. Kantoff, P.R. Carroll and A.V. D'Amico, eds), Williams, & Wilkins, Philadelphia, Lippincott, 1, 29–52.

Febbo, P.G. and Sellers, W.R. (2003). Use of expression analysis to predict outcome after radical prostatectomy. *J Urol* 170(6 Pt 2), S11–S19. discussion S19–20.

Febbo, P.G., Mulligan M.G., Slonina, D.A., Stegmaier, K., Di Vizio, D., Martinez, P.R., Loda, M. and Taylor, S.c. (2007). Literature Lab: A method of automated literature interrogation to infer biology from microarray analysis, *BMC Genomics* 8, 461.

Febbo, P.G., Richie, J.P., George, D.J., Loda, M., Manola, J., Shankar, S., Barnes, A.S., Tempany, C., Catalona, W., Kantoff, P.W. et al. (2005). Neoadjuvant docetaxel before radical prostatectomy in patients with high-risk localized prostate cancer. *Clin Cancer Res* 11(14), 5233–5240.

Febbo, P.G., Thorner, A., Rubin, M.A., Loda, M., Kantoff, P.W., Oh, W.K., Golub, T. and George, D. (2006). Application of oligonucleotide microarrays to assess the biological effects of neoadjuvant imatinib mesylate treatment for localized prostate cancer. *Clin Cancer Res* 12(1), 152–158.

Freedland, S.J., Humphreys, E.B., Mangold, L.A., Eisenberger, M., Dorey, F.J., Walsh, P.C. and Partin, A.W. (2005). Risk of prostate cancer-specific mortality following biochemical recurrence after radical prostatectomy. *JAMA* 294(4), 433–439.

Freedman, M.L., Haiman, C.A., Patterson, N., McDonald, G.J., Tandon, A., Waliszewska, A., Penney, K., Steen, R.G., Ardlie, K., John, E.M. et al. (2006). Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proc Natl Acad Sci USA* 103(38), 14068–14073.

Gelmann, E.P. and Semmes, O.J. (2004). Expression of genes and proteins specific for prostate cancer. *J Urol* 172(5 Pt 2), S23–S26, discussion S26-7.

Ghossein, R.A., Carusone, L. and Bhattacharya, S. (1999). Review: Polymerase chain reaction detection of micrometastases and circulating tumor cells: Application to melanoma, prostate, and thyroid carcinomas. *Diagn Mol Pathol* 8(4), 165–175.

Glinsky, G.V., Glinskii, A.B., Stephenson, A.J., Hoffman, R.M. and Gerald, W.L. (2004). Gene expression profiling predicts clinical outcome of prostate cancer. *J Clin Invest* 113(6), 913–923.

Gomez-Zaera, M., Abril, J., Gonzalez, L., Aguilo, F., Condom, E., Nadal, M. and Nunes, V. (2006). Identification of somatic and germline mitochondrial DNA sequence variants in prostate cancer patients. *Mutat Res* 595(1–2), 42–51.

Gudmundsson, J., Sulem, P., Manolescu, A., Amundadottir, L.T., Gudbjartsson, D., Helgason, A., Rafnar, T., Bergthorsson, J.T., Agnarsson, B.A., Baker, A. et al. (2007). Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet* 39(5), 631–637.

Haiman, C.A., Le Marchand, L., Yamamato, J., Stram, D.O., Sheng, X., Kolonel, L.N., Wu, A.H., Reich, D. and Henderson, B.E. (2007a). A common genetic risk factor for colorectal and prostate cancer. *Nat Genet*.

Haiman, C.A., Patterson, N., Freedman, M.L., Myers, S.R., Pike, M.C., Waliszewska, A., Neubauer, J., Tandon, A., Schirmer, C., McDonald, G.J. et al. (2007b). Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet* 39(5), 638–644.

Halvorsen, O.J., Oyan, A.M., Bo, T.H., Olsen, S., Rostad, K., Haukaas, S.A., Bakke, A.M., Marzolf, B., Dimitrov, K., Stordrange, L. et al. (2005). Gene expression profiles in prostate cancer: Association with patient subgroups and tumour differentiation. *Int J Oncol* 26(2), 329–336.

Harashima, K., Akimoto, T., Nonaka, T., Tsuzuki, K., Mitsuhashi, N. and Nakano, T. (2005). Heat shock protein 90 (Hsp90) chaperone complex inhibitor, radicicol, potentiated radiation-induced cell killing in a hormone-sensitive prostate cancer cell line through degradation of the androgen receptor. *Int J Radiat Biol* 81(1), 63–76.

Hemann, M.T., Bric, A., Teruya-Feldstein, J., Herbst, A., Nilsson, J.A., Cordon-Cardo, C., Cleveland, J.L., Tansey, W.P. and Lowe, S.W. (2005). Evasion of the p53 tumour surveillance network by tumour-derived MYC mutants. *Nature* 436(7052), 807–811.

Henshall, S.M., Afar, D.E., Hiller, J., Horvath, L.G., Quinn, D.I., Rasiah, K.K., Gish, K., Willhite, D., Kench, J.G., Gardiner-Garden, M. et al. (2003). Survival analysis of genome-wide gene expression profiles of prostate cancers identifies new prognostic targets of disease relapse. *Cancer Res* 63(14), 4196–4203.

Hermans, K.G., van Marion, R., van Dekken, H., Jenster, G., van Weerden, W.M. and rapman, J. (2006). TMPRSS2:ERG fusion by translocation or interstitial deletion is highly relevant in androgen-dependent prostate cancer, but is bypassed in late-stage androgen receptor-negative prostate cancer. *Cancer Res* 66(22), 10658–10663.

Hessels, D., Smit, F.P., Verhaegh, G.W., Witjes, J.A., Cornel, E.B. and Schalken, J.A. (2007). Detection of TMPRSS2-ERG fusion transcripts and prostate cancer antigen 3 in urinary sediments may improve diagnosis of prostate cancer. *Clin Cancer Res* 13(17), 5103–5108.

Hieronymus, H., Lamb, J., Ross, K.N., Peng, X.P., Clement, C., Rodina, A., Nieto, M., Du, J., Stegmaier, K., Raj, S.M. et al. (2006). Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. *Cancer Cell* 10(4), 321–330.

Hill, R., Song, Y., Cardiff, R.D. and Van Dyke, T. (2005). Selective evolution of stromal mesenchyme with p53 loss in response to epithelial tumorigenesis. *Cell* 123(6), 1001–1011.

Hricak, H. (2005). MR imaging and MR spectroscopic imaging in the pre-treatment evaluation of prostate cancer. *Br J Radiol* 78(Spec No 2), S103–S111.

Huggins, C. and Hodges, C.V. (1941). Studies on prostatic cancer. I. The effect of castration, of estrogen and of androgen injection on serum phosphatases in metastatic carcinoma of the prostate. *Cancer Res* 1, 293–297.

Iljin, K., Wolf, M., Edgren, H., Gupta, S., Kilpinen, S., Skotheim, R.I., Peltola, M., Smit, F., Verhaegh, G., Schalken, J. et al. (2006). TMPRSS2 fusions with oncogenic ETS factors in prostate cancer involve unbalanced genomic rearrangements and are associated with HDAC1 and epigenetic reprogramming. *Cancer Res* 66(21), 10242–10246.

Kawamoto, K., Okino, S.T., Place, R.F., Urakami, S., Hirata, H., Kikuno, N., Kawakami, T., Tanaka, Y., Pookot, D., Chen, Z. et al. (2007). Epigenetic modifications of RASSF1A gene through chromatin remodeling in prostate cancer. *Clin Cancer Res* 13(9), 2541–2548.

Koivisto, P., Hyytinen, E., Palmberg, C., Tammela, T., Visakorpi, T., Isola, J. and Kallioniemi, O.P. (1995). Analysis of genetic changes underlying local recurrence of prostate carcinoma during androgen deprivation therapy. *Am J Pathol* 147(6), 1608–1614.

Krishnan, A.V., Zhao, X.Y., Swami, S., Brive, L., Peehl, D.M., Ely, K.R. and Feldman, D. (2002). A glucocorticoid-responsive mutant androgen receptor exhibits unique ligand specificity: Therapeutic implications for androgen-independent prostate cancer. *Endocrinology* 143(5), 1889–1900.

Lapointe, J., Li, C., Higgins, J.P., van de Rijn, M., Bair, E., Montgomery, K., Ferrari, M., Egevad, L., Rayford, W., Bergerheim, U. et al. (2004). Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc Natl Acad Sci USA* 101(3), 811–816.

Lee, W.H., Morton, R.A., Epstein, J.I., Brooks, J.D., Campbell, P.A., Bova, G.S., Hsieh, W.S., Isaacs, W.B. and Nelson, W.G. (1994). Cytidine methylation of regulatory sequences near the pi-class glutathione S-transferase gene accompanies human prostatic carcinogenesis. *Proc Natl Acad Sci USA* 91(24), 11733–11737.

Li, J., Yen, C., Liaw, D., Podsypanina, K., Bose, S., Wang, S.I., Puc, J., Miliaresis, C., Rodgers, L., McCombie, R. et al. (1997). PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science* 275(5308), 1943–1947.

Li, L.C., Carroll, P.R. and Dahiya, R. (2005). Epigenetic changes in prostate cancer: Implication for diagnosis and treatment. *J Natl Cancer Inst* 97(2), 103–115.

Lieberfarb, M.E., Lin, M., Lechpammer, M., Li, C., Tanenbaum, D.M., Febbo, P.G., Wright, R.L., Shim, J., Kantoff, P.W., Loda, M. et al. (2003). Genome-wide loss of heterozygosity analysis from laser capture microdissected prostate cancer using single nucleotide polymorphic allele (SNP) arrays and a novel bioinformatics platform dChipSNP. *Cancer Res* 63(16), 4781–4785.

Lin, J., Adam, R.M., Santiestevan, E. and Freeman, M.R. (1999). The phosphatidylinositol 3′-kinase pathway is a dominant growth factor-activated cell survival pathway in LNCaP human prostate carcinoma cells. *Cancer Res* 59(12), 2891–2897.

Luo, J., Duggan, D.J., Chen, Y., Sauvageot, J., Ewing, C.M., Bittner, M.L., Trent, J.M. and Isaacs, W.B. (2001). Human prostate cancer and benign prostatic hyperplasia: Molecular dissection by gene expression profiling. *Cancer Res* 61(12), 4683–4688.

Luo, J., Zha, S., Gage, W.R., Dunn, T.A., Hicks, J.L., Bennett, C.J., Ewing, C.M., Platz, E.A., Ferdinandusse, S., Wanders, R.J. et al. (2002). Alpha-methylacyl-CoA racemase: A new molecular marker for prostate cancer. *Cancer Res* 62(8), 2220–2226.

M'Koma, A.E., Blum, D.L., Norris, J.L., Koyama, T., Billheimer, D., Motley, S., Ghiassi, M., Ferdowsi, N., Bhowmick, I., Chang, S.S. et al. (2007). Detection of pre-neoplastic and neoplastic prostate disease by MADI profiling of urine. *Biochem Biophys Research Commun* 353, 829–834.

Mayo, L.D., Dixon, J.E., Durden, D.L., Tonks, N.K. and Donner, D.B. (2002). PTEN protects p53 from Mdm2 and sensitizes cancer cells to chemotherapy. *J Biol Chem* 277(7), 5484–5489.

Moreno, J., DeBono, J.S., Shaffer, D.R., Montgomery, B., Miller, M.C., Tissing, H., Doyle, G., Terstappen, L.W., Pienta, K.J. and Raghavan, D. (2007). *Multi-center study evaluating circulating tumor cells (CTCs) as a surrogate for surivial in men treated for castration refractory prostate cancer (CRPC).* ASCO 2007 Annual Meeting, Chicago, IL.

Moreno, J.G., O'Hara, S.M., Gross, S., Doyle, G., Fritsche, H., Gomella, L.G. and Terstappen, L.W. (2001). Changes in circulating carcinoma cells in patients with metastatic prostate cancer correlate with disease status. *Urology* 58(3), 386–392.

Moreno, J.G., Miller, M.C., Gross, S., Allard, W.J., Gomella, L.G. and Terstappen, L.W. (2005). Circulating tumor cells predict survival in patients with metastatic prostate cancer. *Urology* 65(4), 713–718.

Mosquera, J.M., Perner, S., Demichelis, F., Kim, R., Hofer, M.D., Mertz, K.D., Paris, P.L., Simko, J., Collins, C., Bismar, T.A. et al. (2007). Morphological features of TMPRSS2-ERG gene fusion prostate cancer. *J Pathol* 212(1), 91–101.

Mostaghel, E.A., Page, S.T., Lin, D.W., Fazli, L., Coleman, I.M., True, L.D., Knudsen, B., Hess, D.L., Nelson, C.C., Matsumoto, A.M. et al. (2007). Intraprostatic androgens and androgen-regulated gene expression persist after testosterone suppression: Therapeutic implications for castration-resistant prostate cancer. *Cancer Res* 67(10), 5033–5041.

Murabito, J.M., Rosenberg, C.L., Finger, D., Kreger, B.E., Levy, D., Splansky, G.L., Antman, K. and Hwang, S.J. (2007). A genome-wide association study of breast and prostate cancer in the NHLBI's Framingham Heart Study. *BMC Med Genet* 8(Suppl 1), S6.

Nelson, W.G., De Marzo, A.M. and Isaacs, W.B. (2003). Prostate cancer. *N Engl J Med* 349(4), 366–381.

Nelson, W.G., Yegnasubramanian, S., Agoston, A.T., Bastian, P.J., Lee, B.H., Nakayama, M. and De Marzo, A.M. (2007). Abnormal DNA methylation, epigenetics, and prostate cancer. *Front Biosci* 12, 4254–4266.

Okino, S.T., Pookot, D., Li, L.C., Zhao, H., Urakami, S., Shiina, H., Igawa, M. and Dahiya, R. (2006). Epigenetic inactivation of the dioxin-responsive cytochrome P4501A1 gene in human prostate cancer. *Cancer Res* 66(15), 7420–7428.

Ornstein, D.K. and Tyson, D.R. (2006). Proteomics for the identification of new prostate cancer biomarkers. *Urol Oncol* 24(3), 231–236.

Ouyang, X., DeWeese, T.L., Nelson, W.G. and Abate-Shen, C. (2005). Loss-of-function of Nkx3.1 promotes increased oxidative damage in prostate carcinogenesis. *Cancer Res* 65(15), 6773–6779.

Paris, P.L., Andaya, A., Fridlyand, J., Jain, A.N., Weinberg, V., Kowbel, D., Brebner, J.H., Simko, J., Watson, J.E., Volik, S. et al. (2004). Whole genome scanning identifies genotypes associated with recurrence and metastasis in prostate tumors. *Hum Mol Genet* 13(13), 1303–1313.

Parsons, J.K., Nelson, C.P., Gage, W.R., Nelson, W.G., Kensler, T.W. and De Marzo, A.M. (2001). GSTA1 expression in normal, preneoplastic, and neoplastic human prostate tissue. *Prostate* 49(1), 30–37.

Partin, A.W., Pearson, J.D., Landis, P.K., Carter, H.B., Pound, C.R., Clemens, J.Q., Epstein, J.I. and Walsh, P.C. (1994). Evaluation of serum prostate-specific antigen velocity after radical prostatectomy to distinguish local recurrence from distant metastases. *Urology* 43(5), 649–659.

Petricoin, E.F., 3rd, Hackett, J.L., Lesko, L.J., Puri, R.K., Gutman, S.I., Chumakov, K., Woodcock, J., Feigal, D.W., Jr, Zoon, K.C. and Sistare, F.D. (2002). Medical applications of microarray technologies: A regulatory science perspective. *Nat Genet* 32(Suppl), 474–479.

Petros, J.A., Baumann, A.K., Ruiz-Pesini, E., Amin, M.B., Sun, C.Q., Hall, J., Lim, S., Issa, M.M., Flanders, W.D., Hosseini, S.H. et al. (2005). mtDNA mutations increase tumorigenicity in prostate cancer. *Proc Natl Acad Sci USA* 102(3), 719–724.

Petrylak, D.P., Tangen, C.M., Hussain, M.H., Lara, P.N., Jr, Jones, J.A., Taplin, M.E., Burch, P.A., Berry, D., Moinpour, C., Kohli, M. et al. (2004). Docetaxel and estramustine compared with mitoxantrone and prednisone for advanced refractory prostate cancer. *N Engl J Med* 351(15), 1513–1520.

Phillips, S.M., Barton, C.M., Lee, S.J., Morton, D.G., Wallace, D.M., Lemoine, N.R. and Neoptolemos, J.P. (1994). Loss of the retinoblastoma susceptibility gene (RB1) is a frequent and early event in prostatic tumorigenesis. *Br J Cancer* 70(6), 1252–1257.

Pomerantz, M.M., Freedman, M.L. and Kantoff, P.W. (2007). Genetic determinants of prostate cancer risk. *BJU Int* 100(2), 241–243.

Potti, A., Dressman, H.K., Bild, A., Riedel, R.F., Chan, G., Sayer, R., Cragun, J., Cottrill, H., Kelley, M.J., Petersen, R. et al. (2006). Genomic signatures to guide the use of chemotherapeutics. *Nat Med* 12(11), 1294–1300.

Pulukuri, S.M. and Rao, J.S. (2006). CpG island promoter methylation and silencing of 14-3-3sigma gene expression in LNCaP and Tramp-C1 prostate cancer cell lines is associated with methyl-CpG-binding protein MBD2. *Oncogene* 25(33), 4559–4572.

Pulukuri, S.M., Patibandla, S., Patel, J., Estes, N. and Rao, J.S. (2007). Epigenetic inactivation of the tissue inhibitor of metalloproteinase-2 (TIMP-2) gene in human prostate tumors. *Oncogene* 26(36), 5229–5237.

Ramaswamy, S., K. Ross, N., Lander, E.S. and Golub, T.R. (2003). A molecular signature of metastasis in primary solid tumors. *Nat Genet* 33(1), 49–54.

Rehman, I., Azzouzi, A.R., Catto, J.W., Allen, S., Cross, S.S., Feeley, K., Meuth, M. and Hamdy, F.C. (2004). Proteomic analysis of voided urine after prostatic massage from patients with prostate cancer: A pilot study. *Urology* 64(6), 1238–1243.

Rhodes, D.R., Sanda, M.G., Otte, A.P., Chinnaiyan, A.M. and Rubin, M.A. (2003). Multiplex biomarker approach for determining risk of prostate-specific antigen-defined recurrence of prostate cancer. *J Natl Cancer Inst* 95(9), 661–668.

Rogers, C.G., Yan, G., Zha, S., Gonzalgo, M.L., Isaacs, W.B., Luo, J., De Marzo, A.M., Nelson, W.G. and Pavlovich, C.P. (2004). Prostate cancer detection on urinalysis for alpha methylacyl coenzyme a racemase protein. *J Urol* 172(4 Pt 1), 1501–1503.

Rogers, C.G., Gonzalgo, M.L., Yan, G., Bastian, P.J., Chan, D.Y., Nelson, W.G. and Pavlovich, C.P. (2006). High concordance of gene methylation in post-digital rectal examination and post-biopsy urine samples for prostate cancer detection. *J Urol* 176(5), 2280–2284.

Rubin, M.A., Zhou, M., Dhanasekaran, S.M., Varambally, S., Barrette, T.R., Sanda, M.G., Pienta, K.J., Ghosh, D. and Chinnaiyan, A.M. (2002). alpha-Methylacyl coenzyme A racemase as a tissue biomarker for prostate cancer. *JAMA* 287(13), 1662–1670.

Sansal, I. and Sellers, W.R. (2004). The biology and clinical relevance of the PTEN tumor suppressor pathway. *J Clin Oncol* 22(14), 2954–2963.

Saramaki, O.R., Porkka, K.P., Vessella, R.L. and Visakorpi, T. (2006). Genetic aberrations in prostate cancer by microarray analysis. *Int J Cancer* 119(6), 1322–1329.

Schaid, D.J. (2004). The complex genetic epidemiology of prostate cancer. *Hum Mol Genet* 13(Spec No 1), R103–R121.

Scher, H.I. and Sawyers, C.L. (2005). Biology of progressive, castration-resistant prostate cancer: Directed therapies targeting the androgen-receptor signaling axis. *J Clin Oncol* 23(32), 8253–8261.

Schumacher, F.R., Feigelson, H.S., Cox, D.G., Haiman, C.A., Albanes, D., Buring, J., Calle, E.E., Chanock, S.J., Colditz, G.A., Diver, W.R. et al. (2007). A common 8q24 variant in prostate and breast cancer from a large nested case-control study. *Cancer Res* 67(7), 2951–2956.

Seiden, M.V., Kantoff, P.W., Krithivas, K., Propert, K., Bryant, M., Haltom, E., Gaynes, L., Kaplan, I., Bubley, G., DeWolf, W. et al. (1994). Detection of circulating tumor cells in men with localized prostate cancer. *J Clin Oncol* 12(12), 2634–2639.

Seligson, D.B., Horvath, S., Shi, T., Yu, H., Tze, S., Grunstein, M. and Kurdistani, S.K. (2005). Global histone modification patterns predict risk of prostate cancer recurrence. *Nature* 435(7046), 1262–1266.

Shaffer, D.R., Leversha, M.A., Danila, D.C., Lin, O., Gonzalez-Espinoza, R., Gu, B., Anand, A., Smith, K., Maslak, P., Doyle, G.V. et al. (2007). Circulating tumor cell analysis in patients with progressive castration-resistant prostate cancer. *Clin Cancer Res* 13(7), 2023–2029.

Singh, A.S., Chau, C.H., Price, D.K. and Figg, W.D. (2005). Mechanisms of disease: Polymorphisms of androgen regulatory genes in the development of prostate cancer. *Nat Clin Pract Urol* 2(2), 101–107.

Singh, D., Febbo, P.G., Ross, K., Jackson, D.G., Manola, J., Ladd, C., Tamayo, P., Renshaw, A.A., D'Amico, A.V., Richie, J.P. et al. (2002). Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* 1(2), 203–209.

Solit, D.B., Zheng, F.F., Drobnjak, M., Munster, P.N., Higgins, B., Verbel, D., Heller, G., Tong, W., Cordon-Cardo, C., Agus, D.B. et al. (2002). 17-Allylamino-17-demethoxygeldanamycin induces the degradation of androgen receptor and HER-2/neu and inhibits the growth of prostate cancer xenografts. *Clin Cancer Res* 8(5), 986–993.

Solit, D.B., Scher, H.I. and Rosen, N. (2003). Hsp90 as a therapeutic target in prostate cancer. *Semin Oncol* 30(5), 709–716.

Sorensen, A.G. (2006). Magnetic resonance as a cancer imaging biomarker. *J Clin Oncol* 24(20), 3274–3281.

Stambolic, V., MacPherson, D., Sas, D., Lin, Y., Snow, B., Jang, Y., Benchimol, S. and Mak, T.W. (2001). Regulation of PTEN transcription by p53. *Mol Cell* 8(2), 317–325.

Stanbrough, M., Bubley, G.J., Ross, K., Golub, T.R., Rubin, M.A., Penning, T.M., Febbo, P.G. and Balk, S.P. (2006). Increased expression of genes converting adrenal androgens to testosterone in androgen-independent prostate cancer. *Cancer Res* 66(5), 2815–2825.

Subramanian, A., Tanayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. (in press).

Tannock, I.F., de Wit, R., Berry, W.R., Horti, J., Pluzanska, A., Chi, K.N., Oudard, S., Theodore, C., James, N.D., Turesson, I. et al. (2004). Docetaxel plus prednisone or mitoxantrone plus prednisone for advanced prostate cancer. *N Engl J Med* 351(15), 1502–1512.

Taplin, M.E. and Balk, S.P. (2004). Androgen receptor: A key molecule in the progression of prostate cancer to hormone independence. *J Cell Biochem* 91(3), 483–490.

Taplin, M.E., Bubley, G.J., Shuster, T.D., Frantz, M.E., Spooner, A.E., Ogata, G.K., Keer, H.N. and Balk, S.P. (1995). Mutation of the androgen-receptor gene in metastatic androgen-independent prostate cancer. *N Engl J Med* 332(21), 1393–1398.

Taplin, M.E., Bubley, G.J., Ko, Y.J., Small, E.J., Upton, M., Rajeshkumar, B. and Balk, S.P. (1999). Selection for androgen receptor mutations in prostate cancers treated with androgen antagonist. *Cancer Res* 59(11), 2511–2515.

Theodorescu, D., Broder, S.R., Boyd, J.C., Mills, S.E. and Frierson, H.F., Jr (1997). p53, bcl-2 and retinoblastoma proteins as long-term prognostic markers in localized carcinoma of the prostate. *J Urol* 158(1), 131–137.

Tinzl, M., Marberger, M., Horvath, S. and Chypre, C. (2004). DD3PCA3 RNA analysis in urine – a new perspective for detecting prostate cancer. *Eur Urol* 46(2), 182–186. discussion 187.

Tomlins, S.A., Rhodes, D.R., Perner, S., Dhanasekaran, S.M., Mehra, R., Sun, X.W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R. et al. (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* 310(5748), 644–648.

Tomlins, S.A., Mehra, R., Rhodes, D.R., Cao, X., Wang, L., Dhanasekaran, S.M., Kalyana-Sundaram, S., Wei, J.T., Rubin, M.A., Pienta, K.J. et al. (2007). Integrative molecular concept modeling of prostate cancer progression. *Nat Genet* 39(1), 41–51.

Trotman, L.C., Niki, M., Dotan, Z.A., Koutcher, J.A., Di Cristofano, A., Xiao, A., Khoo, A.S., Roy-Burman, P., Greenberg, N.M., Van Dyke, T. et al. (2003). Pten dose dictates cancer progression in the prostate. *PLoS Biol* 1(3), E59.

True, L., Coleman, I., Hawley, S., Huang, C.Y., Gifford, D., Coleman, R., Beer, T.M., Gelmann, E., Datta, M., Mostaghel, E. et al. (2006). A molecular correlate to the Gleason grading system for prostate adenocarcinoma. *Proc Natl Acad Sci USA* 103(29), 10991–10996.

Tu, J.J., Rohan, S., Kao, J., Kitabayashi, N., Mathew, S. and Chen, Y.T. (2007). Gene fusions between TMPRSS2 and ETS family genes in prostate cancer: frequency and transcript variant analysis by RT-PCR and FISH on paraffin-embedded tissues. *Mod Pathol* 20, 921–928.

Urisman, A., Molinaro, R.J., Fischer, N., Plummer, S.J., Casey, G., Klein, E.A., Malathi, K., Magi-Galluzzi, C., Tubbs, R.R., Ganem, D. et al. (2006). Identification of a novel Gammaretrovirus in prostate tumors of patients homozygous for R462Q RNASEL variant. *PLoS Pathog* 2(3), e25.

Varambally, S., Dhanasekaran, S.M., Zhou, M., Barrette, T.R., Kumar-Sinha, C., Sanda, M.G., Ghosh, D., Pienta, K.J., Sewalt, R.G., Otte, A.P. et al. (2002). The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* 419(6907), 624–629.

Varambally, S., Yu, J., Laxman, B., Rhodes, D.R., Mehra, R., Tomlins, S.A., Shah, R.B., Chandran, U., Monzon, F.A., Becich, M.J. et al. (2005). Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. *Cancer Cell* 8(5), 393–406.

Villanueva, J., Shaffer, D.R., Philip, J., Chaparro, C.A., Erdjument-Bromage, H., Olshen, A.B., Fleisher, M., Lilja, H., Brogi, E., Boyd, J. et al. (2006). Differential exoprotease activities confer tumor-specific serum peptidome patterns. *J Clin Invest* 116(1), 271–284.

Visakorpi, T., Hyytinen, E., Koivisto, P., Tanner, M., Keinanen, R., Palmberg, C., Palotie, A., Tammela, T., Isola, J. and Kallioniemi, O.P. (1995). *In vivo* amplification of the androgen receptor gene and progression of human prostate cancer. *Nat Genet* 9(4), 401–406.

Wang, L., Hricak, H., Kattan, M.W., Chen, H.N., Scardino, P.T. and Kuroiwa, K. (2006). Prediction of organ-confined prostate cancer: Incremental value of MR imaging and MR spectroscopic imaging to staging nomograms. *Radiology* 238(2), 597–603.

Wang, L., McDonnell, S.K., Slusser, J.P., Hebbring, S.J., Cunningham, J.M., Jacobsen, S.J., Cerhan, J.R., Blute, M.L., Schaid, D.J. and Thibodeau, S.N. (2007). Two common chromosome 8q24 variants are associated with increased risk for prostate cancer. *Cancer Res* 67(7), 2944–2950.

Wang, S., Gao, J., Lei, Q., Rozengurt, N., Pritchard, C., Jiao, J., Thomas, G.V., Li, G., Roy-Burman, P., Nelson, P.S. et al. (2003). Prostate-specific deletion of the murine Pten tumor suppressor gene leads to metastatic prostate cancer. *Cancer Cell* 4(3), 209–221.

Wang, X., Yu, J., Sreekumar, A., Varambally, S., Shen, R., Giacherio, D., Mehra, R., Montie, J.E., Pienta, K.J., Sanda, M.G. et al. (2005). Autoantibody signatures in prostate cancer. *N Engl J Med* 353(12), 1224–1235.

Welsh, J.B., Sapinoso, L.M., Su, A.I., Kern, S.G., Wang-Rodriguez, J., Moskaluk, C.A., Frierson, H.F., Jr and Hampton, G.M. (2001). Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. *Cancer Res* 61(16), 5974–5978.

Yeager, M., Orr, N., Hayes, R.B., Jacobs, K.B., Kraft, P., Wacholder, S., Minichiello, M.J., Fearnhead, P., Yu, K., Chatterjee, N. et al. (2007). Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet* 39(5), 645–649.

Zhao, X.Y., Boyle, B., Krishnan, A.V., Navone, N.M., Peehl, D.M. and Feldman, D. (1999). Two mutations identified in the androgen receptor of the new human prostate cancer cell line MDA PCa 2a. *J Urol* 162(6), 2192–2199.

CHAPTER

Genomic Assessment of Ovarian Cancer

Tanja Pejovic, Matthew L. Anderson and Kunle Odunsi

INTRODUCTION

The normal human ovary is lined by a cuboidal epithelium. These cells remain quiescent for most of their lifespan, proliferating only to repair the ovarian surface after ovulation. Epidemiologic evidence demonstrates a clear correlation between the frequency of ovulation and the incidence of epithelial ovarian cancer. As a result, particular attention in the search for precursors to ovarian cancer has focused on small inclusion clefts that persist after repair of the ovarian surface. The molecular environment influencing these clefts is unique, the epithelia lining these clefts express specific cell adhesion molecules, including cadherins, not normally observed in other areas of the ovarian surface epithelium.

While in other organs, such as colon, distinct pre-malignant lesions have been identified and found to accumulate, genetic defects that ultimately result in the pre-malignant ovarian change has not yet been clearly identified. Histologic findings consistent with a pre-invasive lesion for ovarian cancer have been described by a number of investigators in ovaries from high-risk women undergoing prophylactic oophorectomy and in areas of ovarian epithelium adjacent to early stage ovarian cancers that demonstrate a transition from normal to malignant cells (Schlosshauer et al., 2003). The hypothesis that these lesions are pre-malignant is strengthened by observations that regions of epithelial irregularity express levels of p53 and Ki-67 intermediate between those found in normal ovary epithelium and ovarian cancers. Each of these observations is consistent with the hypothesis that, similar to cancers originating in other organs, ovarian cancer evolves from an intraepithelial precursor. If so, improved means to detect and/or eradicate these lesions may prove fruitful for preventing ovarian cancer. Research efforts, designed to improve the early detection of ovarian cancer or optimize its clinical management at later stages, are likely to be more successful based on the understanding of molecular events responsible for this disease.

INHERITED OVARIAN CANCER SYNDROMES

Linkage analysis of familial breast and ovarian cancers provided some of the first insights into the molecular basis of ovarian cancer. These efforts ultimately identified two genes, *BRCA1* and *BRCA2*, each clearly associated with an increased incidence of ovarian cancer (Miki et al., 1994; Wooster et al., 1994). Although only a minority (8–10%) of diagnosed ovarian cancers are familial, about 60% of familial ovarian cancers are associated with mutations at the *BRCA1* locus, located on 17q21 (Easton et al., 1995). Hundreds of mutations in *BRCA1* have been identified, most commonly loss of function nonsense or frameshift mutations. Two of the three specific founder mutations, 185delAG and 5382insC, are found in 1% and 0.1% of Ashkenazi Jewish women. The population-based studies have suggested that the lifetime risk of ovarian cancer in *BRCA1* mutation carriers is about 20–30%, but this increased risk is not manifest until the age of 40.

The functional consequences of loss of function mutations in *BRCA1* are potentially profound. *BRCA1* regulates p53, an

Copyright © 2010, Elsevier Inc.
All rights reserved.

oncogene frequently implicated in ovarian cancer. Thus, loss of BRCA1 function allows DNA damage to accumulate via a loss of its activation of p53 check point. Other mechanisms of BRCA1-associated ovarian carcinogenesis such as site-specific dysregulation of X-linked gene expression in BRCA1-associated epithelial ovarian malignancies have been described.

BRCA2 gene plays a role in DNA repair by homologous recombination. About 35% of hereditary ovarian cancers may be attributed to *BRCA2* mutations and both female and male carriers of the mutation are at risk for breast cancer. The third founder mutation 6174delT occurs in *BRCA2*. Mutations in *BRCA1* and *BRCA2* are only rarely observed in sporadic, non-familial ovarian cancers. Characterization of genome-wide patterns of gene expression in sporadic breast cancers has allowed investigators to classify these tumors as either *BRCA1*-like or *BRCA2* like in the patterns of their gene expression. These observations implicate alterations in other components of the BRCA1/2-regulated pathways that contribute to sporadic ovarian cancer.

OPTIONS FOR SCREENING AND PREVENTION

The standard recommendations for women at increased risk for ovarian cancer include pelvic examination, serum tumor markers, and pelvic ultrasound. Unfortunately, none of these strategies have proven effective as a screening modality. Most cancers detected in high-risk populations are detected at an advanced stage. The high risk combined with the ineffectiveness of current screening methods has led to the recommendation that women with high risk undergo risk-reducing salpingo-oophorectomy (RRSO) after completion of child-bearing.

To be considered at high-risk for ovarian cancer, subjects must satisfy one of the following criteria (Table 38.1) (US Preventive Services Task Force recommendations, 2005).

Unfortunately risk-reducing surgery is not fail-safe. Cases of peritoneal surface serous carcinoma have been described in 2% of cases occurring from 1 to 27 years after risk reducing oophorectomy (Offit, 1998).

TABLE 38.1 Criteria for high-risk for ovarian cancer population

1.	The family of the subject has a documented deleterious *BRCA1* or *BRCA2* mutation – in either the subject herself, or a first- or second-degree relative.
OR	
2.	For non-Ashkenazi Jewish women: The family of the subject contains at least two ovarian and/or breast cancers among the first- or second-degree relatives within the same lineage. This condition can be satisfied by multiple primary cancers in the same person. Where breast cancer is required to meet this criterion, at least one breast cancer must have been diagnosed at the age of 50 years or younger.
OR	
3.	For non-Ashkenazi Jewish women: A combination of three or more first- or second-degree relatives with breast cancer regardless of age at diagnosis.
OR	
4.	For non-Ashkenazi Jewish women: A first-degree relative with bilateral breast cancer.
OR	
5.	For non-Ashkenazi Jewish women: A history of breast cancer in a male relative.
OR	
6.	For women of Ashkenazi Jewish heritage: Any first-degree relative or two second-degree relatives within the same lineage with breast or ovarian cancer.
OR	
7.	The subject is of Ashkenazi Jewish heritage and has had breast cancer herself. To meet this criterion, her breast cancer must have been diagnosed at the age of 50 years or younger.

GENOMIC INSTABILITY AND OVARIAN CANCER

Genomic instability is a hallmark of all cancer, including epithelial ovarian cancers. In order to become genetically unstable the cell has to become intolerant to DNA damage. The cell can achieve this by inactivating in any of the six major DNA repair pathways: base excision repair (BER), mismatch repair (MRM), nucleotide excision repair (NER), homologous recombination (HR), non-homologous recombination (NHR), and translesion DNA synthesis (TLS). The specific DNA pathway affected often predicts the specific type of mutations observed in particular cancers, its sensitivity to drugs, as well as clinical outcome of affected patients.

SOMATIC MUTATIONS IN OVARIAN CANCER

Ovarian carcinoma is a monoclonal disease that progresses through a series of genetic alterations that successively accumulate and transform a normal to a neoplastic cell. The molecular complexity of ovarian cancer has become apparent through the use of novel technologies that explore the genome, transcriptome, and proteome. These novel high throughput technologies are expected to provide information that allows for earlier diagnosis and individualized treatment of ovarian cancer. Biomarkers could be also used to stratify ovarian cancer into disease groups of varying aggressiveness. Stage III and IV ovarian cancer has a consistently high mortality rate, while Stage I and II ovarian cancer has a widely variable rate of mortality. It may be possible to

stratify ovarian cancer in its earliest stages into groups of cancer with varying degrees of aggressiveness and subsequent mortality.

Gene Expression Profiling of Ovarian Cancer

Ovarian cancers have been subjected to transcriptional profiling using *complementary DNA* (*cDNA*) or *oligonucleotide* array types. Alterations in gene expression have been suggested as useful markers of histologic differentiation in ovarian cancer. Marquez and coworkers (2005) showed distinct expression profiles of different ovarian cancer histologic subtypes and identified several markers of mucinous tumors, including *TFF1*, *AGR2*, *LGALS4*, *CEACAM6*, and *CTSE*. Several of these genes are also upregulated in normal colon and are important for protection and healing of the human gastrointestinal tract (Devine et al., 2000). In addition, Wamunyokoli et al. (2006) identified upregulation of *NET1* and *ERBB3* in low-malignant potential mucinous ovarian tumors and invasive mucinous carcinoma of the ovary leading to the hypothesis that these two genes may participate in the initiation of the transformation process in mucinous ovarian cancer. *ERB-B2/HER-2/NEU* was also found to be differentially upregulated in clear cell ovarian cancer. This gene encodes the target for the humanized anti-HER2/neu antibody, trastuzumab (Herceptin), that is showing promise for treatment of patients with ovarian cancers showing overexpression of Her-2/neu protein (Wang et al., 2006).

Recently, gene expressing profiling has successfully been used to predict response to platinum-based chemotherapy (Dressman et al., 2007). The accuracy of detecting platinum resistant disease based on expression profiling was 89%. Within this group expression signatures consistent with SRC and Rb/E2F pathways were identified and successfully targeted *in vitro*.

Proteomics

Proteomics is the study of the proteome of a population of cells. The proteome is a result of many factors including DNA alterations, mRNA splicing, post-translational modifications, and functional regulation of gene expression (An et al., 2006; Banderal et al., 2003). Technology platforms incorporating mass spectrometry (MS) for proteomic biomarker discovery include both pattern-based methods that produce MS-derived protein pattern via SELDI (surface-enhanced laser desorption and ionization), MALDI, or electrospray and identity-based methods that yield lists of sequence-identified peptides from LC-MS/MS analysis of proteolytically digested proteins (Zhang et al., 2006). Pattern and identity combine in MS/MS analysis of selected spots from differential protein displays such as two-dimensional polyacrylamide gel electrophoresis (2D-PAGE). Each method has strengths and limitations. The SELDI technique was used by Petricoin et al. (2002) for the search for ovarian cancer detection markers. This group studied serum samples of 50 ovarian cancer patients and 50 unaffected women. A protein pattern specific to ovarian cancer was identified and applied to a set of 116 serum samples from 50 women with cancer and 66 unaffected women. The proteomic pattern correctly identified all patients including 18 Stage I cancers. Only 3 of the 66 unaffected women were wrongly diagnosed as having ovarian cancer. This study awaits further validation.

ONCOGENES AND GROWTH FACTORS

Increased mutagenic signaling by receptor tyrosine kinases play a major role in ovarian carcinogenesis. Overexpression of *ERBB1*, *ERBB2/HER2/neu*, and *c-FMS* has been reported repeatedly in ovarian cancer. One of the major downstream mediators of signaling initiated by these receptors is the phosphatidylinositiol 3-kinase (PIK3K)-AKT/mammalian target of rapamycin (mTOR) pathway. This pathway is activated by multiple genomic aberrations in up to 70% of ovarian carcinomas. For example, with the *PIK3CA* gene at chromosome 3q26 being amplified in 25–40% of the cases. This proto-oncogene encodes the p110α catalytic subunit of PI3K and, when amplified or mutated, activates signaling through the PI3K/AKT/mTOR pathway. This pathway is thought to be a critical target for therapy of ovarian cancer. Indeed PI3K inhibitor molecules decrease ovarian cancer proliferation and ascites formation and increase chemotherapy induced apoptosis (Table 38.2).

It has been reported that >75% of ovarian carcinomas are resistant to transforming growth factor-β (TGF-β) (Hu et al., 2000), and the loss of TGF-β responsiveness may play an important role in the pathogenesis and/or progression of ovarian cancer. In addition, it has been shown that TGF-β_1, the TGF-β receptors (TβR-II and TβR-I), and the TGF-β-signaling component Smad2 are altered in ovarian cancer (Wang et al., 2006).

TABLE 38.2 Genetic alterations in sporadic ovarian carcinoma

Gene	Function	Mechanism	Frequency (%)
HER2/neu	Tyrosine kinase	Amplification/ overexpression	20–30
EGFR	Tyrosine kinase	Amplification/ overexpression	12–17
KRAS	G-protein	Mutation	15–60
CMYC	Transcription factor	Overexpression	30
PIK3CA	Kinase	Amplification	25–40
AKT2	Kinase	Amplification	10
p53	Tumor suppressor	Mutation/ deletion Overexpression	30–60
p16	Tumor suppressor	Homozygous deletion	15

Alterations in TβR-II have been identified in 25% of ovarian carcinomas (Lynch et al., 1998) whereas mutations in TβR-I were reported in 33% of such cancers (Chen et al., 2001). Loss of function mutations of TGF-β1, TβR-I, and TβR-II can lead to disruption of TGF-β-signaling pathways and subsequent loss of cell cycle control novel TGF-β-signaling component, termed km23. Sequence alterations in human km23 in epithelial ovarian cancer were found in 42% of ovarian carcinomas (Ding et al., 2005). The expression of IGF-II and IGFBP-3 also vary substantially by clinical and pathologic features of ovarian cancer. High levels of IGF-II expression are associated with unfavorable prognostic indicators of the disease, whereas high expression of IGFBP-3 is related to favorable prognostic variables. Patients with high IGF-II expression tend to have higher risk of disease progression and death regardless of the level of IGFBP-3 expression (Lu et al., 2006). However, the IGF-II does not seem to be an independent marker for ovarian cancer prognosis.

Ras pathway also appears to be activated in ovarian cancers. The KRAS is mutated in 15% of borderline serous ovarian tumors and 47% of mucinous neoplasm. While the mutation rate is higher in borderline mucinous (30–60%) than in invasive mucinous cancers (19%), it appears that these mutations play a role in the progression of the mucinous adenoma to borderline tumor to invasive mucinous carcinoma (Ren et al., 2006).

KRAS activates PKCi on 3q26.2, a key regulator of a complex required for localization of E-cadherin to cell junctions, for maintenance of tight junctions and for cell polarity, loss of which would allow aberrant interactions of cell signaling molecules and autocrine cell activation (Macara, 2004). In ovarian cancer PKCi is associated with cyclin E induction. Amplification of the CMYC oncogene also occurs in about 30% of ovarian cancers, however the functional significance of this amplification is unclear if overexpression of the corresponding protein occurs at the same time.

TUMOR SUPPRESSOR GENES

Loss of tumor suppressor gene function plays a major role in the development of most cancers. This involves a two-step process in which both copies of tumor suppressor gene allele are inactivated. In most cases there is mutation of one allele of the gene and loss of the other allele due to deletion of the large region of the chromosome where the gene localizes. Some tumor suppressor genes are inactivated via epigenetic changes, such as hypermethylation of GpC islands in the promoter area, while the gene structure remains intact.

Aberrations of p53 tumor suppressor gene are the most frequent genetic event in ovarian carcinoma (Berchuck et al., 1994). The frequency of p53 mutations is higher in advanced versus early ovarian carcinomas (50–60% versus 10–20%) and is uncommon in borderline ovarian tumors.

While in most cases of ovarian cancer, single amino acid change in the DNA binding domain of p53 (missense mutation) result in overexpression of non-functional protein, 20% of advanced stage ovarian carcinoma contain mutations that result in a truncated protein that is not over-expressed.

Other tumor suppressor genes known to be inactivated in ovarian cancer include *PTEN* and p16 (Table 38.2). *PTEN* is inactivated in some endometrioid ovarian carcinomas (Obata et al., 1998). In addition p16, an inhibitor of cyclin-dependent kinases, may be inactivated via homozygous deletion or promoter methylation in a small fraction of ovarian cancer (Havrilesky et al., 2001).

EPIGENETICS IN OVARIAN CARCINOGENESIS

It has become increasingly apparent that epigenetic events can lead to cancer as frequently as loss of gene function due to mutations or loss of heterozygosity. The overall level of genomic methylation is reduced in cancer (global hypomethylation), but hypermethylation of promoter regions of specific genes is a common event (Jones and Baylin, 2002) that is often associated with transcriptional inactivation of specific genes (Costello et al., 2000). This is critical because the silenced genes are often tumor suppressor genes and their loss of function can be evident in early stages of cancer, and can also drive neoplastic progression and metastasis. Epigenetic gene silencing is a complex series of events that includes DNA hypermethylation of CpG-islands within gene promoter regions, histone deacetylation, methylation or phosphorylation, recruitment of methyl-binding domain proteins and other chromatin remodeling factors to suppress gene transcription. Global hypermethylation of CpG islands appears to be prevalent but highly variable in ovarian cancer tissue (Ahluwalia et al., 2001). Specific aberrant methylation of cancer-associated genes such as *p16INK4A*, *RASSF1A*, *BRCA1*, and *hMLH1* have been reported for ovarian tumors or cell lines, albeit in various degrees and in a non-tumor type-dependent fashion (Esteller et al., 2000; Rathi et al., 2002). While a higher degree of DNA methylation is associated with drug resistance and believed to be the reason for treatment failure and death of 90% of patients with metastatic disease, the demethylation activity of chemotherapeutic drugs can elevate the expression of proteins such as MDR1 (Bell et al., 1985) that lead to a more frequent disease recurrence after chemotherapy (Wei et al., 2002). Thus, the specificity of demethylation of select genes is important to ensure the success of treatment and prevent disease recurrence.

OVARIAN CANCER METASTASES

Metastasis is the functional hallmark of all cancer. In contrast to cancers where metastasis clearly depends on the ability of cells to invade blood or lymphatic vessels, direct dissemination within the peritoneal cavity plays a critical role in the progression of ovarian cancer. A wide variety of gene products have been implicated in the metastasis of ovarian cancer. These include growth factor receptors such as epidermal growth factor receptor (EGFR), insulin-like growth factor receptors (IGFRs) and kinases, such as

jak/stat, focal adhesion kinase, PI-3 kinase and c-met. Comparisons of primary and metastatic ovarian cancers by transcriptional profiling have failed to reveal significant differences in the expression of gene products likely related to the metastatic process.

Particular attention has recently focused on the role of lysophospatidic acid (LPA) in promoting the metastasis of ovarian cancers. LPA is constitutively produced by mesothelial cells lining the peritoneal cavity; its levels are increased in the ascites of women with both early and late stage ovarian cancers (Ren et al., 2006). At a molecular level, exogenous LPA enhances ovarian cancer invasiveness both by activating matrix metalloproteinase-2 via membrane-type-1-matrix metalloproteinase (MT1-MMP) and downregulating the expression of specific tissue inhibitors of metalloproteinases (TIMP-2 and -3) (Sengupta et al., 2006). LPA also promotes dissemination of ovarian cancer by loss of cell adhesion (Do et al., 2007). However, LPA has also been shown to promote the invasiveness of ovarian cancers by additional mechanisms dependent on interleukin-8 (IL-8) (So et al., 2004). The G12/13-RhoA and cyclooxygenase pathways have also been implicated in the LPA – induced migration of ovarian cancers. These mechanisms appear to be independent of the ability of LPA to induce changes in MMP2 expression.

Until recently, the metastasis of ovarian cancer has been almost exclusively studied as a process involving individual cells. However, multicellular clusters of self-adherent cells, known as spheroids, can be isolated from the ascitic fluid of women with ovarian cancer. Spheroids readily adhere to both extracellular matrix proteins, such as collagen IV, and mesothelial cells in monolayer culture using beta1 integrins. Once adherent, the cells contained in spheroids disaggregate, allowing them to invade underlying mesothelial cells and create invasive foci. These observations are consistent with the hypothesis that ovarian cancer spheroids play an important role in the metastasis of ovarian cancer. Recent evidence has shown that a loss of circulating gonadotropins result in a dose-dependent decrease in the expression of VEGF in their outer proliferating cells of ovarian cancer spheroids, that these cell cluster remain responsive to signals in their microenvironment that may further promote metastasis (Schiffenbauer et al., 1997).

The presence of spheroids in ascites may also help to explain the frequent persistence and frequent recurrence of ovarian cancer after treatment. Spheroids express high levels of p27 and P-glycoprotein which contribute, at least in part, to their relative resistance to the cytotoxic effects of paclitaxel when compared to ovarian cancer cells in monolayer culture (Xing et al., 2007). However, the mechanisms by which the aggregation of malignant cells promote or enhance cell survival remain unclear. However, these observations are consistent with *in vitro* studies that demonstrate that the signals generated by adhesion to specific components of the extracellular matrix, such as collagen IV, can modify the sensitivity of ovarian cancers to chemotherapy.

It is also unclear, at present, how the aggregation of these malignant cells might promote or enhance the migration, attachment or invasion of ovarian cancer cells. However, insight into these questions is likely to come from genetic models, such as the migration of the border cell cluster in *Drosophila*. Analyses of border cell migration indicate specific shifts in epithelial polarity and changes in the patterns of signals arising at junctional proteins are necessary for the invasion and migration of epithelial clusters. Signals arising from these junctional proteins appear to be integrated by specific steroid receptor coactivators, known as AIB1 (SRC3). Overexpression of AIB1 is a frequent feature of ovarian cancers suggesting that the pathways regulated by this transcriptional coactivators may indeed play a critical role in promoting ovarian cancer metastasis.

2009 UPDATE

MicroRNAs (miRNAs) and short interfering RNAs (siRNAs) are classes of regulatory small RNA molecules, ranging from 18 to 24 nucleotides in length, whose roles in cancer are becoming increasingly recognized. They function by altering the stability or translational efficiency of messenger RNAs (mRNAs) with which they share sequence complementarities and are predicted to affect about one-third of all human genes.

MicroRNAs

miRNAs represent a class of small, 22 nucleotide long, noncoding RNAs that control gene expression by targeting mRNAs and triggering either translation repression or RNA degradation in the cell (Lagos-Qintana et al., 2001). They differ from siRNA in that they are processed from single-stranded RNA precursors and show only partial complementarities to mRNA targets. The human genome contains about 1000 miRNAs (Zamore and Haley, 2005). Each miRNA may target several different genes simultaneously or in sequence (Pillai et al., 2007; Wang et al., 2006). Recent evidence clearly suggests that miRNAs may participate in key cellular processes such as proliferation, apoptosis, and stem cell division (Brennecke et al., 2003; Hatfield et al., 2005; Xu et al., 2003). Therefore, any aberration in miRNAs may contribute to carcinogenesis by changing the final protein profile of cells.

Comparison between ovarian cancers and its normal counterparts have revealed distinct miRNA expression profiles (Iorio et al., 2007). Most prominently overexpressed miRNAs were miR-200a, miR-141, miR-200c, and miR-200b, whereas miR-199a, miR-140, miR-145, and miR-125b1 were underexpressed. In ovarian cancer, miRNA is deregulated by both genetic and epigenetic mechanisms. In 2006, Zhang et al., showed a high frequency of DNA copy number aberrations in regions containing miRNA in 109 ovarian cancer cases. In addition, they showed that miRNA copy number

changes directly correlated with the levels of miRNA expression (Zhang et al., 2006). Using miRNA microarrays, Dahyia et al. identified several miRNAs aberrantly expressed in human ovarian cancer tissues and cell lines. For example, miR-221 was found to be highly elevated in ovarian cancer, whereas miR-21 and several members of the let-7 family were downregulated (Dahiya et al., 2008).

The expression of miRNAs may be altered by mechanisms such as imprinting or via different epigenetic processes including methylation. Experimental and computer-assisted approaches have led to the identification of hundreds of imprinted small RNA genes, mainly clustered in two chromosomal domains (human 15q11→q13 and 14q32 loci) (Royo et al., 2006). Also, many miRNA genes, including let-7a-3, are located in CpG islands, suggesting possible epigenetic regulation of their expression.

Investigations into the miRNA expression profiles and interaction of miRNA and their targets may provide novel therapeutic options where miRNA are therapeutic targets.

Short interfering RNAs

Small interfering RNA (siRNA) has a well-defined structure: a short (usually 21-nt) double strand of RNA (dsRNA) with 2-nucleotide 3' overhangs on both the ends (Elbashir et al., 2001; Fire et al., 1998). The siRNA are produced in vivo by the action of the enzyme Dicer and can complex with complementary sequences on an mRNA activating a post-transcriptional destruction cascade spearheaded by RNA-induced silencing complex (RISC), hence rendering it non-viable (McManus and Sharp, 2002). Inducing destruction of specific mRNA using siRNA is a powerful tool in analysis of protein function, but until recently its use as a therapeutic modality has been limited. Landen et al. have used efficient siRNA incorporated into the neutral liposome 1, 2-dioleoyl-sn-glycero-3-phosphatidylcholine (DOPC) for efficient *in vivo siRNA* delivery. DOPC-encapsulated siRNA targeting the oncoprotein EphA2 was highly effective in reducing *in vivo* EphA2 expression in Nu/nu mice. Therapy experiments in an orthotopic mouse model of ovarian cancer with EphA2-targeting siRNA-DOPC reduced tumor growth when compared with a nonsilencing siRNA. When EphA2-targeting siRNA-DOPC was combined with paclitaxel, tumor growth was dramatically reduced compared with treatment with paclitaxel and a nonsilencing siRNA (Landen et al., 2005).

Most recently, Merritt et al. examined the effect of IL-8 gene silencing with siRNAs incorporated into neutral liposomes alone and in combination with docetaxel, on *in vivo* tumor growth, angiogenesis (microvessel density), and tumor cell proliferation and found that in this mouse model, siRNA-mediated IL-8 gene silencing decreased tumor growth through antiangiogenic mechanisms (Merritt et al., 2008).

The emerging therapies of ovarian cancer hint at exploiting these mechanisms to target signaling proteins, their ligands, and their pathways. Used alone or in combination with existing chemotherapy, these biologics have potential for more robust tumor response and less toxicity.

2009 UPDATE REFERENCES

Brennecke, J., Hipfner, D.R., Stark, A. et al. (2003). Bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates proapoptotic gene hid in Drosophila. *Cell* 113, 25–35.

Dahiya, N., Sherman-Baust, C.A., Wang, T.L. et al. (2008). MicroRNA expression and identification of putative miRNA targets in ovarian cancer. *PLoS One* 3(6), e2436.

Elbashir, S., Harborth, J. Lendeckel W, et al. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* 411, 494–498.

Fire, A., Xu, S., Montgomery, M. et al. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* 391, 806–811.

Hatfield, S.D., Ahcherbata, H.R., Fischer, K.A. et al. (2005). Stem cell division is regulated by the microRNA pathway. *Nature* 435, 974–978.

Iorio, M.V., Visone, R., Di Leva, G. et al. (2007). MicroRNA signatures in human ovarian cancer. *Cancer Res* 67, 10117–10122.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W. et al. (2001). Identification of novel genes coding for small expressed RNAs. *Science* 294, 853–858.

Landen, C.N., Jr., Chavez-Reyes, A., Bucana, C. et al. (2005). Therapeutic EphA2 gene targeting *in vivo* using neutral liposomal small interfering RNA delivery. *Cancer Res* 56, 6910–6918.

McManus, M.T. and Sharp PA (2002). Gene silencing in mammals by small interfering RNAs. *Nat Rev* 3, 737–746.

Merritt, W.M., Lin, W.G., Spannuth, W.A. et al. (2008). Effect of interleukin-8 gene silencing with liposome-encapsulated small interfering RNA on ovarian cancer cell growth. *J Natl Cancer Inst* 100, 359–372.

Pillai, R.S., Bhattacharyya, S.N. and Filipowcz, W. (2007). Repression of protein synthesis by miRNAs: How many mechanisms. *Trends Cell Biol* 17, 118–126.

Royo, H., Botolin, M.L., Seitz, H. et al. (2006). Small non-soding RNAs and genomic imprinting. *Cytogenet Genome Res* 113, 99–108.

Xu, P., Vernooy, S.Y., Guo, M. et al. (2003). The Drosophila microRNA Mir-14 suppresses cell death and is required for normal fat metabolism. *Curr Biol* 13, 790–795.

Wang, B., Love, T.M., Call, M.E. et al. (2006). Recapitulation of short RNA-directed translational gene silencing *in vitro*. *Mol Cell* 22, 553–560.

Zamore, P.D. and Haley, B. (2005). Ribo-gnome: The big world of small RNAs. *Science* 309, 1519–1524.

Zhang, L., Huang, J., Yang, N. et al. (2006). MicroRNAs exhibit high frequency genomic alterations in human cancer. *Proc Natl Acad Sci U S A* 103, 9136–9141.

REFERENCES

Ahluwalia, A. et al. (2001). DNA methylation and ovarian cancer. I. Analysis of CpG island hypermethylation in human ovarian cancer using differential methylation hybridization. *Gynecol Oncol* 82(2), 261–268.

An, H.J. et al. (2006). Comparative proteomics of ovarian epithelial tumors. *Journal of proteome research* 5(5), 1082–1090.

Bandera, C.A., Ye, B. and Mok, S.C. (1985). New technologies for the identification of markers for early detection of ovarian cancer. *Curr Opin Obstet Gynecol* 15(1), 51–55.

Bell, D.R. et al. (1985). Detection of P-glycoprotein in ovarian cancer: A molecular marker associated with multidrug resistance. *J Clin Oncol* 3(3), 311–315.

Berchuck, A., Kohler, M.F., Marks, J.R. et al. (1994). The p53 tumor suppressor gene frequently is altered in gynecologi cancers. *Am J Obstet Gyn* 12, 64–69.

Chen, T., Triplett, J., Dehner, B., Hurst, B., Colligan, B., Pemberton, J. et al. (2001). Transforming growth factor-beta receptor type I gene is frequently mutated in ovarian carcinomas. *Cancer Res* 61(12), 4679–4682.

Costello, J.F. et al. (2000). Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat Genet* 24(2), 132–138.

Devine, D.A. et al. (2000). Trefoil factor expression in normal and diseased human salivary glands. *Human Pathology* 31(4), 509–515.

Ding, W., Tang, Q., Espina, V., Liotta, L.A., Mauger, D.T. and Mulder, K.M. (2005). A transforming growth factor-beta receptor-interacting protein frequently mutated in human ovarian cancer. *Cancer Res* 65(15), 6526–6533.

Hu, W., Wu, W., Nash, M.A., Freedman, R.S., Kavanagh, J.J. and Verschraegen, C.F. (2000). Anomalies of the TGF-beta postreceptor signaling pathway in ovarian cancer cell lines. *Anticancer Res* 20(2A), 729–733.

Do, T.V., Symowicz, J.C., Berman, B.M., Liotta, L.A., Petricoin, E.F., Stack, M.S. and Fishman, D.A. (2007). Lysophosphatic acid down regulates stress fibers and upregulates pro-matrix metalloproteinase-2 activation. *Mol Cancer Research* 5(2), 121–131.

Dressman, H.K., Berchuck, A., Chan, G., Zhai, J., Bild, A., Sayer, R., Cragun, J., Clarke, J., Whitaker, R.S., LiHua, L. et al. (2007). An integrated genomic-base dapproach to individualized treatment of patients with advanced-stage ovarian cancer. *J Clin Oncol* 25, 517–524.

Easton, D.F., Ford, D. and Bishop, D.T. (1995). Breast and ovarian cancer incidence in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. *Am J Hum Genet* 56(1), 265–271, 15.

Esteller, M. et al. (2000). Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. *J Natl Cancer Inst* 92(7), 564–569.

Ganesan, S., Richardson, A.L., Wang, Z.C., Iglehart, J.D., Miron, A., Feunteun, J. et al. (2005). Abnormalities of the inactive X chromosome are a common feature of BRCA1 mutant and sporadic basal-like breast cancer. *Cold Spring Harb Symp Quant Biol* 70, 93–97.

Havrilesky, L., Alvarez, A.A., Whitaker, R.S. et al. (2001). Loss of expression of the p16 tumor suppressor gene is more frequent in advanced ovarian cancers lacking p53 mutation. *Gynecol Oncol* 83, 491–500.

Jones, P.A. and Baylin, S.B. (2002). The fundamental role of epigenetic events in cancer. *Nat Rev Genet* 3(6), 415–428.

Lu, L., Katsaros, D., Wiley, A., Rigault de la Longrais, I.A., Risch, H.A., Puopolo, M. et al. (2006). The relationship of insulin-like growth factor-II, insulin-like growth factor binding protein-3, and estrogen receptor-alpha expression to disease progression in epithelial ovarian cancer. *Clin Cancer Res* 12(4), 1208–1214.

Lynch, M.A., Nakashima, R., Song, H., DeGroff, V.L., Wang, D., Enomoto, T. et al. (1998). Mutational analysis of the transforming growth factor beta receptor type II gene in human ovarian carcinoma. *Cancer Res* 58(19), 4227–4232.

Macara, I.G. (2004). Parsing the polarity code. *Nat Rev Mol Cell Biol* 5, 220–231.

Marquez, R.T. et al. (2005). Patterns of gene expression in different histotypes of epithelial ovarian cancer correlate with those in normal fallopian tube, endometrium, and colon. *Clinical cancer research: an official journal of the American Association for Cancer Research* 11(17), 6116–6126.

Miki, Y., Swensen, J., Shattuc-Eidens, et al. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 266, 66–71.

Obata, K., Morland, S.J., Watson, R.H. et al. (1998). Frequent PTEN/MMAC mutations in endometrioid but not serous or mucinous epithelial ovarian tumors. *Cancer Res* 58, 2095–2097.

Petricoin, E.F. et al. (2002). Use of proteomic patterns in serum to identify ovarian cancer [see comment]. *Lancet* 359(9306), 572–577.

Rathi, A. et al. (2002). Methylation profiles of sporadic ovarian tumors and nonmalignant ovaries from high-risk women. *Clin Cancer Res* 8(11), 3324–3331.

Ren, J., Xiao, Y.J., Singh, L.S., Zhao, X., Zhao, Z., Feng, L. et al. (2006). Lysophosphatidic acid is constitutively produced by human peritoneal mesothelial cells and enhances adhesion, migration, and invasion of ovarian cancer cells. *Cancer Res* 66, 3006–3014.

Schiffenbauer, Y.S., Abramovitch, R., Meir, G., Nevo, N., Holzinger, M., Itin, A., Keseht, E. and Neeman, M. (1997). Loss of ovarian function promotes angiogenesis in ovarian carcinomas. *Proc. Natl Acad Sci* 94(24), 13202–13208.

Schlosshauer, P.W., Cohen, C.J., Penault-Llorca, F., Miranda, C.R., Bignon, Y.J., Dauplat, J. et al. (2003). Prophylactic oophorectomy: A morphologic and immunohistochemical study. *Cancer* 8(12), 2599–2606.

Sengupta, S., Kim, K.S., Berk, M.P., Oates, R., Escobar, P., Belinson, J., Li, W., Lindmer, D.J., Williams, B. and Xu, Y. (2006). Lysophosphatidic acid downregulates tissue inhibitor of metalloproteinases, which are negatively involved in lysophosphatidic acid-induced cell invasion. *Oncogene* 26, 2894–2901.

Shih, M. and Kurman, R.J. (2004). Ovarian tumorigenesis: A proposed model based on morphological and molecular genetic analysis. *Am J Pathol* 16, 1511–1518.

So, J., Navari, J. and Fishman, D.A. (2004). Lysophosphatidic acid enhances epithelial ovarian carcinoma invasion through increased expression of interleukin 8. 95(2), 314–322.

US Preventive Services Task Force recommendations (2005). Summaries for patients. Genetic risk assessment and BRCA mutation testing for breast and ovarian cancer susceptibility. *Ann Intern Med* 143(5), I47.

Wamunyokoli, F.W. et al. (2006). Expression profiling of mucinous tumors of the ovary identifies genes of clinicopathologic importance. *Clinical cancer research: an official journal of the American Association for Cancer Research* 12(3), 690–700.

Wang, D., Kanuma, T., Mizunuma, H., Takama, F., Ibuki, Y., Wake, N. et al. (2006). Analysis of specific gene mutations in the transforming growth factor-beta signal transduction pathway in human ovarian cancer. *Cancer Res* 60(16), 4507–4512.

Wei, S.H. et al. (2002). Methylation microarray analysis of late-stage ovarian carcinomas distinguishes progression-free survival in patients and identifies candidate epigenetic markers. *Clin Cancer Res* 8(7), 2246–2252.

Wooster, R., Neuhausen, S.L., Mangiou, J. et al. (1994). Localization of a breast cancer susceptability gene, BRCA2, to chromosome 13q12–13. *Science* 265, 2088–2090.

Xing, H., Wang, S., Weng, D., Chen, G., Yang, X., Zhou, J., Xu, G., Lu, Y. and Ma, D. (2007). *Oncol Rep* 17(1), 117–122.

Zhang, H. et al. (2006). Biomarker discovery for ovarian cancer using SELDI-TOF-MS. *Gynecol Oncol* 102(1), 61–66.

CHAPTER

Genomic Evaluation of Pancreatic Neoplasms

Asif Khalid and Kevin McGrath

INTRODUCTION

Pancreatic cancer is a lethal disease, where yearly prevalence equals mortality. It currently is the fourth leading cause of cancer-related death in the United States (Jemal et al., 2006). Unfortunately, despite medical and surgical advances, overall survival has not changed over the last several decades. Owing to the insidious onset of non-specific symptoms, pancreatic cancer generally presents in an advanced state, where the overall 5-year survival rate remains at 5% (Jemal et al., 2006).

Since detection at the earliest stage provides the best chance for cure, recent research efforts have focused on the improved diagnosis, early detection and screening for pancreatic cancer. There is abundant literature supporting the development of pancreatic cancer in parallel with accumulation of genetic alterations. As such, molecular analysis has emerged as an area of intense interest in this regard. Pancreatic intraductal precursor lesions have been re-classified as pancreatic intraepithelial neoplasia (PanIN) based on degree of cellular atypia (Hruban et al., 2001). The prevalence of genetic mutations appears to increase with increasing atypia of these precursor lesions, temporally correlating to PanIN grade progression (Figure 39.1) (Hruban et al., 2000). Understanding the genetic basis of the pancreatic cancer progression model will hopefully provide targets for innovative molecular-based screening and diagnostic testing.

Oncogene activation (e.g., K-ras), tumor-suppressor gene losses (e.g., p53, p16, DPC4, HER-2/neu) and genome-maintenance gene mutations (e.g., BRCA2, microsatellite instability, telomerase) appear to parallel the cellular evolution of pancreatic cancer. Histologically, these changes progress through the PanIN classification, where PanIN-1A is characterized by flat mucinous epithelium, PanIN-1B by papillary change, PanIN-2 by atypical papillary change, and PanIN-3 by carcinoma-*in-situ*. K-ras oncogene point mutations and over-expression of the HER-2/neu gene product appear to be early events in pancreatic carcinogenesis, as these are prevalent in lower grade PanIN lesions (Day et al., 1996; Hruban et al., 1993). p16/INK4a tumor-suppressor gene inactivation is found more frequently in higher grade PanIN lesions, suggesting it is an intermediate event (Wilentz et al., 1998). Inactivation of p53, DPC4/SMAD4 and BRCA2 is prevalent in PanIN-3 lesions and rarely found in low-grade lesions, representing late molecular events (Goggins et al., 2000; Luttges et al., 1999; Wilentz et al., 2000;). Thus, this temporal model of molecular carcinogenesis can serve as a template to allow integration of histologic or cytologic information with that derived from mutational analysis of DNA present in cellular samples, cyst fluid and pancreatic juice with the ultimate goal of increasing early detection rates and diagnostic accuracy.

There is a wealth of information rapidly accumulating regarding the role of molecular analysis in the early detection and diagnosis of pancreatic cancer. This chapter will discuss current molecular capabilities for screening, early detection and improved diagnosis of pancreatic cancer. The focus will be on direct clinical applicability in this rapidly evolving field, amounting

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

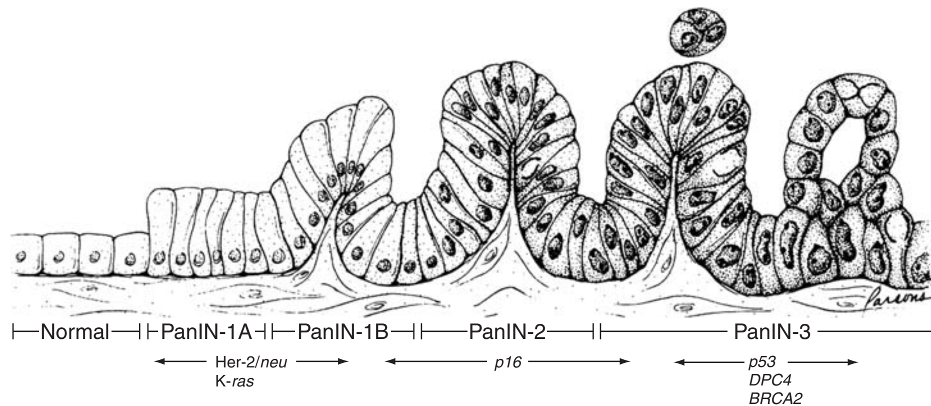


Figure 39.1 The pancreatic carcinogenesis model of PanIN progression. From Hruban, R.H., Goggins, M., Parsons, J., Kern, S.E. (2000). Progression model for pancreatic cancer. *Clin Cancer Res* 6, 2969–2972, with permission.

to translational research in motion. Diagnosis and evaluation of mucinous cystic neoplasms (MCN), which are premalignant pancreatic cysts, will also be discussed, along with new insights into molecular prognostication and tumor-directed gene therapy.

PREDISPOSITION (GENETIC AND NON-GENETIC)

It is estimated that up to 10% of pancreatic cancers may be linked to an inherited component (Lynch et al., 1996). Two distinct groups of patients are considered to be at risk for pancreatic cancer: (1) familial pancreatic cancer, requiring at least two first-degree relatives to be afflicted with the disease in the absence of other familial cancers, and (2) hereditary tumor syndromes carrying an increased incidence of pancreatic cancer. Unfortunately, there has been no definitively identified gene responsible for familial pancreatic cancer, and there is likely phenotypic influence in addition to genetic predisposition. Additionally, the absolute risk is unknown. In the latter group of hereditary tumor syndromes, genetic alterations have been discovered, therefore at-risk patients can be readily identified based on genetic testing. Hereditary syndromes with an increased incidence of pancreatic cancer include familial atypical multiple mole melanoma (FAMMM), Peutz-Jeghers syndrome (PJS), hereditary non-polyposis colorectal carcinoma (HNPCC), familial breast and ovarian cancer (FOBC), cystic fibrosis (CF), ataxia-telangiectasia (AT) and familial adenomatous polyposis (FAP) (Hahn and Bartsch, 2004).

The vast majority of pancreatic cancers are sporadic, without an apparent genetic linkage. Risk factors for sporadic pancreatic cancer include increasing age, tobacco use, chronic pancreatitis and long-standing diabetes mellitus. Additionally, dietary factors such as high intake of meat and fats appear to increase the risk.

SCREENING

Pancreatic Cancer

The ability to reliably screen patients for pancreatic cancer would be a major clinical advance. As with other clinical screening tests, the goal is early detection of disease, ideally in the premalignant stage, as only early detection offers any chance of curative treatment with this deadly disease. Given the relatively low prevalence of pancreatic cancer (33,700 cases estimated for 2006) (Jemal et al., 2006), screening of the general population is not practical. However, secondary screening of high-risk patients, such as those with hereditary pancreatic carcinoma syndromes (up to 10% of cases) may prove feasible. Current screening strategies employ various imaging modalities (computed tomography [CT], endoscopic retrograde cholangiopancreatography [ERCP], endoscopic ultrasound [EUS]) in high-risk patients, such as those with hereditary pancreatic carcinoma syndromes.

An early experience using EUS in high-risk families histologically correlated subtle EUS and ERCP findings suggestive of chronic pancreatitis with dysplasia, or PanIN lesions. Twelve patients from high-risk families were referred for pancreatic resection based on EUS and ERCP findings. Ten had total pancreatectomy and two had distal pancreatectomy; all had widespread dysplasia by surgical pathology, mostly involving small- and medium-sized ducts. CT and serum markers (CEA, CA19-9) had low sensitivity for pancreatic dysplasia (Brentnall et al., 1999; Kimmey et al., 2002). The Seattle group's current screening protocol employs EUS beginning 10 years before the earliest age of cancer development in the affected relative. Normal EUS findings are followed up every 2–3 years; an abnormal EUS is evaluated with ERCP. As the age of the surveyed patient approaches that of the affected relative, the surveillance interval increases to yearly. It is acknowledged that the ideal surveillance interval remains unknown. Given the high prevalence of PanIN lesions in pancreata (Stelow et al., 2006), this aggressive surgical strategy should only be reserved for those high-risk individuals with hereditary pancreatic carcinoma syndromes.

In a larger prospective controlled series, 78 asymptomatic high-risk patients were evaluated with both CT and EUS at baseline and 12 months (Canto et al., 2006). ERCP was performed if EUS detected an abnormality. To date, eight patients have undergone surgery with a confirmed diagnosis of pancreatic neoplasia: six benign Intraductal papillary mucinous neoplasm (IPMN), one malignant IPMN and one focal PanIN. Thus, there was a 10% yield in screening this high-risk population. EUS findings of chronic pancreatitis were more commonly seen in the study group as compared to matched controls. EUS detected more pancreatic lesions than multidetector CT; however, CT scan discovered extra-pancreatic neoplasms not evident on EUS. Hence, the Johns Hopkins group currently recommends both CT and EUS for screening these high-risk patients, as the tests appear to be complementary. Surveillance intervals for lesions detected but not resected are not currently defined, however three patients in this study had IPMN enlargement over the course of 1 year (Canto et al., 2006).

Screening programs for pancreatic cancer, such as the above mentioned, currently depend on imaging modalities to detect radiographic abnormalities. Ideally, molecular analysis of pancreatic juice to detect key mutations involved in pancreatic carcinogenesis could serve as a stand-alone screening test or an adjunct to EUS examination. To this end, the oncogene K-ras has been the most thoroughly studied in pancreatic juice. Studies to date report a widely variable sensitivity and specificity for detection in juice of pancreatic cancer patients, ranging from 30% to 100% (Furuya et al., 1997; Kondo et al., 1994; Kondoh et al., 1998; Matsubayashi et al., 2006; Tada et al., 1993; Trumper et al., 2002; Watanabe et al., 1998; Yamashita et al., 1999). Further troubling has been the detection of K-ras in pancreatic juice of control patients and those with chronic pancreatitis (Furuya et al., 1997; Trumper et al., 2002), where it may reflect the existence of PanIN lesions. Mutational and methylation status analysis of p16 and p53 have also been performed; however, in isolation, these have a low sensitivity for detecting pancreatic cancer (≤42%) (Fukushima et al., 2003; Klump et al., 2003; Matsubayashi et al., 2003; Yamaguchi et al.,1999). It therefore seems unlikely that any one marker will have a future role as a stand-alone screening or diagnostic test.

A recent study evaluated a combination of K-ras, p16 and p53 mutations for predicting the presence of pancreatic cancer (Yan et al., 2005). Pancreatic juice was aspirated at the time of ERCP in 48 patients with pancreatic cancer, 49 patients with chronic pancreatitis and 49 patients with biliary stone disease. p53 mutations were detected in 42% of cancer cases, 4% of patients with chronic pancreatitis, and 0% of controls. K-ras mutations were detected in 54%, 34% and 21% of these cohorts, respectively. Sixty-two percent of cancer patients had p16 promoter methylation levels >12%, compared to 8% of those with chronic pancreatitis and 13% of controls. Although individual mutational analysis lacked sensitivity and specificity, combination analysis increased the discrimination between benign and malignant disease. The authors concluded that the level of discrimination would allow high-risk patients to be stratified from a negligible risk to over a 50% probability of harboring malignancy (Yan et al., 2005). Although case-controlled, the above study was conducted in symptomatic pancreatic cancer patients. The level of discrimination was modeled for screening high-risk patients; however, it remains unproven in asymptomatic patients at risk.

The Johns Hopkins group analyzed pancreatic juice to detect aberrantly methylated DNA for use as a marker of pancreatic neoplasia, as promoter methylation is a common mechanism for gene silencing during carcinogenesis (Matsubayashi et al., 2006). Patients with known pancreatic cancer, IPMN, chronic pancreatitis and controls were evaluated, where juice was obtained via pre-operative ERCP or at surgical resection. Pancreatic juice DNA was analyzed for promoter methylation using methylation-specific PCR assays for 17 genes. There was a significantly higher percentage of genes methylated in the juice of pancreatic cancer patients as compared to the controls and those with chronic pancreatitis. Additionally, a higher percentage of genes were methylated in pancreatic juice obtained at ERCP as compared to surgically obtained juice. Reasoning for this is unclear, but may reflect the use of secretin to stimulate juice production during ERCP (Matsubayashi et al., 2006). Deserving of further study, this analysis may serve as an adjunct to aid in the diagnosis of pancreatic cancer, and may also have screening applications.

Molecular analysis of pancreatic juice for pancreatic cancer screening is currently an intense area of translational research. It is likely that broad-based molecular panels targeting multiple mutations commonly involved in pancreatic carcinogenesis will have the highest yield, and results of ongoing studies are anxiously awaited. It is presumed that these analyses will be performed on pancreatic juice collected at the time of screening EUS examinations.

DIAGNOSIS

Pancreatic Cancer

Short of surgery, pancreatic cancer can prove difficult to diagnose in a subset of patients. Current modalities for diagnosing pancreatic cancer include ERCP with bile duct or pancreatic duct brush cytology, percutaneous ultrasound- (US) or CT-guided fine needle aspiration (FNA) and endoscopic ultrasound-guided FNA (EUS-FNA). Definitive pre-operative diagnosis of pancreatic malignancy is thus dependent on morphologic cellular criteria of cytologic specimens. Cytology-based diagnosis of pancreatic cancer from bile duct brushings has a diagnostic sensitivity of <60% (Farrell et al., 2001; Glasbrenner et al., 1999; Macken et al., 2000; Pugliese et al., 2001; Stewart et al., 2001), where EUS-FNA has a diagnostic sensitivity of 60–95% based on precise targeting of the tumor (Brandwein et al., 2001; Eloubeidi et al., 2003; Shin et al., 2002; Wiersema et al., 1997). In a percentage of cases, sampling error, paucicellular samples, confounding inflammation and desmoplasia and cell drying artifact can result in an indeterminate diagnosis, necessitating repeat invasive procedures which may result in the delay of a definitive diagnosis and ultimately treatment for several months. Molecular

applications applied to indeterminate cytology have potential to increase the diagnostic accuracy for pancreatic cancer.

Microdissection-Based Genotyping

Khalid and colleagues developed an innovative method to evaluate indeterminate cytologic specimens to increase diagnostic yield (Khalid et al., 2004, 2006). Microdissection-based genotyping (MBG) coupled with loss of heterozygosity (LOH) analysis was used to evaluate for a broad panel of tumor-suppressor gene-linked microsatellite markers and k-ras point mutations. This application involves dissecting individual cell aggregates from existing slides and subjecting them to polymerase chain reaction (PCR) to generate DNA for broad panel genotyping of microsatellite allele loss markers. Tumor-suppressor gene loss is based on determination of LOH for informative loci situated within or adjacent to specific genes of interest using fluorescent capillary electrophoresis. Fractional allelic loss (FAL) or fractional mutational rate (FMR), defined as the number of mutations (k-ras-2 point mutation +/− significant allelic imbalance) divided by the total number of informative microsatellite markers plus 1 for k-ras status, is used to measure overall mutation accumulation. As each individual possesses a unique panel of informative polymorphic microsatellite markers, the FAL or FMR allows comparison across cohorts with respect to cumulative acquired LOH.

This technique was first applied to biliary stricture brush cytology obtained at ERCP, which has an overall sensitivity <60% for diagnosing pancreatic cancer. LOH and k-ras codon 12 mutational analysis of PCR amplified DNA from brush cytology specimens discriminated reactive from malignant cells with 100% sensitivity, specificity and accuracy (Khalid et al., 2004).

MBG has since been applied to EUS-FNA cytology samples from pancreatic mass lesions. Benign inflammatory pseudotumors, due to focal chronic pancreatitis or autoimmune pancreatitis, can be impossible to differentiate from malignancy based on imaging alone. All too frequently, EUS-FNA cytology samples are indeterminate due to cellular atypia, and patients are subjected to surgical resection for a definitive diagnosis. Indeed, approximately 10% of pancreaticoduodenectomies performed for presumed malignancy reveal benign disease on pathological evaluation (Abraham et al., 2003; Weber et al., 2003). If surgical pathology proves a benign entity, resection in retrospect is considered unnecessary. LOH analysis of FNA samples shows promise in accurately differentiating benign from malignant pancreatic masses and increasing the diagnostic sensitivity of inconclusive cytologic samples (Khalid et al., 2006). Based on comparison to final surgical pathology, LOH analysis of indeterminate cytology specimens, as compared to positive controls, accurately differentiated benign pseudotumors from pancreatic cancer with high-statistical significance. It additionally improved diagnostic accuracy (100%) for suspicious but inconclusive samples. The FMR for cases ultimately proven malignant was significantly higher than benign cases. Five of six benign cases carried no mutations, where one case of autoimmune pancreatitis harbored a k-ras mutation in the setting of coexisting PanIN lesions (Khalid et al., 2006). Although a currently small experience in this regard, the adjunct genomic analysis will prove to be a powerful tool in the clinical evaluation and management of these difficult cases.

Proteomics

Proteomics is an emerging field incorporating large-scale analysis of proteins in biologic fluids or cells by biochemical techniques. This is an attractive arena for pancreatic cancer research, specifically early detection and diagnosis. Candidate biomarkers have been discovered by analyzing resected pancreatic cancer, pancreatic juice and serum. Analysis of pancreatic juice seems the appropriate strategy when screening high-risk patient groups, whereas serum analysis would be most appropriate for wide-based screening. Although still in its early stage, the use of proteomic profiling for pancreatic cancer biomarker discovery is encouraging.

Two-dimensional electrophoresis for protein separation followed by mass spectrometric identification of proteins has been utilized in analysis of pancreatic cancer and pancreatic juice. However, this method is limited to a relatively low-throughput scale of research (Chen et al., 2005). A more appealing approach for disease biomarker development uses proteomic pattern analysis. This approach uses the pattern of signals within a mass spectrum to identify differentially abundant peaks within normal and disease samples for distinguishing the two groups. Surface-enhanced laser desorption/ionization time-of-flight (SELDI-TOF) mass spectrometry (MS) has been employed in this regard, whereby MS is used to generate proteomic patterns in biological fluids and then pattern recognition algorithms are applied to distinguish cancer patients from normal controls. This technique is suited for high-throughput research and can be applied to body fluids such as pancreatic juice or serum using biochip technology and antibody microarrays to enable multiplexed and rapid protein measurements (Chen et al., 2005).

SELDI-TOF has also been applied to pancreatic tissue to identify protein peaks to differentiate benign from malignant states. In one study, a training model was developed that could accurately distinguish pancreatic cancer from benign pancreatic tissue based on the protein peak pattern (Scarlett et al., 2006). A similar study found a large number of proteins differentially expressed in chronic pancreatitis and pancreatic cancer as compared to normal pancreatic tissue (Crnogorac-Jurcevic et al., 2005). It is unclear whether this technology will enhance preoperative diagnostic accuracy in this regard, as these methodologies have yet to be applied to EUS-FNA cytology samples. More importantly, these protein profiles may potentially be applied to pancreatic juice or serum to screen high-risk patients.

As such, the first comprehensive study of the pancreatic juice proteome revealed numerous proteins associated with pancreatic cancer (Chen et al., 2006). This study identified candidate biomarkers differentially expressed in cancer patients as compared to normal controls. This group has since evaluated pancreatic juice proteins in the setting of pancreatitis, in hopes of eliminating false-positive results for cancer given the potential for overlapping biomarkers (Chen et al., 2007). Similar preliminary work has also been applied to serum via plasma protein

profiling. Honda and colleagues compared plasma proteomes between pancreatic cancer patients and matched controls using SELDI-TOF MS. A learning algorithm was applied to discriminate proteomic patterns, identifying a set of four mass peaks that differentiated sera from pancreatic cancer patients with high accuracy rates (Honda et al., 2005). Although exciting, this work will need to be confirmed in a large-scale multi-center trial to truly determine clinical utility. There is also a potential role for proteomics to monitor for disease recurrence in patients who have undergone pancreatic cancer resection. An innovative study using two-dimensional electrophoresis identified a group of plasma proteins that consistently change following resection. Furthermore, this group identified proteins that correlated with recurrence of disease (Lin et al., 2006). Although a small study, the clinical applications of this type of technology could revolutionize post-operative monitoring and surveillance.

Pancreatic Cystic Neoplasms

Incidental pancreatic cystic lesions are being detected with increasing frequency due to the widespread use of high-quality abdominal imaging. Contrary to previous belief, the majority of cysts detected today are MCN (Fernandez-del Castillo et al., 2003). MCN are considered premalignant and encompass IPMN and MCN. Surgical resection of mucinous pancreatic cysts is generally recommended given the concern for malignant degeneration. However, the natural history regarding the frequency and timing of malignant change of MCN remains unknown.

Evaluation of pancreatic cysts involve cross-sectional imaging (CT or MRI) and EUS-FNA with cyst aspirate analysis. Unfortunately, in the absence of an associated solid mass, imaging modalities cannot reliably differentiate benign from malignant cysts (Ahmad et al., 2001, 2003; Brugge et al., 2004). Given the paucicellular nature of the cyst aspirate, the sensitivity of cytologic analysis has proven suboptimal. Currently, an elevated cyst fluid CEA level is considered the most reliable indicator of a mucinous cyst; however it does not distinguish premalignant from malignant state (Brugge et al., 2004). Given the unknown natural history of MCN and associated morbidity of pancreatic resection, better tools to assess presence of malignancy are desirable. This is especially important in marginal surgical candidates with small incidental cystic neoplasms.

As the cellular content of pancreatic cyst aspirates is frequently suboptimal, molecular analysis of the cyst fluid itself is attractive in hopes of detecting mutations that correlate with malignant degeneration. It is hypothesized that cyst epithelial cells undergoing malignant change would have a higher turnover rate, thereby releasing more DNA into the fluid that bathes the cells. PCR amplification of DNA from these whole or lysed cells shed into the fluid from the cyst lining may therefore be predictive of the cyst pathology, where a high level of accumulated mutational damage would reflect an underlying malignancy, and similar alterations would not be seen in benign cysts.

To investigate this hypothesis, a single center study was conducted, applying LOH and K-ras mutational analysis to pancreatic cyst aspirates targeting markers for pancreatic carcinogenesis (Khalid et al., 2005). Based on comparison to surgical pathology of the resected cyst, LOH analysis of the cyst fluid aspirate accurately predicted the existence of malignancy. Thirty-six cysts with confirmed histology were analyzed: 11 malignant, 15 premalignant and 10 benign cysts. The malignant cysts could be differentiated from premalignant cysts based on the number of mutations and the temporal sequence in which the mutations were acquired ($p < 0.001$). Early K-ras mutation followed by allelic loss was the most predictive of a malignant cyst (sensitivity 91%, specificity 93%) (Khalid et al., 2005). Thus, it appears that pancreatic cyst fluid contains adequate DNA to allow mutational analysis that can serve as an ancillary tool to the conventional workup of pancreatic cysts. A multi-center trial to further evaluate this technology has been completed, with the interim analysis showing that the quantity of DNA and temporal sequence of mutations continue to predict cyst pathology (Khalid et al., 2006a, b).

Several studies have also investigated whether telomerase activity can serve as a marker for malignancy in cystic neoplasms. In one small study evaluating IPMN, pancreatic juice aspirated at the time of ERCP was analyzed. Cytology alone diagnosed malignancy in 4/13 patients, but when combined with analysis for telomerase activity, the yield increased to 11 of 13 patients being correctly diagnosed with malignant IPMN. As telomerase activity was not detected in benign tumors, the authors suggest telomerase may be a useful adjunct to standard cytologic analysis (Inoue et al., 2001).

In another study analyzing a spectrum of cystic lesions, Yeh and colleagues found expression of telomerase activity in malignant and borderline malignant cysts, but not in benign cysts (mucinous or serous) or pseudocysts. The sensitivity of telomerase for predicting at least borderline malignant change was 67% with a specificity of 100% (Yeh et al., 1999). Thus, it appears that telomerase activity takes part in malignant transformation and may be a useful marker to distinguish malignant from benign cystic neoplasms.

The Johns Hopkins group has evaluated gene expression profiles to help differentiate invasive from non-invasive IPMNs. Employing RT-PCR and oligonucleotide microarrays, they were able to identify genes that were overexpressed in IPMNs associated with invasive carcinoma. Immunohistochemical validation revealed that claudin 4, CXCR4, S100A4 and mesothelin were associated with the invasive phenotype. This analysis was performed on resected specimens; thus further work is required to see if these results can be reproduced pre-operatively on aspirated cyst fluid, as they could be used to guide management regarding surgery or observation (Sato et al., 2004).

PROGNOSIS

Pancreatic Cancer

Pre-operative prognostication for malignant disease is currently based on the cancer stage and grade. Prognostic markers are desired to stratify treatment protocols, to enable individualized

therapy, and to develop new treatment strategies. For pancreatic adenocarcinoma, in which the majority of cases present with locally advanced and unresectable disease, prognostication unfortunately has not been of significant clinical importance to date given the high-mortality rate with only few long-term survivors (Jemal et al., 2006). However, gene expression profiling of the tumor may predict clinical course and outcome in select patients with potentially resectable disease.

In a recent study from MD Anderson Cancer Center, single nucleotide polymorphisms (SNPs) in DNA repair genes were evaluated to see if they affected clinical prognosis. Previous studies have shown that individual variation in DNA repair capacity can affect response to therapy and overall survival. All patients underwent neoadjuvant therapy with gemcitabine and radiotherapy after gene expression was determined from whole blood. Genotypes *RecQ1* A159C, *RAD54L* C157T, *XRCC1* R194W and *ATM* T77C had a significant effect on overall survival. The overall mean survival of the group was 20.2 months. However, patients with none of the adverse genotypes had a mean survival of 62.1 months. Those with one, two or three or more at risk alleles had median survival times of 27.5, 14.4 and 9.9 months, respectively (Li et al., 2006). Thus, for these earlier stage tumors, it appears that polymorphic variants of DNA repair genes affect the clinical course of disease, which hopefully will translate to novel targeted therapy in the future.

Other studies have been performed *ex vivo*, where resected pancreatic cancer has been profiled and compared to the outcome of the patient. Cytokeratins (CK) 7 and 20 were found to be overexpressed in pancreatic cancer, where CK 20 expression defined a subset of tumors with a worse prognosis (Matros et al., 2006). Peroxisome proliferator-activated receptor gamma (PPARgamma) is a ligand-activated transcription factor shown to be overexpressed and associated with a higher tumor stage and grade, thus correlating to a worse prognosis (Kristiansen et al., 2006). Secreted protein acidic and rich in cysteine (SPARC), a protein involved in cell migration and cell matrix interactions, is frequently silenced in pancreatic cancer but expressed in stromal fibroblasts. Peritumoral SPARC expression has also been associated with a poor clinical prognosis (Infante et al., 2007).

Thus, many genes have been identified that are differentially expressed in pancreatic cancer, some of which are predictive of clinical outcome. If this could be determined routinely in a pre-operative fashion, a patient could be counseled as to the true benefit of major pancreatic surgery. A study evaluating EUS-FNA cytology samples revealed that RT-PCR detected increased expression of lipocalin 2 (LCN 2) and tissue-type plasminogen activator (PLAT) in pancreatic cancer (Laurell et al., 2006). Not only could this technology increase diagnostic accuracy, but more importantly, could provide pre-operative prognostic information. The field of gene expression profiling is expanding quickly in the pancreatic cancer arena and will continue to evolve rapidly. Not only is it anticipated to provide reliable prognostic information, it may also give insight into the pancreatic carcinogenesis pathway and provide novel targets for therapy.

Pancreatic Endocrine Tumors

In less aggressive pancreatic tumors, the ability to pre-operatively prognosticate may be very significant, especially for marginal surgical candidates given the morbidity associated with pancreatic surgery. Pancreatic endocrine tumors (PET) have been reported to occur with an incidence of 1 per 100,000 persons per year (Barakat et al., 2004); however, it is expected that incidental PETs will be discovered with increasing frequency given widespread use of cross-sectional imaging, much like the phenomenon occurring with incidental pancreatic cystic lesions (Warner, 2005). The biological behavior of PETs can vary widely from clinically indolent to highly aggressive. Elevated Ki-67 proliferative index >2%, mitotic rate ≥2, size ≥4 cm, nuclear atypia, capsular penetration with local invasion, and/or metastatic disease define a malignant PET. The Ki-67 index and mitotic rate also correlate with survival (La Rosa et al., 1996; Panzuto et al., 2005). Unfortunately, these indices and histologic assessment come from surgical resection specimens.

Similar to pancreatic adenocarcinoma, the developmental progression and malignant transformation of PET also appear to correlate with accumulation of genetic alterations. Recent studies indicate that chromosomal losses occur more commonly than gains or amplifications, and the LOH profile correlates with tumor size, extent and malignant phenotype (Barghorn et al., 2001a, b; Guo et al., 2002a, b; Hessman et al., 1999, 2001; Speel et al., 1999, 2001; Stumpf et al., 2000; Zhao et al., 2001). Furthermore, molecular markers such as telomerase activity, Her-2/neu overexpression, hMLH1 methylation and microsatelliteinstability may predict PET behavior independent of the tumor's functional status or histopathologic features (Furlan et al., 2004; Goebel et al., 2002; House et al., 2003; Lam et al., 2000). These data suggest that molecular analysis of PET may provide relevant information of clinical and prognostic utility. To date, however, these analyses have only been performed on surgically resected specimens. The molecular investigation of pre-operative EUS-FNA samples from PET are very appealing for predicting the clinical course specifically in an era where more incidental lesions are being detected.

To that effect, the Pittsburgh group performed broad panel microsatellite loss analysis on pre-operative cytologic samples obtained via EUS-FNA. Twenty-five patients were studied; 13 with "benign" or indolent and 12 with malignant disease, respectively, based on pathologic assessment and clinical course. As previously reported, tumor size greater than 3 cm and high Ki-67 index and mitotic rate correlated well with disease progression. However, FAL appeared to be the strongest factor associated with disease progression. Four of 13 "benign" PETs contained a single microsatellite loss; the other nine were without mutation (mean FAL 0.028). All 12 malignant PETs harbored multiple allelic losses (mean FAL 0.37). The LOH profile not only differentiated benign from malignant PET (FAL 0.2, sensitivity 83%, specificity 100%), but also correlated best with disease progression and survival. Thus, for the first time, this molecular information can be used for clinical decision-making and prognostication *pre-operatively*, especially in those patients who are marginal surgical candidates (Nodit et al., 2006).

2009 UPDATE

Pancreatic Cancer Development

The role of microRNA (miRNA) in various aspects of cancer pathogenesis, including cell growth and apoptosis, has become an area of intense interest. A recent study analyzing the differential expression of 95 miRNA by real time RT-PCR found that the expression of most miRNA was substantially changed in pancreatic cancer specimens and pancreatic cancer cell lines compared to normal pancreatic tissue. Additionally each pancreatic cancer specimen or cell line had a substantially different profiling pattern. A detailed analysis on 10 pancreatic cancer cell lines and 17 pairs of pancreatic cancer and normal tissues revealed eight miRNAs (miR-196a, miR-190, miR-186, miR-221, miR-222, miR-200b, miR-15b, and miR-95) that were significantly up regulated (70-100%) in most pancreatic cancer specimens and cell lines (Zhang et al., 2008).

Pancreatic Cancer Diagnosis

In a study from Johns Hopkins Hospital, ERCP brush DNA concentrations of methylated Cyclin D2, NPTX2, and TFPI2 promoter DNA were measured in 130 individuals with biliary tract strictures (64 malignant; 41 pancreatic cancer and 10 bile duct cancer, 66 benign) by real time quantitative PCR. A significantly higher number of pancreatic cancer (73.2%) and bile duct cancer (80%) samples had at least 1 gene positive for methylation compared with 13.6% of patients with a benign stricture (Parsi et al., 2008). This study adds to the existing literature on the incremental yield of molecular techniques on routine cytology.

There have been two studies that provide promising data that, although in its infancy, may serve as a platform for the development of a "screening test" for pancreatic cancer. In the first study, serum-based LOH analysis of 12 microsatellites on chromosomal arms 2p, 5q, 9p, 13q, 17p and 17q was performed in 35 pancreatic cancer patients, 22 patients with chronic pancreatitis and 20 healthy individuals. Thirty-two patients manifested LOH with the majority (30) with two or more regions of LOH. Single LOH was observed in 7/22 patients with chronic pancreatitis and none in the healthy individuals (Wachowiak et al., 2007). In the second study, differential methylation profiles of plasma DNA utilizing multiplexed array-mediated analysis in 30 pancreatic cancer patients and an equal number of matched controls could identify cancer, albeit with a sensitivity of 76% and specificity of only 59% (Melnikov et al., 2008).

Pancreatic Cancer Prognosis

In a study correlating the presence of k-ras mutation and LOH affecting 5q, 6q, 9p, 10q, 12q, 17p and 18q in resected pancreatic cancer with postoperative survival, once again the importance of genetic events on tumor biology is illustrated. In this study with a median f/u of 18 months, tumor stage, nodal and margin status were not predictive of survival. However, the presence of LOH and k-ras and the combination of both were significantly associated with shorter median survival (Franko et al., 2008).

Pancreatic Endocrine Tumors

The clinical course of endocrine pancreatic tumors (PET) is often difficult to predict as previously alluded to. As further evidence for the key role specific DNA alterations play in this context, a study of 47 insulinomas and 24 non-insulinoma PET revealed that chromosomal alterations affecting 3p, 6q and 12q were most significantly associated with a poor outcome in insulinomas, and 7q gain in non-insulinoma PET. CK19 expression was the most optimal predictor of tumor-specific death (Jonkers et al., 2007).

2009 UPDATE REFERENCES

Franko, J., Krasinskas, A.M., Nikiforova, M.N., Zarnescu, N.O., Lee, K.K., Hughes, S.J., Bartlett, D.L., Zeh, H.J., 3rd and Moser, A.J. (2008). Loss of heterozygosity predicts poor survival after resection of pancreatic adenocarcinoma. *J Gastrointest Surg* 12, 1664-1672.

Jonkers, Y.M., Claessen, S.M., Perren, A., Schmitt, A.M., Hofland, L.J., de Herder, W., de Krijger, R.R., Verhofstad, A.A., Hermus, A.R., Kummer. J.A. et al. (2007). DNA copy number status is a powerful predictor of poor survival in endocrine pancreatic tumor patients. *Endocr Relat Cancer* 14, 769-779.

Melnikov, A.A., Scholtens, D., Talamonti, M.S., Bentrem, D.J. and Levenson, V.V. Methylation profile of circulating plasma DNA in patients with pancreatic cancer. *J Surg Oncol* 99(2), 119-122.

Parsi, M.A., Li, A., Li, C.P. and Goggins, M. (2008). DNA methylation alterations in endoscopic retrograde cholangiopancreatography brush samples of patients with suspected pancreaticobiliary disease. *Clin Gastroenterol Hepatol* 6, 1270-1278.

Wachowiak, R., Kaifi, J., Schwarzenbach, H., Yekebas, E., Merkert, P., Schurr, P., Hansen, B., Reichelt, U., Strate, T., Pantel, K. et al. (2007). Microsatellite analysis in serum DNA as a diagnostic tool for distinction of patients with unknown pancreatic masses. *Diagn Mol Pathol* 16, 174-178.

Zhang, Y., Li, M., Wang, H., Fisher, W.E., Lin, P.H., Yao, Q. and Chen, C. Profiling of 95 microRNAs in pancreatic cancer cell lines and surgical specimens by real-time PCR analysis. *World J Surg* 33(4), 698-709.

REFERENCES

Abraham, S.C., Wilentz, R.E., Yeo, C.J., Sohn, T.A., Cameron, J.L., Boitnott, J.K. and Hruban, R.H. (2003). Pancreaticoduodenectomy (Whipple resections) in patients without malignancy: Are they all 'chronic pancreatitis'? *Am J Surg Pathol* 27, 110–120.

Ahmad, N.A., Kochman, M.L., Lewis, J.D. and Ginsberg, G.G. (2001). Can EUS alone differentiate between malignant and benign cystic lesions of the pancreas? *Am J Gastroenterol* 96, 3295–3300.

Ahmad, N.A., Kochman, M.L., Brensinger, C., Brugge, W.R., Faigel, D.O., Gress, F.G., Kimmey, M.B., Nickl, N.J., Savides, T.J., Wallace, M.B. et al. (2003). Interobserver agreement among endosonographers for the diagnosis of neoplastic versus non-neoplastic pancreatic cystic lesions. *Gastrointest. Endosc* 58, 59–64.

Barakat, M.T., Meeran, K. and Bloom, S.R. (2004). Neuroendocrine tumors. *Endocr Relat Cancer* 11, 1–18.

Barghorn, A., Komminoth, P., Bachmann, D., Rutimann, K., Saremaslani, P., Muletta-Feurer, S., Perren, A., Roth, J., Heitz, P.U. and Speel, E.J. (2001a). Deletion at 3p25. *J Pathol* 194, 451–458.

Barghorn, A., Speel, E.J., Farspour, B., Saremaslani, P., Schmid, S., Perren, A., Roth, J., Heitz, P.U. and Komminoth, P. (2001b). Putative tumor suppressor loci at 6q22 and 6q23-q24 are involved in the malignant progression of sporadic endocrine pancreatic tumors. *Am J Pathol* 158, 1903–1911.

Brandwein, S.L., Farrell, J.J., Centeno, B.A. and Brugge, W.R. (2001). Detection and tumor staging of malignancy in cystic, intraductal and solid tumors of the pancreas by EUS. *Gastrointest Endosc* 53, 722–727.

Brentnall, T.A., Bronner, M.P., Byrd, D.R., Haggitt, R.C. and Kimmey, M.B. (1999). Early diagnosis and treatment of pancreatic dysplasia in patients with a family history of pancreatic cancer. *Ann Intern Med* 131, 247–255.

Brugge, W.R., Lewandrowski, K., Lee-Lewandrowski, E., Centeno, B.A., Szydlo, T., Regan, S., del Castillo, C.F. and Warshaw, A.L. (2004). Diagnosis of pancreatic cystic neoplasms: A report of the cooperative pancreatic cyst study. *Gastroenterology* 126, 1330–1336.

Canto, M.I., Goggins, M., Hruban, R.H., Petersen, G.M., Giardiello, F.M., Yeo, C., Fishman, E.K., Brune, K., Axilbund, J., Griffin, C. et al. (2006). Screening for early pancreatic neoplasia in high-risk individuals: A prospective controlled study. *Clin Gastroenterol Hepatol* 4, 766–781.

Chen, R., Pan, S., Cooke, K., Moyes, K.W., Bronner, M.P., Goodlett, D.R., Aebersold, R. and Brentnall, T.A. (2007). Comparison of pancreas juice proteins from cancer versus pancreatitis using quantitative proteomic analysis. *Pancreas* 34, 70–79.

Chen, R., Pan, S., Yi, E.C., Donohoe, S., Bronner, M.P., Potter, J.D., Goodlett, D.R., Aebersold, R. and Brentnall, T.A. (2006). Quantitative proteomic profiling of pancreatic cancer juice. *Proteomics* 6, 3871–3879.

Chen, R., Sheng, P., Brentnall, T.A. and Aebersold, R. (2005). Proteomic profiling of pancreatic cancer for biomarker discovery. *Mol Cell Proteomics* 4, 523–533.

Crnogorac-Jurcevic, T., Gangeswaran, R., Bhakta, V., Capurso, G., Lattimore, S., Akada, M., Sunamura, M., Prime, W., Campbell, F., Brentnall, T.A. et al. (2005). Proteomic analysis of chronic pancreatitis and pancreatic adenocarcinoma. *Gastroenterology* 129, 1454–1463.

Day, J.D., Digiuseppe, J.A., Yeo, C., Lai-Goldman, M., Anderson, S.M., Goodman, S.N., Kern, S.E. and Hruban, R.H. (1996). Immunohistochemical evaluation of HER-2/neu expression in pancreatic adenocarcinoma and pancreatic intraepithelial neoplasms. *Hum Pathol* 27, 119–124.

Eloubeidi, M.A., Jhala, D., Chieng, D.C., Chen, V.K., Eltoum, I., Vickers, S., Wilcox, C.M. and Jhala, N. (2003). Yield of endoscopic ultrasound-guided fine-needle aspiration biopsy in patients with suspected pancreatic carcinoma. *Cancer* 99, 285–292.

Farrell, R.J., Jain, A.K., Brandwein, S.L., Wang, H., Chuttani, R. and Pleskow, D.K. (2001). The combination of stricture dilation, endoscopic needle aspiration, and biliary brushings significantly improves diagnostic yield from malignant bile duct strictures. *Gastrointest Endosc* 54, 587–594.

Fernandez-del Castillo, C., Targarona, J., Thayer, S.P., Rattner, D.W., Brugge, W.R. and Warshaw, A.L. (2003). Incidental pancreatic cysts: Clinicopathologic characteristics and comparison with symptomatic patients. *Arch Surg* 138, 423–427.

Fukushima, N., Walter, K.M., Uek, T., Sato, N., Matsubayashi, H., Cameron, J.L., Hruban, R.H., Canto, M., Yeo, C.J. and Goggins, M. (2003). Diagnosing pancreatic cancer using methylation specific PCR analysis of pancreatic juice. *Cancer Biol Ther* 2, 78–83.

Furlan, D., Cerutti, R., Uccella, S., La Rosa, S., Rigoli, E., Genasetti, A. and Capella, C. (2004). Different molecular profiles characterize well-differentiated endocrine tumors and poorly differentiated endocrine carcinomas of the gastroenteropancreatic tract. *Clin Cancer Res* 10, 947–957.

Furuya, N., Kawa, S., Akamatsu, T. and Furihata, K. (1997). Long-term follow-up of patients with chronic pancreatitis and K-*ras* gene mutation detected detected in pancreatic juice. *Gastroenterology* 113, 593–598.

Glasbrenner, B., Ardan, M., Boeck, W., Preclik, G., Moller, P. and Adler, G. (1999). Prospective evaluation of brush cytology of biliary strictures during endoscopic retrograde cholangiopancreatography. *Endoscopy* 31, 712–717.

Goebel, S.U., Iwamoto, M., Raffeld, M., Gibril, F., Hou, W., Serrano, J. and Jensen, R.T. (2002). Her-2/neu expression and gene amplification in gastrinomas: Correlations with tumor biology, growth, and aggressiveness. *Cancer Res* 62, 3702–3710.

Goggins, M., Hruban, R.H. and Kern, S.E. (2000). BRCA2 is inactivated late in the development of pancreatic intraepithelial neoplasia: Evidence and implications. *Am J Pathol* 156, 1767–1771.

Guo, S.S., Arora, C., Shimoide, A.T. and Sawicki, M.P. (2002a). Frequent deletion of chromosome 3 in malignant sporadic pancreatic endocrine tumors. *Mol Cell Endocrinol* 190, 109–114.

Guo, S.S., Wu, A.Y. and Sawicki, M.P. (2002b). Deletion of chromosome 1, but not mutation of MEN-1, predicts prognosis in sporadic pancreatic endocrine tumors. *World J Surg* 26, 843–847.

Hahn, S.A. and Bartsch, D.K. (2004). Genetics of hereditary pancreatic carcinoma. *Gastroenterol Clin North Am* 33(4), 919–934.

Hessman, O., Lindberg, D., Einarsson, A., Lillhager, P., Carling, T., Grimelius, L., Eriksson, B., Akerstrom, G., Westin, G. and Skogseid, B. (1999). Genetic alterations on 3p, 11q13, and 18q in nonfamilial and MEN 1-associated pancreatic endocrine tumors. *Genes Chromosomes Cancer* 26, 258–264.

Hessman, O., Skogseid, B., Westin, G. and Akerstrom, G. (2001). Multiple allelic deletions and intratumoral genetic heterogeneity in men1 pancreatic tumors. *J Clin Endocrinol Metab* 86, 1355–1361.

Honda, K., Hayashida, Y., Umaki, T., Okusaka, T., Kosuge, T., Kikuchi, S., Endo, M., Tsuchida, A., Aoki, T., Itoi, T. et al. (2005). Possible detection of pancreatic cancer by plasma protein profiling. *Cancer Res* 65, 10613–10622.

House, M.G., Herman, J.G., Guo, M.Z., Hooker, C.M., Schulick, R.D., Cameron, J.L., Hruban, R.H., Maitra, A. and Yeo, C.J. (2003). Prognostic value of hMLH1 methylation and microsatellite instability in pancreatic endocrine neoplasms. *Surgery* 134, 902–908.

Hruban, R.H., van Mansfeld, A.D., Offerhaus, G.J., van Weering, D.H., Allison, D.C., Goodman, S.N., Kensler, T.W., Bose, K.K., Cameron, J.L. and Bos, J.L. (1993). K-ras oncogene activation in adenocarcinoma

of the human pancreas: A study of 82 carcinomas using a combination of mutant-enriched polymerase chain reaction analysis and allele-specific oligonucleotide hybridization. *Am J Pathol* 143, 545–554.

Hruban, R.H., Wilentz, R.E. and Kern, S.E. (2000). Genetic progression in the pancreatic ducts. *Am J Pathol* 156, 1821–1825.

Hruban, R.H., Adsay, N.V., Albores-Saavedra, J., Comptom, C., Garrett, E.S., Goodman, S.N., Kern, S.E., Klimstra, D.S., Kloppel, G., Longnecker, D.S. et al. (2001). Pancreatic intraepithelial neoplasia: A new nomenclature and classification system for pancreatic duct lesions. *Am J Surg Pathol* 25, 579–586.

Infante, J.R., Matsubayashi, H., Sato, N., Tonascia, J., Klein, A.P., Riall, T.A., Yeo, C., Iacobuzio-Donahue, C. and Goggins, M. (2007). Peritumoral fibroblast SPARC expression and patient outcome with resectable pancreatic adenocarcinoma. *J Clin Oncol* 25, 319–325.

Inoue, H., Tsuchida, A., Kawasaki, Y., Fujimoto, Y., Yamasaki, S. and Kajiyama, G. (2001). Preoperative diagnosis of intraductal papillary-mucinous tumors of the pancreas with attention to telomerase activity. *Cancer* 91, 35–41.

Jemal, A., Siegel, R., Ward, E., Murray, T., Xu, J., Smigal, C. and Thun, M. (2006). Cancer Statistics 2006. *CA Cancer J Clin* 56, 106–130.

Khalid, A., McGrath, K.M., Zahid, M., Wilson, M., Brody, D., Swalsky, P., Moser, A.J., Lee, K.K., Slivka, A., Whitcomb, D.C. et al. (2005). The role of pancreatic cyst fluid molecular analysis in predicting cyst pathology. *Clin Gastroenterol Hepatol* 3, 967–973.

Khalid, A., Pal, R., Sasatomi, E., Swalsky, P., Slivka, A., Whitcomb, D. and Finkelstein, S. (2004). Use of microsatellite marker loss of heterozygosity in accurate diagnosis of pancreaticobiliary malignancy from brush cytology samples. *Gut* 53, 1860–1865.

Khalid, A., Nodit, L., McGrath, K., Jani, N., Carty, S. and Finkelstein, S. (2006a). Endoscopic ultrasound-guided fine needle aspirate microsatellite loss analysis in predicting pancreatic endocrine tumor behavior and clinical course. *Gastrointest Endosc* 63, AB252.

Khalid, A., Finkelstein, S., Kaushik, N., Leblanc, J., Ahmad, N., Brugge, W., Edmundowicz, S., Hawes, R. and McGrath, K. (2006b). Pancreatic cyst DNA analysis (PANDA) study: An interim report. *Gastrointest Endosc* 63, AB93.

Khalid, A., Nodit, L., Zahid, M., Bauer, K., Brody, D., Finkelstein, S., McGrath, K. (2006). Endoscopic ultrasound fine needle aspirate DNA analysis to differentiate malignant and benign pancreatic masses. *Am J Gastroenterol* 101, 2493–2500

Kimmey, M.B., Bronner, M.P., Byrd, D.R. and Brentnall, T.A. (2002). Screening and surveillance for hereditary pancreatic cancer. *Gastrointest Endosc* 56, S82–S86.

Klump, B., Hsieh, C.J., Nehls, O., Dette, S., Holzmann, K., Kiesslich, R., Jung, M., Sinn, U., Ortner, M., Porschen, R. et al. (2003). Methylation status of p14ARF and p16INK4a as detected in pancreatic secretions. *Br J Cancer* 88, 217–222.

Kristiansen, G., Jacob, J., Buckendahl, A.C., Grutzmann, R., Alldinger, I., Sipos, B., Kloppel, G., Bahra, M., Langrehr, J.M., Neuhaus, P. et al. (2006). Peroxisome proliferator-activated receptor gamma is highly expressed in pancreatic cancer and is associated with shorter overall survival times. *Clin Cancer Res* 12, 6444–6451.

Kondo, H., Sugano, K., Fukayama, N., Kyogoku, A., Nose, H., Shimada, K., Ohkura, H., Ohtsu, A., Yoshida, S. and Shimosato, Y. (1994). Detection of point mutations in the K-*ras* oncogene at codon 12 in pure pancreatic juice for diagnosis of pancreatic carcinoma. *Cancer* 73, 1589–1594.

Kondoh, S., Kaino, M., Okita, S., Ryozawa, S., Akiyama, T. and Okita, K. (1998). Detection of Ki-ras and p53 gene mutations in tissue and pancreatic juice from pancreatic adenocarcinomas. *J Gastroenterol* 33, 390–396.

Lam, K.Y., Lo, C.Y., Fan, S.T. and Luk, J.M. (2000). Telomerase activity in pancreatic endocrine tumors: A potential marker for malignancy. *Mol Pathol* 53, 133–136.

La Rosa, S., Sessa, F., Capella, C., Riva, C., Leone, B.E., Klersy, C., Rindi, G. and Solcia, E. (1996). Prognostic criteria in nonfunctioning pancreatic endocrine tumors. *Virchows Arch* 429, 323–333.

Laurell, H., Bouisson, M., Berthelemy, P., Rochaix, P., Dejean, S., Besse, P., Susini, C., Pradayrol, L., Vaysse, N. and Buscail, L. (2006). Identification of biomarkers of human pancreatic adenocarcinomas by expression profiling and validation with gene expression analysis in endoscopic ultrasound-guided fine needle aspiration samples. *World J Gastroenterol* 12, 3344–3351.

Li, D., Frazier, M., Evans, D.B., Hess, K.R., Crane, C.H., Jiao, L. and Abbruzzese, J.L. (2006). Single nucleotide polypmorphisms of RecQ1, RAD 54L and ATM genes are associated with reduced survival of pancreatic cancer. *J Clin Oncol* 24, 1720–1728.

Lin, Y., Goedegebuure, P.D., Tan, M.C., Gross, J., Malone, J.P., Feng, S., Larson, J., Phommaly, C., Trinkaus, K., Townsend, R.R. and Linehan, D.C. (2006). Proteins associated with disease and clinical course in pancreas cancer: a proteomic analysis of plasma in surgical patients. *J Proteome Res* 5, 2169–2176.

Luttges, J., Schlehe, B., Menke, M.A., Vogel, I., Henne-Bruns, D. and Kloppel, G. (1999). The K-ras mutation pattern in pancreatic ductal adenocarcinoma usually is identical to that in associated normal, hyperplastic, and metaplastic ductal epithelium. *Cancer* 85, 1703–1710.

Lynch, H.T., Smyrk, T., Kern, S.E., Hruban, R.H., Lightdale, C.J., Lemon, S.J., Lynch, J.F., Fusaro, L.R., Fusaro, R.M. and Ghadirian, P. (1996). Familial pancreatic cancer: A review. *Semin Oncol* 23, 251–275.

Macken, E., Drijkoningen, M., Van Aken, E. and Van Steenbergen, W. (2000). Brush cytology of ductal strictures during ERCP. *Acta Gastroenterol Belg* 63, 254–259.

Matros, E., Bailey, G., Clancy, T., Zinner, M., Ashley, S., Whang, E. and Redston, M. (2006). Cytokeratin 20 expression identifies a subtype of pancreatic adenocarcinoma with decreased overall survival. *Cancer* 106, 693–702.

Matsubayashi, H., Sato, N., Fukushima, N., Yeo, C.J., Walter, K.M., Brune, K., Sahin, F., Hruban, R.H. and Goggins, M. (2003). Methylation of cyclin D2 is observed frequently in pancreatic cancer but is also an agerelated phenomenon in gastrointestinal tissues. *Clin Cancer Res* 9, 1446–1452.

Matsubayashi, H., Canto, M., Sato, N., Klein, A., Abe, T., Yamashita, K., Yeo, C.J., Kalloo, A., Hruban, R. and Goggins, M. (2006). DNA methylation alterations in the pancreatic juice of patients with suspected pancreatic disease. *Cancer Res* 66, 1208–1217.

Nodit, L., McGrath, K.M., Brody, D., Ohori, N.P., Schoedel, K.E., Finkelstein, S. and Khalid, A. (2005). EUS-FNA microsatellite loss predicts pancreatic endocrine tumor aggressiveness. *Gastrointest Endosc* 61, AB286.

Panzuto, F., Nasoni, S., Falconi, M., Corleto, V.D., Capurso, G., Cassetta, S., Di Fonzo, M., Tornatore, V., Milione, M., Angeletti, S. et al. (2005). Prognostic factors and survival in endocrine tumor patients: Comparison between gastrointestinal and pancreatic localization. *Endocr Relat Cancer* 12, 1083–1092.

Pugliese, V., Pujic, N., Saccomanno, S., Gatteschi, B., Pera, C., Aste, H., Ferrara, G.B. and Nicolo, G. (2001). Pancreatic intraductal sampling during ERCP in patients with chronic pancreatitis and pancreatic cancer: cytologic studies and k-ras-2 codon 12molecular analysis in 47 cases. *Gastrointest Endosc* 54, 595–599.

Sato, N., Fukushima, N., Maitra, A., Iacobuzio-Donahue, C.A., van Heek, N.T., Cameron, J.L., Yeo, C.J., Hruban, R.H. and Goggins, M. (2004). Gene expression profiling identifies genes associated with invasive intraductal papillary mucinous neoplasms of the pancreas. *Am J Pathol* 164, 903–914.

Scarlett, C.J., Smith, R.C., Saxby, A., Nielsen, A., Samra, J.S., Wilson, S.R. and Baxter, R.C. (2006). Proteomic classification of pancreatic adenocarcinoma tissue using protein chip technology. *Gastroenterology* 130, 1670–1678.

Shin, H.J., Lahoti, S. and Sneige, N. (2002). Endoscopic ultrasound-guided fine-needle aspiration in 179 cases: The M.D. Anderson Cancer Center experience. *Cancer* 96, 174–180.

Speel, E.J., Richter, J., Moch, H., Egenter, C., Saremaslani, P., Rutimann, K., Zhao, J., Barghorn, A., Roth, J., Heitz, P.U. et al. (1999). Genetic differences in endocrine pancreatic tumor subtypes detected by comparative genomic hybridization. *Am J Pathol* 155, 1787–1794.

Speel, E.J., Scheidweiler, A.F., Zhao, J., Matter, C., Saremaslani, P., Roth, J., Heitz, P.U. and Komminoth, P. (2001). Genetic evidence for early divergence of small functioning and nonfunctioning endocrine pancreatic tumors: gain of 9Q34 is an early event in insulinomas. *Cancer Res* 61, 5186–5192.

Stelow, E.B., Adams, R.B. and Moskaluk, C.A. (2006). The prevalence of pancreatic intraepithelial neoplasia in pancreata with uncommon types of primary neoplasms. *Am J Surg Pathol* 30, 36–41.

Stewart, C.J., Mills, P.R., Carter, R., O'Donohue, J., Fullarton, G., Imrie, C.W. and Murray, W.R. (2001). Brush cytology in the assessment of pancreatico-biliary strictures: A review of 406 cases. *J Clin Pathol* 54, 449–455.

Stumpf, E., Aalto, Y., Hoog, A., Kjellman, M., Otonkoski, T., Knuutila, S. and Andersson, L.C. (2000). Chromosomal alterations in human pancreatic endocrine tumors. *Genes Chromosomes Cancer* 29, 83–87.

Tada, M., Omata, M., Kawai, S., Saisho, H., Ohto, M., Saiki, R.K. and Sninsky, J.J. (1993). Detection of ras gene mutations in pancreatic juice and peripheral blood of patients with pancreatic adenocarcinoma. *Cancer Res* 53, 2472–2474.

Trumper, L., Menges, M., Daus, H., Kohler, D., Reinhard, J.O., Sackmann, M., Moser, C., Sek, A., Jacobs, G., Zeitz, M. et al. (2002). Low sensitivity of the ki-*ras* polymerase chain reaction for diagnosing pancreatic cancer from pancreatic juice and bile: A multicenter prospective trial. *J Clin Oncol* 20, 4331–4337.

Warner, R.R. (2005). Enteroendocrine tumors other than carcinoid: A review of clinically significant advances. *Gastroenterology* 128, 1668–1684.

Watanabe, H., Yamaguchi, Y., Ha, A., Hu, Y.X., Motoo, Y., Okai, T., Yoshimura, T. and Sawabu, N. (1998). Quantitative determination of K-*ras* mutations in pancreatic juice for diagnosis of pancreatic cancer using hybridization protection assay. *Pancreas* 17, 341–347.

Weber, S.M., Cubukcu-Dimopulo, O., Palesty, J.A., Suriawinata, A., Klimstra, D., Brennan, M.F. and Conlon, K. (2003). Lymphoplasmacytic sclerosing pancreatitis: Inflammatory mimic of pancreatic carcinoma. *J Gastrointest Surg* 7, 129–137.

Wiersema, M.J., Vilmann, P., Giovannini, M., Chang, K.J. and Wiersema, L.M. (1997). Endosonography guided fine-needle aspiration biopsy: Diagnostic accuracy and complication assessment. *Gastroenterology* 112, 1087–1095.

Wilentz, R.E., Geradts, J., Maynard, R., Offerhaus, G.J.A., Kang, M., Goggins, M., Yeo, C.J., Kern, S.E. and Hruban, R.H. (1998). Inactivation of the p16 (INK4A) tumor-suppressor gene in pancreatic duct lesions: Loss of intranuclear expression. *Cancer Res* 58, 4740–4744.

Wilentz, R.E., Iacobuzio-Donahue, C.A., Argani, P., McCarthy, D.M., Parsons, J.L., Yeo, C.J., Kern, S.E. and Hruban, R.H. (2000). Loss of expression of Dpc4 in pancreatic intraepithelial neoplasia: Evidence that DPC4 inactivation occurs late in neoplastic progression. *Cancer Res* 60, 2002–2006.

Yamaguchi, Y., Watanabe, H., Yrdiran, S., Ohtsubo, K., Motoo, Y., Okai, T. and Sawabu, N. (1999). Detection of mutations of p53 tumor suppressor gene in pancreatic juice and its application to diagnosis of patients with pancreatic cancer: Comparison with K-*ras* mutation. *Clin Cancer Res* 5, 1147–1153.

Yamashita, K., Kida, Y., Shinoda, H., Kida, M. and Okayasu, I. (1999). K-*ras* point mutations in the supernatants of pancreatic juice and bile are reliable for diagnosis of pancreas and biliary tract carcinomas complementary to cytologic examination. *Jpn J Cancer Res* 90, 240–248.

Yan, L., McFaul, C., Howes, N., Leslie, J., Lancaster, G., Wong, T., Threadgold, J., Evans, J., Gilmore, I., Smart, H. et al. (2005). Molecular analysis to detect pancreatic ductal adenocarcinoma in high-risk groups. *Gastroenterology* 128, 2124–2130.

Yeh, T.S., Cheng, A.J., Chen, T.C., Jan, Y.Y., Hwang, T.L., Jeng, L.B., Chen, M.F. and Wang, T.C. (1999). Telomerase activity is a useful marker to distinguish malignant pancreatic cystic tumors from benign neoplasms and pseudocysts. *J Surg Res* 87, 171–177.

Zhao, J., Moch, H., Scheidweiler, A.F., Baer, A., Schaffer, A.A., Speel, E.J., Roth, J., Heitz, P.U. and Komminoth, P. (2001). Genomic imbalances in the progression of endocrine pancreatic tumors. *Genes Chromosomes Cancer* 32, 364–372.

CHAPTER

Genomic Evaluation of Head and Neck Cancer

Giovana R. Thomas and Yelizaveta Shnayder

INTRODUCTION

Head and neck squamous cell carcinoma (HNSCC) of the upper aerodigestive tract (UADT) represent approximately 4% of all cancers, with an estimated 40,500 new cases of oral cavity, pharynx and larynx cancer and 12,000 expected deaths in the United States in 2006 (Ries et al., 2006). Approximately half of all patients with HNSCC have advanced stage disease at the time of diagnosis, with an expected 5-year survival rate between 10% and 40%. Despite treatments that may consist of mutilating surgery, radiotherapy, and/or chemotherapy, overall long-term survival remains low due to uncontrollable persistent or recurrent HNSCC. The low rate of survival of patients with locoregional and distant recurrences has highlighted the need for new approaches for diagnosis and treatment.

As a result of exposure to carcinogens, molecular analyses of normal, precancerous, and head and neck cancers have revealed over-time accumulation of specific genetic alterations in proto-oncogenes and tumor suppressor genes (TSG) that are associated with potential for locoregional spread, time to recurrence, and overall survival.

Several experimental models of molecular epithelial carcinogenesis have been proposed. However, there are certain genetic alterations common to these models.

Epithelial carcinogenesis and progression has been characterized by promoter hypermethylation or loss of heterozygosity (LOH) at 9p21 leading to loss of TSG *p16*, mutation at 17p13 leading to loss of *TP53* TSG function, and gene amplification at 11q13 leading to overexpression of cyclin-D1. Importantly, changes in expression levels of EGFR and VEGF have also been associated with malignant progression. DNA microarray technology has been useful in identifying a specific pattern of deregulated network of genes and proteins that may provide a means for early detection, a unique target therapeutic intervention or used to monitor HNSCC disease progression.

In this chapter, we summarize and discuss recent advances in our knowledge concerning the pathogenesis, diagnosis, monitoring, prognosis, and treatment of head and neck cancer obtained by using modern molecular biological tools to study gene expression and chromosomal aberrations. In addition, we discuss novel molecular targeting strategies in the treatment of HNSCC and suggest where future progress may occur.

HEAD AND NECK SQUAMOUS CELL CARCINOMA

Predisposition

The predominant environmental risk factors for developing HNSCC identified to date are the use of alcohol and tobacco, immunosuppression, chewing betel quid nuts, and exposure to high-risk human papilloma virus (HPV). However, not all smokers and drinkers develop cancers; therefore, genetic predisposition and other host factors may play an equally important role in tumorigenesis.

Copyright © 2010, Elsevier Inc.
All rights reserved.

Benzo(a)pyrene diol epoxide (BPDE) is a toxic metabolite of benzo(a)pyrene, the main carcinogenic component of tobacco smoke. BPDE has been shown to exert its mutagenic effect mostly by irreversibly binding to DNA, forming BPDE-DNA adducts, or by DNA oxidation. Other components of tobacco smoke such as catechols and nitrosamines produce free radical compounds known to cause DNA fragmentation. The individual capacity for nucleotide excision and repair of normal DNA after BPDE-induced damage has been shown to correlate with HNSCC development independent of age, sex, ethnicity, smoking status, or alcohol use in a case-control study (Cheng et al., 1998).

Mutagen hypersensitivity or individual's capacity of DNA repair against free radical damage and its relevance to the development of HNSCC was examined by Schantz et al. (1997). The relationship between the risk of HNSCC, nutrition, and mutagen hypersensitivity as determined by the quantity of bleomycin-induced chromosomal breaks within peripheral blood lymphocytes was examined in HNSCC patients and controls. Dietary intake of some antioxidants such as vitamins C, E, and carotenoids; cigarette smoking; alcohol drinking; and body mass index (BMI) were also examined. The study found that mutagen hypersensitivity was strongly associated with increased risk of HNSCC. Also, high intakes of vitamins C and E and some carotenoids were independently related to a decreased risk of HNSCC.

The role of genomic instability on the development of HNSCC was further elucidated in a study of patients with Fanconi anemia, a rare autosomal recessive disorder characterized by a high degree of spontaneous chromosomal aberrations (Kutler et al., 2003). The overall incidence of HNSCC in Fanconi anemia patients was 3% as compared to the incidence of 0.038% in general population, and median age of these patients with HNSCC was 31. The cumulative rate of disease recurrence after treatment by the age of 40 was 50% in this patient population.

In recent years, there has been mounting epidemiologic and experimental evidence of a role for HPV as the etiologic agent in a small proportion of HNSCC, mostly in non-drinkers and non-smokers. A recent study identified high-risk HPV16 DNA inside the nuclei of cancer cells of 90% of HPV-positive oropharyngeal HNSCC specimens by *in situ* hybridization (Gillison et al., 2000). These cancers, as compared to non-HPV-associated oropharyngeal HNSCC, tended to be in non-drinkers, non-smokers, have basaloid features on histology, not associated with p53 mutation, and have a 59% reduction in risk of dying from cancer, after adjustment for stage of disease, morbidity, and other confounding factors. In addition, the association between HPV16 infection and HNSCC in specific sites suggests the strongest and most consistent association is with tonsil cancer, and the magnitude of this association is consistent with an infectious etiology (Hobbs et al., 2006).

Screening

The multistep carcinogenesis process results in epigenetic and metabolic changes that give rise to histologically distinct precursor phenotypes that harbor specific genetic alterations. Pre-malignant lesions of the UADT can present as leukoplakia (white patch) or erythroplakia (red patch). Histologically, leukoplakia may demonstrate benign hyperkeratosis of the surface epithelium, epithelial hyperplasia, dysplasia, or invasive carcinoma. Up to 8% of leukoplastic lesions may demonstrate invasive squamous cell carcinoma (SCC), whereas erythroplakia has a much greater potential for malignancy. Approximately 90% of erythroplastic lesions may demonstrate severe dysplasia, carcinoma *in situ* or invasive SCC. The transformation rate of dysplasia to cancer has been reported as high as 36.4%.

Improvements in overall survival in patients with HNSCC rests on early identification of pre-malignant lesions and intervention in patients at risk prior to development of advanced stage disease. Since epithelial carcinogenesis is a multistep process directed by complex molecular events that involve specific genetic defects in proto-oncogenes and TSG, early identification of genetic alterations that may represent early transition into malignant phenotype may be possible through various measures recently described.

Early detection of oral and oropharyngeal SCC using brush biopsies to detect exfoliated cytology and DNA cytometry have been investigated. Cytologic diagnosis combined with DNA-image cytometry to measure DNA aneuploidy has proven to have high specificity and sensitivity for malignancy in studies of patients evaluated with pre-malignant lesions in the oral cavity (Maraki et al., 2004; Remmerbach et al., 2001). In fact, these investigators have shown that cytology combined with DNA-image cytometry may predict malignant transformation up to 15 months before its histologic confirmation.

The measurement of DNA ploidy in epithelial cells of oral leukoplakia has been studied by Sudbo and colleagues (reviewed in Sudbo and Reith, 2005). These investigators have shown that patients with aneuploid dysplastic oral lesions had a 96% rate of oral cancer with 70% rate within 3 years, an 81% rate of subsequent cancer, and 74% rate of death from cancer. Moreover, newer molecular methods such as microsatellite analysis to detect genetic alterations in exfoliated oral mucosal cells (Spafford et al., 2001) or in serum (Nawroz-Danish et al., 2004) also suggest that this approach is feasible to detect occult carcinoma in patients at risks for developing HNSCC. These studies suggest that biomarkers of genomic instability, such as aneuploidy and allelic imbalance, can accurately measure the cancer risk of oral pre-malignant lesions.

The use of molecular imaging studies for detection of HNSCC is still in the very early pre-clinical stages. Less than a handful of studies have addressed this potential method for detection of HNSCC. Hsu et al. (2005) using a simple fluorescence spectroscopy system developed a molecular-specific contrast agent targeted against EGFR as a screening technique for oral cancer. The system was tested in four *in vitro* models including fresh tissue slices from normal and abnormal oral cavity biopsies and whole normal and abnormal oral cavity biopsies. These investigators proved that an inexpensive and simple spectroscopy system can be used in biological models of living

systems to detect the optical signal from a contrast agent targeted toward a cancer-related biomarker with good signal-to-noise ratios. This system has the potential to improve the early detection of oral neoplasia by providing a low-cost screening tool. Scher (2007) analyzed whether *in vivo* videomicroscopy (IVVM) is useful for the study of distant metastasis and studied the possible role of nitric oxide in the development of metastasis from HNSCC. Videomicroscopic images of a human squamous cell carcinoma cell line (FaDu) labeled with an intracytoplasmic fluorescent marker were analyzed in the microcirculation for cell adhesion, morphology, deformation, circulation, location of adhesion within the microcirculation, and alteration of microvascular circulation. He concluded that IVVM allows direct assessment of circulating HNSCC with the microcirculation and is a powerful model for the study of distant metastasis. In addition, he found that nitric oxide and IL-1 play a role in increasing the arrest of HNSCC in the liver and are important in the generation of distant metastasis in patients with HNSCC.

In a study by El-Sayed et al. (2005), nanoparticle technology was used as potential biosensors in malignant oral epithelial cells. These investigators used a simple and inexpensive technique to record surface plasmon resonance (SPR) scattering images and SPR absorption spectra from both colloidal gold nanoparticles and from gold nanoparticles conjugated to monoclonal anti-epidermal growth factor receptor (anti-EGFR) antibodies after incubation in cell cultures with a nonmalignant epithelial cell line (HaCaT) and two malignant oral epithelial cell lines (HOC 313 clone 8 and HSC 3). They showed that the anti-EGFR antibody conjugated nanoparticles specifically and homogeneously bind to the surface of the cancer type cells with 600% greater affinity than to the noncancerous cells and suggested that SPR scattering imaging or SPR absorption spectroscopy generated from antibody conjugated gold nanoparticles can be useful in molecular biosensor techniques for the diagnosis and investigation of oral epithelial living cancer cells *in vivo* and *in vitro*.

Diagnosis

Because the probability of curing patients with HNSCC is related to the stage of disease, the importance of diagnosing HNSCC at an early stage cannot be overemphasized. Visual inspection is the most common method of detecting oral and oropharyngeal squamous carcinoma. The diagnosis of HNSCC, however, is frequently delayed because symptoms for which patients will seek medical attention such as pain, dysphagia, and shortness of breath occur late in the stage of disease. Laryngeal squamous cell carcinoma remains one of a few subsite that is likely to be detected at earlier stages due to the presenting symptom of persistent hoarseness, which can occur with mild alterations of the vibratory surfaces of the true vocal cords.

Radiologic imaging modalities such as computed tomography (CT), magnetic resonance imaging, positron-emission tomography (PET), combined PET/CT, ultrasound, and lymphoscintigraphy are critical tools that provide information on extent of tissue invasion, involvement of regional lymph nodes, and presence of distant metastatic disease. The information provided by imaging modalities is critical for staging and subsequent treatment planning. However, there remains significant variability in outcomes for patients within the same tumor-node-metastasis (TNM) stage. Therefore, many attempts have been made to identify high-risk patients through molecular markers.

Exposure to carcinogens such as tobacco and alcohol may result in pre-malignant epithelial changes over a wide field of epithelial surface within the aerodigestive tract. This clinical phenomenon has been called "field cancerization" and may lead to frequent occurrence of multiple primary tumors in epithelial areas affected by widespread pre-malignant disease and possibility of distant-related primary tumors in UADT. For this reason, it is imperative to evaluate mucosa of the UADT including the esophagus and trachea with directed biopsies of suspicious areas during endoscopic examination.

By identifying patients at risk for cervical lymph node metastasis and extracapsular spread without the need for surgical node dissection, tissue or serum biomarkers can play a vital role in clinical decision-making. The molecular profiling of primary tumors from HNSCC has been investigated for the potential of predicting the presence of lymph node metastasis at the time of diagnosis. DNA microarray gene expression of primary tumors of the oral cavity and oropharynx found that signature or predictor gene sets can detect local lymph node metastases using material from primary HNSCC with better performance than current clinical diagnosis (O'Donnell et al., 2005; Roepman et al., 2005). Additionally, the gene expression profile of 53 genes with roles in cell differentiation, adhesion, signal transduction, and transcription regulation have been associated with depth of invasion in patients with oral SCC (Toruner et al., 2004).

A large number of genes that are abnormally expressed compared to control have been found in tumors from patients with oral SCC. A large percentage of these were upregulated proteins including metalloproteinases (MMPs), proteins involved in regulation of cell adhesion and cell-cycle-related proteins; and downregulation of potential tumor suppressors, suggesting that oral SCC show aberrant expression of genes involved in proliferation, apoptosis, extracellular matrix degradation, and other cellular pathways (Kornberg et al., 2005).

Although expression signatures may potentially be able to accurately identify subjects harboring HNSCC and may represent an advance in the classification of these tumors, application of this methodology to the clinical setting has not yet been successfully put to practice.

Prognosis

The most important factor correlating with prognosis of HNSCC currently remains TNM staging. Large primary tumor size, presence of lymph node metastasis, positive margins after surgical excision, and perineural invasion have been reliable indicators of poor clinical outcome in patients with HNSCC. However, the single most important factor that determines survival is the metastatic status of the cervical lymph nodes at the time of diagnosis. Particularly, the presence of extracapsular spread in cervical lymph node metastasis remains the most

Table 40.1 Prognostic markers in HNSCC

Marker	Chromosome	Function	Mutation in HNSCC	Result of mutation	Method of study in HNSCC tumors	Correlation with:	Possible clinical applications	References
p53	17p13	Tumor suppressor	50–69%	Overexpression	IHC	Decreased disease-free survival, overall survival in larynx SCC	Cisplatin resistance: consider other agents Increased radiosensitivity	Cabelquenne et al. (2000); Ogawa et al. (1998)
EGFR		Transmembrane tyrosine kinase	34–80%	Overexpression	IHC	Short disease free survival and overall survival	New chemotherapy agents: 1. Inhibitors of tyrosine kinase (Gefitinib, Erlotinib), 2. Antibodies against EGFR (Cetuximab)	Grandis and Tweardy (1993); Bonner et al. (2006)
VEGF		Angiogenic factor		Overexpression	IHC	Poor overall survival	Antibodies against VEGF-A (Bevacizumab)	Teknos et al. (2002); Caponigro et al. (2005)
Cyclin D1	11q13	Proto-oncogene	17–79%	Gene amplification Overexpression	IHC FISH RT-PCR	Tumor extension, regional lymph node metastases, advanced clinical stage		Capaccio et al. (2000); Bova et al. (1999)
P16	9p21	Tumor suppressor	52–82%	Loss of expression	IHC RT-PCR	Disease relapse		Bazan et al. (2002); Yuen et al. (2002)

significant clinical prognostic indicator of survival, local–regional recurrence and distant metastasis in patients with HNSCC. Although these clinical prognostic parameters provide the best possible criteria for deciding treatment modalities, they are limited in discerning future behavior of aggressive HNSCC.

The search for novel molecular prognostic markers with potentially significant predictive value for biological aggressiveness of HNSCC has exploded in the recent years. Better prediction of the risk of developing distant metastases would help introduce a more selective treatment approach according to the biological aggressiveness of the tumor. Several molecular mediators of tumor progression, invasion, and metastasis, which function in growth factor signaling, metastasis, and suppressor genes have been well investigated in HNSCC (Table 40.1) and will be described below.

TP53 is a tumor suppressor gene located on 17p13 and consists of 11 exons that encode a protein, p53, and function in carcinogenesis by initiating G1 arrest in response to certain DNA damage and apoptosis. The prevalence of *TP53* mutations is 50–69% in HNSCC. A number of studies have shown that *TP53* gene mutations are associated with increased risk for locoregional recurrence and poor outcome (Erber et al., 1998; Mineta et al., 1998). Mutated p53 protein overexpression has also been associated with tumor recurrence, poor disease-free survival rate, poor overall survival rate, increased rates of local control, and decreased disease-free survival in HNSCC. A significant correlation between p53 expression and clinical outcome appears to be the strongest in the subsite of patients with laryngeal SCC. In these patients, overexpression of mutated p53 predicts poor disease-free and poor overall survival rates.

Epidermal growth factor receptor (EGFR) is a transmembrane tyrosine kinase capable of promoting neoplastic transformation. The downstream signaling events upon ligand binding include activation of tyrosine kinase and activation of intracellular Ras, Raf, and mitogen-activated protein kinase (MAPK) cascades. EGFR expression has been extensively studied in HNSCC and its overexpression reported in 34–80% of HNSCC using IHC (Beckhardt et al., 1995; Grandis and Tweardy, 1993). This marker has been significantly associated with short disease-free survival, overall survival, and poor prognosis in patients with HNSCC. The EGFR family includes EGFR (*c-erbB1 or Her1*), *c-erbB2* (Her2-*neu*), *c-erbB3* (Her3), and *c-erbB4* (Her4), which have the ability to form receptor heterodimers and crosstalk between them. Both EGF and transforming-growth factor-alpha (TGF-α) are ligands that bind to EGFR. Co-expression of Her2 and Her3, which do not have intrinsic tyrosine kinase activity, is reported to increase transforming activity and their expression has been strongly associated with shortened patient survival. In addition, the co-expression of Her2 and Her3 significantly improved the predictive power in patients with oral SCC and suggest a common regulatory mechanism.

Vascular endothelial growth factor (VEGF) induces proliferation, migration, and survival of endothelial cells during tumor growth by binding to specific tyrosine receptor kinases. Thus far, six members of the VEGF family have been identified; VEGF-A/vascular permeability factor (four isoforms), VEGF-B/VEGF-related factor (VRF), VEGF-C/VEGF-related protein (VRP), VEGF-D/c-fos-induced growth factor (FIGF), VEGF-E and placenta growth factor (PlGF) and abnormal regulation of angiogenic factors have been implicated in the pathogenesis of cancer. Overexpression of VEGF-A and VEGF-C has been correlated to poor overall survival in patients with advanced stage HNSCC (Teknos et al., 2002). Furthermore, pre-treatment levels of VEGF-A and/or VEGF-C in early stage disease appear to be independent predictors of presence of cervical lymph node metastasis (O-Charoenrat et al., 2001).

Cyclin D1 (*CCND1*) is a proto-oncogene located on chromosome 11q13 that modulates a critical step in cell cycle control progression. Amplification and/or overexpression of *CCND1* have been demonstrated in 17–79% of tumor specimens from patients with HNSCC with immunohistochemistry, FISH, or RT-PCR. *CCND1* amplification and/or overexpression of the structurally normal protein has been shown to correlate significantly with tumor extension, regional lymph node metastases and advanced clinical stage of HNSCC (Capaccio et al., 2000). In addition, many recent studies show convincing data of *CCND1* aberration as a prognostic marker for disease-free survival and overall survival in patients with this disease (Bova et al., 1999; Michalides et al., 1995).

The role of *P16* in HNSCC carcinogenesis and progression has not been clearly established. The p16 protein, which is encoded by the *CDKN2A* (*MTS1*, *INK4A*) tumor suppressor gene on chromosome 9p21, inactivates the function of cdk4-cdk6-cyclin D complexes. Using fluorescence *in situ* hybridization (FISH) or PCR-based techniques, *P16* deletion or mutation was detected in 48–52% of tumor specimens from patients with HNSCC and has been significantly associated with decreased survival, development of distant metastases, and as an independent predictive factor for disease relapse and death (Bazan et al., 2002). However, downregulation of p16, although found in 48% of tumors and associated with a more locally advanced tumor, had no prognostic significance for nodal metastasis and survival in studies examining p16 expression by immunohistochemical techniques or *P16* gene alterations by RT-PCR in specimens of patients with HNSCC (Danahey et al., 1999; Yuen et al., 2002). The contrast in these study outcomes may not only reflect differences in the type of tissue and methods used for examination, but also depend on differences in treatment protocols.

In larger studies of patients with HNSCC, DNA microarray analysis has been used to identify distinct gene expression signatures associated with clinical outcome in HNSCC. Distinct subtypes of HNSCC with different clinical outcomes based on gene expression patterns obtained from tumor samples from patients with HNSCC were recently described. These subtypes had significant differences in clinical outcomes including recurrence-free survival and overall survival, and were able to identify patterns of expression that could predict the presence of lymph node metastases in HNSCC tumors (Chung et al., 2004).

Pharmacogenomics

The response to chemotherapy, radiation and their side effects in patients with head and neck cancer is dependent on several factors including pharmacogenetics (see Chapter 15).

Organ preservation therapy using chemotherapy with cisplatin-based chemotherapy and 5-fluorouracil combined with radiotherapy have been used effectively in patients with advanced HNSCC. In addition, studies have shown that patients that achieve a complete response to chemotherapy have a better prognosis than those that do not. Ideally, molecular markers could be used to determine individual pharmacogenetic profiles to identify patients most likely to have chemotherapeutic benefit and patients with the highest risk of suffering genotoxic side effects. These profiles will ideally lead to individualized therapies, improved treatment outcomes, and a movement toward clinically applied pharmacogenetics.

Cisplatin (*cis*-diamminedichloroplatinum [CDDP])-based chemotherapy as part of multimodality treatment has shown significant activity against HNSCC, as a DNA-damaging agent. However, its effectiveness in the treatment of recurrent HNSCC has been limited due to acquired or intrinsic resistance. The mechanisms of resistance to cisplatin-based chemotherapy are not well understood. Molecular markers involved in cisplatin resistance phenotype has been investigated by various groups in cisplatin resistant and sensitive HNSCC cell lines using cDNA microarray analysis. Upregulation of the gene expression of glycoprotein hormone alpha subunit and downregulation of human folate receptor and tumor-associated antigen L6 were described (Higuchi et al., 2003). Decreased expression of caveolin-1, a novel TSG correlated with cisplatin resistance in patients with oral SCC (Nakatani et al., 2005).

Resistance to cisplatin chemotherapy has also been shown to be significantly correlated to expression of mutant p53 (Bradford et al., 2003), which is frequently present in head and neck cancer. Tumor cell lines from patients with HNSCC with mutant p53 have shown significantly more sensitivity to cisplatin than those that do not contain these mutations. The authors of these studies state that tumors that are resistant to cisplatin also overexpress anti-apoptotic proteins Bcl-2 and Bcl-$x_{L.}$ Other studies have shown that p53 mutated tumors was higher in the group of patients with nonresponse to cisplatin-fluorouracil neoadjuvant chemotherapy than in the group of responders (Cabelquenne et al., 2000). Overexpression of mutant-type p53 expression in HNSCC was also shown to be associated with increased sensitivity to ionizing radiation (Ogawa et al., 1998; Servomaa et al., 1996).

Recently, low expression of the apoptosis-blocking protein family members has been shown to be a good predictor of chemotherapy response in patients with head and neck cancer. Low expression of Bcl-x_L in tumor specimens from patients with HNSCC enrolled in the Department of Veterans Affairs Laryngeal Cancer Group Study correlated with response to induction chemotherapy (Bradford et al., 2003). In addition, it was shown that induction of mutant p53 in HNSCC lines resulted in decreased expression of Bcl-2 and increased susceptibility to cisplatin-induced apoptosis and implicates Bcl-2 in the deregulation of p53-induced apoptosis (Andrews et al., 2004).

Increased tumor resistance to cytotoxic agents, including radiotherapy (Milas et al., 2003, 2004; Baumann and Krause, 2004) has been associated with EGFR overexpression in HNSCC in addition to its association with more aggressive tumor behavior. This suggest that evaluation of EGFR status at the time of diagnosis may help identify subset of patient who are at increased risk of neck node metastasis, may have an unfavorable treatment outcome with radiotherapy and may, therefore, benefit from more aggressive treatments.

Monitoring

Patients treated for HNSCC are followed clinically for evidence of recurrent disease, development of second primary lesions or distant metastasis. The chance of developing a second primary tumor has been estimated at 2–3% per year in patients with HNSCC. In addition, 20–30% of patients treated for HNSCC will develop recurrent disease at primary site and is the most common cause of treatment failure. Because prognosis of late stage recurrent disease is dismal, early detection is imperative. Distant metastatic disease occurs in 11–15% of patients treated for HNSCC and, at this stage, treatment is palliative. Identifying molecular markers in primary tumors that are associated with locoregional relapse may allow for early identification of patients needing additional surveillance and treatment and may have the potential of decreasing probability of distant disease. For example, EGFR overexpression has been shown to be an independent prognostic factor for neck node relapse in primary specimens of patients with laryngeal SCC undergoing primary resection (Almadori et al., 1999). In addition, when surgical margins of primary HNSCC are examined for mutational changes, there is an increased risk of local recurrence when positive margins demonstrating clonal alterations in *TP53* are observed (van Houten et al., 2004).

Promoter hypermethylation is an important mechanism to silence TSG in cancer. Aberrant DNA methylation of *P16*, O6-methylguanine-DNA-methyltransferase, glutathione S-transferase P1, and death-associated protein kinase (DAP-kinase), key genes involved in critical pathways in head and neck tumor progression, have been detected in the serum and in the saliva of patients with HNSCC (Sanchez-Cespedes et al., 2000; Rosas et al., 2001). Thus, promoter hypermethylation of key genes may be promising serum and saliva markers for monitoring affected patients with HNSCC.

Gene expression signatures using DNA microarray technology have potential utility as biomarkers to predict patients at risk for locoregional recurrence. Several gene expression signatures from HNSCC tumors from various anatomical sites in the head and neck have been identified. In one study, deregulated gene expression of the met-proto-oncogene and its ligand hepatocyte growth factor/scatter factor; snail homologue 2 (SNA12/SLUG), a transcriptional repressor involved in epithelial/mesenchymal transitions; and genes important for tumor cell/extracellular matrix interactions such as laminins and integrins were associated with risk of local treatment failure (Ginos et al.,

2004). In addition, six novel poorly characterized differentially genes potentially involved in acquisition of metastatic potential were identified when tumors from patients with SCC of the hypopharynx were studied for identification of biomarkers of aggressive clinical behavior (Cromer et al., 2004). These studies show the feasibility of monitoring HNSCC patients at risk for developing locoregional recurrence by molecular characterization of genes in primary tumors or in serum and saliva of patients with HNSCC.

Because surgical salvage rates are greatly diminished when occult nodal disease becomes clinically manifest, planned post-treatment neck dissection is advocated but may not be necessary in all patients. The role of PET-CT in this scenario remains unproven but holds promise in being able to identify which patients with advanced nodal disease prior to treatment may be harboring residual disease in the neck after chemoradiotherapy. These patients may then go on to planned post-treatment neck dissection. The implementation of as yet unidentified molecular tumor markers in combination with PET-CT may ultimately prove to be effective in identifying patients who will best benefit from post-therapy neck dissection.

Proteomics and Metabolomics of Head and Neck Cancer

Whereas genomics offers the opportunity to examine gene expression or variation in gene sequence, proteomics encompasses evaluation of protein expression, activation, modification, and degradation and targets protein function. Likely both proteomics and genomics will provide clinically useful and complementary information that will speed scientific understanding of HNSCC. Proteomic profiling of serum using surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF-MS) is an emerging technique to identify new biomarkers in biological fluids and to establish clinically useful diagnostic computational models.

Proteomics of HNSCC have revealed significant findings. Baker et al. (2005) combined laser-capture microdissection (LCM) with liquid chromatography-tandem mass spectrometry (LC-MS/MS) to identify proteins expressed in histologically normal squamous epithelium and matching SCC. Immunohistochemical analysis of HNSCC tissues revealed lack of keratin 13 in tumor tissues and strong staining in normal epithelia, and high expression of Hsp90 in tumors. Likewise, cellular proteomes of 30 matched normal squamous epithelial cells and carcinoma specimens were analyzed after tissue microdissection using microarray composed of 83 different antibodies by Weber et al. (2007). Of the 83 proteins examined, 14 showed differential expression between HNSCCs and normal epithelium. The protein microarray approach revealed an upregulation of eight proteins and a downregulation of six proteins. Bag-1, Cox-2, Hsp-70, Stat3, pescadillo, MMP-7 (matrilysin), IGF-2, and cyclin D1 were identified to be significantly upregulated, whereas suppressor of cytokine signaling 1, thrombospondin, TGF-beta1, Jun, Fos, and Fra-2 were downregulated. Upon correlation of differentially regulated proteins with the clinicopathologic data of their patients, MMP-7 (matrilysin) was found to be associated with survival in univariate, but not multivariate, analysis.

Protein expression profiles from two laryngeal carcinoma specimens and corresponding adjacent normal tissue were recently investigated as potential biomarkers which may contribute to the pathogenesis of laryngeal carcinoma using two-dimensional differential in-gel electrophoresis and mass spectroscopy. Differentially expressed proteins were identified, and they included stratifin, S100 calcium-binding protein A9, p21-ARC, stathmin, and enolase (Sewell et al., 2007). Protein expression profiles were also analyzed by Roesch-Ely et al. (2007) in 113 HNSCCs and 73 healthy, 99 tumor-distant, and 18 tumor-adjacent squamous mucosa by SELDI-TOF-MS on IMAC30 ProteinChip Arrays. Calgizarrin (S100A11), the Cystein proteinase inhibitor Cystatin A, Acyl-CoA-binding protein, Stratifin (14-3-3 sigma), Histone H4, alpha- and beta-Hemoglobin, a C-terminal fragment of beta-hemoglobin and the alpha-defensins 1-3 were identified by mass spectrometry. Comparison of the protein profiles in the tumor-distant-samples with clinical outcome of 32 patients revealed a significant association between aberrant profiles with tumor relapse events.

Zhou et al. (2006) studied 100 serum samples including 48 from HSCC patients and 52 from normal controls by SELDI-TOF-MS. They found that 45 potential biomarkers could differentiate HSCC patients from normal controls.

Metabolomics, the study of metabolite changes in a biological system, is believed to be a good reflection of the phenotype of any cell or tissue, and studies have reported that tumoral tissue differ in metabolic content from their benign counterpart. Proton nuclear magnetic resonance (HNMR) spectroscopy is emerging as a rapid and noninvasive method in identifying new markers potentially useful for clinical diagnosis in biofluids, such as urine and plasma. However, this new -omic science has not yet been significantly evaluated in clinical studies in HNSCC. Although, metabolomics may be a complimentary tool to genomics and proteonomics in diagnosing cancer, refining the process of targeted drug discovery, defining surrogate markers that may predict response, toxicity, and prognosis, several years of study will be necessary before the impact of this new technology in HNSCC can be truly assessed.

GENOMICS OF HNSCC: CLINICAL APPLICATIONS

Genomic technologies have the potential to change the basis of clinical oncologic practice in various ways. Information on the disease pathology of HNSCC generated from genomic technologies could be translated into: (1) diagnostic tests of biomarkers of cellular proliferation, cell damage or death, and cellular metabolism, and (2) therapeutic products, as is seen with the development of EGFR antagonists, which are presently in clinical practice for treatment of HNSCC.

Genomic approaches are also changing the conduct of clinical trials. Many clinical trials of HNSCC are now being designed

with genomic testing in mind. Importantly, biomarkers in tissue or body fluids identified by genomics offer a means for detecting early or recurrent disease in patients that are at risk for HNSCC. This technology has the added potential to characterize individuals that are likely to have a response to chemotherapy and/or radiotherapy, so that treatment can be tailored to the individual patient. As a preventive approach, genome-based information of the potential risk for HNSCC may allow early intervention with preventive measures, which may ultimately reduce healthcare costs.

The economic value of using genomic technology in clinical setting and in clinical trials have not been fully addressed. However, because of the potential advantages in targeted drug development and drug-safety concerns, significant growth of molecular approaches to clinical medicine will continue. Therefore, the combination of genomic, proteomic, and metabolomic technologies may allow us to monitor a large number of key cellular pathways simultaneously and may provide a better understanding of HNSCC.

2009 UPDATE

Recent data continue to support a strong etiologic role for HPV in the pathogenesis of HNSCC (D'Souza et al., 2007; Ernster et al., 2007). Studies have further defined distinct clinical behavior between HPV-positive and HPV-negative cancers. HPV-positive cancers arise mainly from the lingual and palatine tonsils (oropharynx), are associated with distinct risk factor profile of sexual behavior and marijuana use, typically occur in patients younger than 60 years of age, and are associated with an improved prognosis when compared to HPV-negative patients (Charfi et al., 2008; Chaturvedi et al., 2008; Dahlstrand et al., 2008; Fakhry et al., 2008; Gillison et al., 2008; Hafkamp et al., 2008). Presence of antibodies to HPV-16 E6 and E7 oncogenes in serum has also been associated with better clinical outcomes (Smith et al., 2008), suggesting their role as potential biomarkers for diagnosis and response to treatment for HPV-associated tumors. The p73 variant and p53 codon 72 variant genotypes show promise as potential markers of genetic susceptibility to HPV-16-associated HNSCC of the oropharynx, particularly in never smokers and never drinkers (Chen et al., 2008; Ji et al., 2008), suggesting that p73 and p53 polymorphism may modulate the risk of HPV-16-associated HNSCC of the oropharynx.

To summarize, recent molecular studies suggest that two distinct pathways may be involved in the pathogenesis of HNSCC of the oropharynx: one driven by alcohol and tobacco and associated with cytogenetic changes, such as p53 mutations, loss of p16, cyclin D1 overexpression, and increased numbers of EGFR; and the other driven by HPV-induced genomic instability with minimal molecular alterations.

2009 UPDATE REFERENCES

Charfi, L., Jouffroy, T., de Cremoux, P., Le Peltier, N., Thioux, M., Fréneaux, P., Point, D., Girod, A., Rodriguez, J. and Sastre-Garau, X. (2008). Two types of squamous cell carcinoma of the palatine tonsil characterized by distinct etiology, molecular features and outcome. *Cancer Lett* 260(1–2), 72–78.

Chaturvedi, A.K., Engels, E.A., Anderson, W.F. and Gillison, M.L. (2008). Incidence trends for human papillomavirus-related and -unrelated oral squamous cell carcinomas in the United States. *J Clin Oncol* 26(4), 612–619.

Chen, X., Sturgis, E.M., El-Naggar, A.K., Wei, Q. and Li, G. (2008). Combined effects of the p53 codon 72 and p73 G4C14-to-A4T14 polymorphisms on the risk of HPV16-associated oral cancer in never-smokers. *Carcinogenesis* 29(11), 2120–2125.

D'Souza, G., Kreimer, A.R., Viscidi, R., Pawlita, M., Fakhry, C., Koch, W.M., Westra, W.H. and Gillison, M.L. (2007). Case-control study of human papillomavirus and oropharyngeal cancer. *N Engl J Med* 356(19), 1944–1956.

Dahlstrand, H., Näsman, A., Romanitan, M., Lindquist, D., Ramqvist, T. and Dalianis, T. (2008). Human papillomavirus accounts both for increased incidence and better prognosis in tonsillar cancer. *Anticancer Res* 28(2B), 1133–1138.

Ernster, J.A., Sciotto, C.G., O'Brien, M.M., Finch, J.L., Robinson, L.J., Willson, T. and Mathews, M. (2007). Rising incidence of oropharyngeal cancer and the role of oncogenic human papilloma virus. *Laryngoscope* 117(12), 2115–2128.

Fakhry, C., Westra, W.H., Li, S., Cmelak, A., Ridge, J.A., Pinto, H., Forastiere, A. and Gillison, M.L. (2008). Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. *J Natl Cancer Inst* 100(4), 261–269.

Gillison, M.L., D'Souza, G., Westra, W., Sugar, E., Xiao, W., Begum, S. and Viscidi, R. (2008). Distinct risk factor profiles for human papillomavirus type 16-positive and human papillomavirus type 16-negative head and neck cancers. *J Natl Cancer Inst* 100(6), 407–420.

Hafkamp, H.C., Manni, J.J., Haesevoets, A., Voogd, A.C., Schepers, M., Bot, F.J., Hopman, A.H., Ramaekers, F.C. and Speel, E.J. (2008). Marked differences in survival rate between smokers and nonsmokers with HPV 16-associated tonsillar carcinomas. *Int J Cancer* 122(12), 2656–2664.

Ji, X., Neumann, A.S., Sturgis, E.M., Adler-Storthz, K., Dahlstrom, K.R., Schiller, J.T., Wei, Q. and Li, G. (2008). p53 codon 72 polymorphism associated with risk of human papillomavirus-associated squamous cell carcinoma of the oropharynx in never-smokers. *Carcinogenesis* 29(4), 875–879.

Smith, E.M., Rubenstein, L.M., Ritchie, J.M., Lee, J.H., Haugen, T.H., Hamsikova, E. and Turek, L.P. (2008). Does pretreatment seropositivity to human papillomavirus have prognostic significance for head and neck cancers? *Cancer Epidemiol Biomarkers Prev* 17(8), 2087–2096.

REFERENCES

Almadori, G., Cadoni, G., Galli, J., Ferrandina, G., Scambia, G., Exarchakos, G., Paludetti, G. and Ottaviani, F. (1999). Epidermal growth factor receptor expression in primary laryngeal cancer: an independent prognostic factor of neck node relapse. *Int J Cancer* 20(84 2), 188–191.

Andrews, G.A., Xi, S., Pomerantz, R.G., Lin, C.J., Gooding, W.E., Wentzel, A.L., Wu, L., Sidransky, D. and Grandis, J.R. (2004). Mutation of p53 in head and neck squamous cell carcinoma correlates with Bcl-2 expression and increased susceptibility to cisplatin-induced apoptosis. *Head Neck* 26(10), 870–877.

Baker, H., Patel, V., Molinolo, A.A., Shillitoe, E.J., Ensley, J.F., Yoo, G.H., Meneses-Garcia, A., Myers, J.N., El-Naggar, A.K., Gutkind, J.S. et al. (2005). Proteome-wide analysis of head and neck squamous cell carcinomas using laser-capture microdissection and tandem mass spectrometry. *Oral Oncol* 41(2), 183–199.

Baumann, M. and Krause, M. (2004). Targeting the epidermal growth factor receptor in radiotherapy: Radiobiological mechanisms, preclinical and clinical results. *Radiother Oncol* 72(3), 257–266.

Bazan, V., Zanna, I., Migliavacca, M., Sanz-Casla, M., Maestro, M., Corsale, S., Macaluso, M., Dardanoni, G., Restivo, S., Quintela, P. et al. (2002). Prognostic significance of p16INK4a alterations and 9p21 loss of heterozygosity in locally advanced laryngeal squamous cell carcinoma. *J Cell Physiol* 192, 286–293.

Beckhardt, R., Kiyokawa, N., Xi, L., Liu, T., Hung, M., el-Naggar, A., Zhang, H. and Clayman, G. (1995). HER-2/neu oncogene characterization in head and neck squamous cell carcinoma. *Arch Otolaryngol Head Neck* 121(11), 1265–1270.

Bonner, J.A., Harari, P.M., Giralt, J., Azarnia, N., Shin, D.M., Cohen, R.B., Jones, C.U., Sur, R., Raben, D., Jassem, J. et al. (2006). Radiotherapy plus cetuximab for squamous-cell carcinoma of the head and neck. *N Engl J Med* 9(354 6), 567–578.

Bova, R., Quinn, D., Nankervis, J., Cole, I., Sheridan, B., Jensen, M., Morgan, G., Hughes, C. and Sutherland, R. (1999). Cyclin D1 and p16INK4A expression predict reduced survival in carcinoma of the anterior tongue. *Clin Cancer Res* 5(10), 2810–2819.

Bradford, C.R., Zhu, S., Ogawa, H., Ogawa, T., Ubell, M., Narayan, A., Johnson, G., Wolf, G.T., Fisher, S.G. and Carey, T.E. (2003). P53 mutation correlates with cisplatin sensitivity in head and neck squamous cell carcinoma lines. *Head Neck* 25(8), 654–661.

Cabelguenne, A., Blons, H., de Waziers, I., Carnot, F., Houllier, A.M., Soussi, T., Brasnu, D., Beaune, P., Laccourreye, O. and Laurent-Puig, P. (2000). p53 alterations predict tumor response to neoadjuvant chemotherapy in head and neck squamous cell carcinoma: a prospective series. *J Clin Oncol* 7, 1465–1473.

Capaccio, P., Pruneri, G., Carboni, N., Pagliari, A.V., Quatela, M., Cesana, B.M. and Pignataro, L. (2000). Cyclin D1 expression is predictive of occult metastases in head and neck cancer patients with clinically negative cervical lymph nodes. *Head Neck* 22(3), 234–240.

Caponigro, F., Formato, R., Caraglia, M., Normanno, N. and Iaffaioli, R.V. (2005). Monoclonal antibodies targeting epidermal growth factor receptor and vascular endothelial growth factor with a focus on head and neck tumors. *Curr Opin Oncol* 17(3), 212–217.

Cheng, L., Eicher, S.A., Guo, Z., Hong, W.K., Spitz, M.R. and Wei, Q. (1998). Reduced DNA repair capacity in head and neck cancer patients. *Cancer Epidemiol Biomarkers Prev* 7(6), 465–468.

Chung, C., Parker, J., Karaca, G., Wu, J., Funkhouser, W., Moore, D., Butterfoss, D., Xiang, D., Zanation, A., Yin, X. et al. (2004). Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. *Cancer Cell* 5(5), 489–500.

Cromer, A., Carles, A., Millon, R., Ganguli, G., Chalmel, F., Lemaire, F., Young, J., Dembele, D., Thibault, C., Muller, D. et al. (2004). Identification of genes associated with tumorigenesis and metastatic potential of hypopharyngeal cancer by microarray analysis. *Oncogene* 1(23 14), 2484–2498.

Danahey, D.G., Tobin, E.J., Schuller, D.E., Bier-Laning, C.M., Weghorst, C.M. and Lang, J.C. (1999). p16 mutation frequency and clinical correlation in head and neck cancer. *Acta Otolaryngol* 2, 285–288.

El-Sayed, I.H., Huang, X. and El-Sayed, M.A. (2005). Surface plasmon resonance scattering and absorption of anti-EGFR antibody conjugated gold nanoparticles in cancer diagnostics: applications in oral cancer. *Nano Lett* 5(5), 829–834.

Erber, R., Conradt, C., Homann, N., Enders, C., Finckh, M., Dietz, A., Weidauer, H. and Bosch, F. (1998). TP53 DNA contact mutations are selectively associated with allelic loss and have a strong clinical impact in head and neck cancer. *Oncogene* 16(13), 1671–1679.

Gillison, M.L., Koch, W.M., Capone, R.B., Spafford, M., Westra, W.H., Wu, L., Zahurak, M.L., Daniel, R.W., Viglione, M., Symer, D.E. et al. (2000). Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. *J Natl Cancer Inst* 3(92 9), 709–720.

Ginos, M.A., Page, G.P., Michalowicz, B.S., Patel, K.J., Volker, S.E., Pambuccian, S.E., Ondrey, F.G., Adams, G.L. and Gaffney, P.M. (2004). Identification of a gene expression signature associated with recurrent disease in squamous cell carcinoma of the head and neck. *Cancer Res* 64(1), 55–63.

Grandis, J. and Tweardy, D. (1993). Elevated levels of transforming growth factor alpha and epidermal growth factor receptor messenger RNA are early markers of carcinogenesis in head and neck cancer. *Cancer Res* 53(15), 3579–3584.

Grandis, J., Tweardy, D. and Melhem, M. (1998). Asynchronous modulation of transforming growth factor alpha and epidermal growth factor receptor protein expression in progression of premalignant lesions to head and neck squamous cell carcinoma. *Clin Cancer Res* 4(1), 13–20.

Higuchi, E., Oridate, N., Furuta, Y., Suzuki, S., Hatakeyama, H., Sawa, H., Sunayashiki-Kusuzaki, K., Yamazaki, K., Inuyama, Y. and Fukuda, S. (2003). Differentially expressed genes associated with CIS-diamminedichloroplatinum (II) resistance in head and neck cancer using differential display and CDNA microarray. *Head Neck* 25(3), 187–193.

Hobbs, C.G., Sterne, J.A., Bailey, M., Heyderman, R.S., Birchall, M.A. and Thomas, S.J. (2006). Human papillomavirus and head and neck cancer: A systematic review and meta-analysis. *Clin Otolaryngol* 31(4), 259–266.

Hsu, E.R., Gillenwater, A.M. and Richards-Kortum, R.R. (2005). Detection of the molecular changes associated with oral cancer using a molecular-specific fluorescent contrast agent and single-wavelength spectroscopy. *Appl Spectrosc* 59(9), 1166–1173.

Kornberg, L.J., Villaret, D., Popp, M., Lui, L., McLaren, R., Brown, H., Cohen, D., Yun, J. and McFadden, M. (2005). Gene expression profiling in squamous cell carcinoma of the oral cavity shows abnormalities in several signaling pathways. *Laryngoscope* 115(4), 690–698.

Kutler, D.I., Auerbach, A.D., Satagopan, J., Giampietro, P.F., Batish, S.D., Huvos, A.G., Goberdhan, A., Shah, J.P. and Singh, B. (2003). High

incidence of head and neck squamous cell carcinoma in patients with Fanconi anemia. *Arch Otolaryngol Head Neck Surg* 129(1), 106–112.

Maraki, D., Becker, J. and Boecking, A. (2004). Cytologic and DNA-cytometric very early diagnosis of oral cancer. *J Oral Pathol Med* 33(7), 398–404.

Michalides, R., van Veelen, N., Hart, A., Loftus, B., Wientjens, E. and Balm, A. (1995). Overexpression of cyclin D1 correlates with recurrence in a group of forty-seven operable squamous cell carcinomas of the head and neck. *Cancer Res* 55(5), 975–978.

Milas, L., Mason, K.A. and Ang, K.K. (2003). Epidermal growth factor receptor and its inhibition in radiotherapy: *In vivo* findings. *Int J Radiat Biol* 79(7), 539–545.

Milas, L., Fan, Z., Andratschke, N.H. and Ang, K.K. (2004). Epidermal growth factor receptor and tumor response to radiation: *In vivo* preclinical studies. *Int J Radiat Oncol Biol Phys* 58(3), 966–971.

Mineta, H., Borg, A., Dictor, M., Wahlberg, P., Akervall, J. and Wennerberg, J. (1998). P53 mutation, but not p53 overexpression, correlates with survival in head and neck squamous cell carcinoma. *Br J Cancer* 78(8), 1084–1090.

Nakatani, K., Wada, T., Nakamura, M., Uzawa, K., Tanzawa, H. and Fujita, S. (2005). Expression of caveolin-1 and its correlation with cisplatin sensitivity in oral squamous cell carcinoma. *J Cancer Res Clin Oncol* 131(7), 445–452.

Nawroz-Danish, H., Eisenberger, C.F., Yoo, G.H., Wu, L., Koch, W., Black, C., Ensley, J.F., Wei, W.Z. and Sidransky, D. (2004 Aug 10). Microsatellite analysis of serum DNA in patients with head and neck cancer. *Int J Cancer* 111(1), 96–100.

O-Charoenrat, P., Rhys-Evans, P. and Eccles, S. (2001). Expression of vascular endothelial growth factor family members in head and neck squamous cell carcinoma correlates with lymph node metastasis. *Cancer* 92(3), 556–568.

O'Donnell, R.K., Kupferman, M., Wei, S.J., Singhal, S., Weber, R., O'Malley, B., Cheng, Y., Putt, M., Feldman, M., Ziober, B. et al. (2005). Gene expression signature predicts lymphatic metastasis in squamous cell carcinoma of the oral cavity. *Oncogene* 24(7), 1244–1251.

Ogawa, Y., Nishioka, A., Hamada, N., Terashima, M., Inomata, T., Yoshida, S., Seguchi, H. and Kishimoto, S. (1998). Changes of mutant-type p53 expression in squamous cell carcinoma of the head and neck during radiation therapy and its clinical significance: comparison of an immunohistochemical method and PCR-SSCP assay. *Oncol Rep* 5(5), 1053–1059.

Remmerbach, T.W., Weidenbach, H., Pomjanski, N., Knops, K., Mathes, S., Hemprich, A. and Bocking, A. (2001). Cytologic and DNA-cytometric early diagnosis of oral cancer. *Anal Cell Pathol* 22(4), 211–221.

Ries, L.A.G., Harkins, D., Krapcho, M., Mariotto, A., Miller, B.A., Feuer, E.J., Clegg, L., Eisner, M.P., Horner, M.J., Howlader, N., et al. (eds). SEER Cancer Statistics Review, 1975-2003, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2003/, based on November 2005 SEER data submission, posted to the SEER web site, 2006.

Roepman, P., Wessels, L.F., Kettelarij, N., Kemmeren, P., Miles, A.J., Lijnzaad, P., Tilanus, M.G., Koole, R., Hordijk, G.J., van der Vliet, P.C. et al. (2005). An expression profile for diagnosis of lymph node metastases from primary head and neck squamous cell carcinomas. *Nat Genet* 37(2), 182–186.

Roesch-Ely, M., Nees, M., Karsai, S., Ruess, A., Bogumil, R., Warnken, U., Schnolzer, M., Dietz, A., Plinkert, P.K., Hofele, C. et al. (2007). Proteomic analysis reveals successive aberrations in protein expression from healthy mucosa to invasive head and neck cancer. *Oncogene* 26(1), 54–64.

Rosas, S.L., Koch, W., da Costa Carvalho, M.G., Wu, L., Califano, J., Westra, W., Jen, J. and Sidransky, D. (2001 Feb 1). Promoter hypermethylation patterns of P16, O6-methylguanine-DNA-methyltransferase, and death-associated protein kinase in tumors and saliva of head and neck cancer patients. *Cancer Res* 61(3), 939–942.

Sanchez-Cespedes, M., Esteller, M., Wu, L., Nawroz-Danish, H., Yoo, G.H., Koch, W.M., Jen, J., Herman, J.G. and Sidransky, D. (2000 Feb 15). Gene promoter hypermethylation in tumors and serum of head and neck cancer patients. *Cancer Res* 60(4), 892–895.

Schantz, S.P., Zhang, Z.F., Spitz, M.S., Sun, M. and Hsu, T.C. (1997). Genetic susceptibility to head and neck cancer: Interaction between nutrition and mutagen sensitivity. *Laryngoscope* 107(6), 765–781.

Scher, R.L. (2007). Role of nitric oxide in the development of distant metastasis from squamous cell carcinoma. *Laryngoscope* 117(2), 199–209.

Servomaa, K., Kiuru, A., Grenman, R., Pekkola-Heino, K., Pulkkinen, J.O.and Rytomaa, T. (1996). p53 mutations associated with increased sensitivity to ionizing radiation in human head and neck cancer cell lines. *Cell Prolif* 29(5), 219–230.

Sewell, D.A., Yuan, C.X. and Robertson, E. (2007). Proteomic signatures in laryngeal squamous cell carcinoma. *ORL J Otorhinolaryngol Relat Spec* 69(2), 77–84.

Spafford, M.F., Koch, W.M., Reed, A.L., Califano, J.A., Xu, L.H., Eisenberger, C.F., Yip, L., Leong, P.L., Wu, L., Liu, S.X. et al. (2001). Detection of head and neck squamous cell carcinoma among exfoliated oral mucosal cells by microsatellite analysis. *Clin Cancer Res* 7(3), 607–612.

Sudbo, J. and Reith, A. (2005). The evolution of predictive oncology and molecular-based therapy for oral cancer prevention. *Int J Cancer* 115(3), 339–345.

Teknos, T., Cox, C., Yoo, S., Chepeha, D., Wolf, G., Bradford, C., Carey, T. and Fisher, S. (2002). Elevated serum vascular endothelial growth factor and decreased survival in advanced laryngeal carcinoma. *Head Neck* 24(11), 1004–1011.

Toruner, G.A., Ulger, C., Alkan, M., Galante, A.T., Rinaggio, J., Wilk, R., Tian, B., Soteropoulos, P., Hameed, M.R., Schwalb, M.N. et al. (2004). Association between gene expression profile and tumor invasion in oral squamous cell carcinoma. *Cancer Genet Cytogenet* 154(1), 27–35.

Weber, A., Hengge, U.R., Stricker, I., Tischoff, I., Markwart, A., Anhalt, K., Dietz, A., Wittekind, C. and Tannapfel, A. (2007). Protein microarrays for the detection of biomarkers in head and neck squamous cell carcinomas. *Hum Pathol* 38(2), 228–238.

van Houten, V.M., Leemans, C.R., Kummer, J.A., Dijkstra, J., Kuik, D.J., van den Brekel, M.W., Snow, G.B. and Brakenhoff, R.H. (2004). Molecular diagnosis of surgical margins and local recurrence in head and neck cancer patients: A prospective study. *Clin Cancer Res* 10(11), 3614–3620.

Yuen, P.W., Man, M., Lam, K.Y. and Kwong, Y.L. (2002). Clinicopathological significance of p16 gene expression in the surgical treatment of head and neck squamous cell carcinomas. *J Clin Pathol* 1, 58–60.

Zhou, L., Cheng, L., Tao, L., Jia, X., Lu, Y. and Liao, P. (2006). Detection of hypopharyngeal squamous cell carcinoma using serum proteomics. *Acta Otolaryngol* 126(8), 853–860.

RECOMMENDED RESOURCES

Slaughter, D.P., Southwick, H.W. and Smejkal, W. (1953). Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. *Cancer* 6(5), 963–968.

This is a landmark article describing the concept of field cancerization and its implications in development of second primaries in the upper aerodigestive tract.

Thomas, G.R., Nadiminti, H. and Regalado, J. (2005). Molecular predictors of clinical outcome in patients with head and neck squamous cell carcinoma. *Int J Exp Pathol* 86(6), 347–363.

This presents a comprehensive review of molecular markers involved in carcinogenesis and progression of head and neck squamous cell carcinoma and their associations with clinical outcome.

Patmore, H.S., Cawkwell, L., Stafford, N.D. and Greenman, J. (2005). Unraveling the chromosomal aberrations of head and neck squamous cell carcinoma: a review. *Ann Surg Oncol* 12(10), 831–842.

This is an in-depth evaluation of specific chromosomal aberrations from data generated from comparative genomic hybridization analysis of head and neck squamous cell carcinoma.

CHAPTER

Genomic Evaluation of Brain Tumors and Gliomas

Sean E. Lawler and E. Antonio Chiocca

INTRODUCTION

Approximately 20,000 new patients with primary brain tumors are diagnosed in the United States each year (De Angelis, 2001). These constitute the most common solid tumors in children, and rank first among all cancer types in average years lost. They rarely metastasize outside the central nervous system; however, more than 100,000 patients per year die with symptomatic intracranial metastases due to systemic primary cancer.

Tumor type is currently determined by histopathologic analysis of biopsy samples and graded on a scale of I (benign) to IV (highly malignant) based on a range of histological tumor features (frequency of mitotic figures, necrosis, nuclear atypia and vascularity) according to the WHO classification of nervous system tumors (Kleihues and Cavenee, 2000; Louis et al., 2007). The various brain tumor types show similarities in clinical presentation, diagnosis and treatment. Most are treated by aggressive surgical resection when possible, followed by chemo and/or radiotherapy. Brain tumors are especially challenging, because they are often resistant to therapies, progress rapidly, and infiltrate normal brain tissue. Even a benign brain tumor may seriously compromise normal brain function, and surgical tumor excision must be carried out without damaging vital brain structures. Delivery of drugs to the central nervous system, and therapy induced neurotoxicity present further obstacles to effective treatment.

This article will focus on gliomas, the best studied and most common primary brain tumor type, accounting for approximately half of all cases. Gliomas are usually classified as either astrocytic or oligodendroglial (summarized in Table 41.1). Low-grade gliomas such as pilocytic astrocytomas are benign and have a good prognosis. Survival range for Grade II and III astrocytomas is broad, and these usually progress to higher-grade tumors. The most common and most aggressive astrocytic tumor is the Grade IV glioblastoma multiforme, among the deadliest of human cancers with a median survival of around 12 months, even with aggressive treatment. Oligodendrogliomas and the mixed oligoastrocytomas are less common than pure astrocytomas, and typically have longer survival times. Most low-grade oligodendrogliomas progress to a higher-grade tumor. With early diagnosis by MRI and chemotherapy, the mean survival is 16 years.

There are several well established molecular alterations commonly seen in gliomas, which are also observed in many other cancer types (reviewed in Louis, 2006; Schwartzbaum et al., 2006). Gliomas are typically characterized by increased tyrosine kinase receptor activation through activation of platelet-derived growth factor receptor (PDGFR) or epidermal growth factor receptor (EGFR) signaling. This is reflected in amplified levels of receptors and/or their ligands, or the presence of mutated constitutively active receptors such as EGFRVIII. This leads to activation of downstream intracellular signaling pathways, the

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

TABLE 41.1 Common subtypes of glioma and their basic characteristics

Tumor type	Characteristics	Peak incidence	Survival
Astrocytic			
Pilocytic Astrocyoma (I)	Slow growing, often cystic	Children	Infrequently fatal
Diffuse Astrocytoma (II)	Slow growing, invasive tendency to progress	Young adults	Mean 6–8 years highly variable
Anaplastic Astrocytoma (III)	tendency to progress	40–60 years	Mean TTP 2 years
Glioblastoma (IV)	*De novo*, or secondary mitotic figures, anaplasia, necrosis, vascularity	45–70 years	12 months
Oligodendroglial			
Oligodendroglioma (II) and Anaplastic oligodendroglioma (III)	Grow diffusely in cortex and white matter	50–60 years	3–15 years
Mixed			
Oligoastrocytoma (II) and anaplastic oligoastrocytoma (III)	Oligodendrocytic and astrocytic characteristics	35–50 years	4–7 years

PI3K/Akt pathway being considered as extremely important by promoting cell survival. High-grade gliomas are also associated with loss of the tumor suppressor phosphatase and tensin homolog (PTEN), which acts by further activating PI3K signaling. The cell cycle is deregulated through disruption of p53 (alterations in p53, HDM2, and $p14^{ARF}$) and RB (through alterations in p16 and RB) pathways. Based on age of onset and pathology, two different kinds of glioblastoma exist, with characteristic genetic alterations, although they are clinically similar (Kleihues and Cavenee, 2000). Glioblastoma in older patients shows no sign of previous low-grade tumors, and is known as either *de novo* or primary glioblastoma (95%). Typically these tumors show EGFR amplification, p16 deletions and PTEN deletions. Secondary glioblastoma occurs in younger patients (5%) and is due to progression from lower-grade astrocytoma. These are characterized by p53 mutation and amplification of PDGF signaling. In contrast to astrocytomas, oligodendrogliomas are characterized by allelic loss of 1p and 19q. However, tumor progression is associated with changes similar to those seen in astrocytomas (see Figure 41.1). Emerging data from the use of high-throughput, chip-based global molecular profiling techniques such as array-CGH and gene expression profiling suggest that these changes represent only a fraction of the alterations in gliomas, as described in the following sections.

PREDISPOSITION

Relatively few genetic and environmental factors that influence brain tumor development have been clearly identified (reviewed in Schwartzbaum et al., 2006), and only a small minority of tumors can be attributed to inherited predisposition. For example, a shared susceptibility to breast cancer, brain tumors, and Fanconi anemia was reported in four families with germline BRCA2 mutations (Offit et al., 2003). In addition, genetic predisposition is associated with various familial cancer syndromes (Turcot's, NF1, NF2, and Li-Fraumeni), accounting for 1–2% of all brain tumors. Inherited polymorphisms that may influence brain tumor formation involve oxidation, DNA repair, and immune function. For example, the DNA repair gene XRCC7 G7621T variant, leads to a 1.8-fold increased risk of glioblastoma (Wang et al., 2004) and the hTERT MNS16A allele results in a twofold increase in survival, due to higher expression levels of this allele (Wang et al., 2006). A consistent inverse association between self-reported allergic conditions and glioma has been observed. Asthma-associated polymorphisms in IL-4RA and IL-13 were inversely associated with glioblastoma incidence. This is a provocative observation because these factors have also been shown to inhibit glioma growth, and suggests a link between glioblastoma and immune function (Schwartzbaum et al., 2005).

Epidemiological studies have shown that glioblastoma has higher incidence in the white population and is more common in men than women (ratio 3:2), for unknown reasons. Many environmental risk factors have been investigated, including the use of cell phones, electromagnetic fields, viral infection, diet, tobacco, and previous head trauma/injury (Schwartzbaum et al., 2006). Gamma-radiation is the most widely accepted risk factor for gliomas. Indeed, genetic sensitivity to radiation or environmental carcinogens may play a role in brain tumor pathogenesis (Bondy et al., 2001). The limited findings on genetic and environmental factors involved in brain tumor development may reflect the complexity of genetic interactions with environmental factors. In addition, there is a need for studies to be carried out

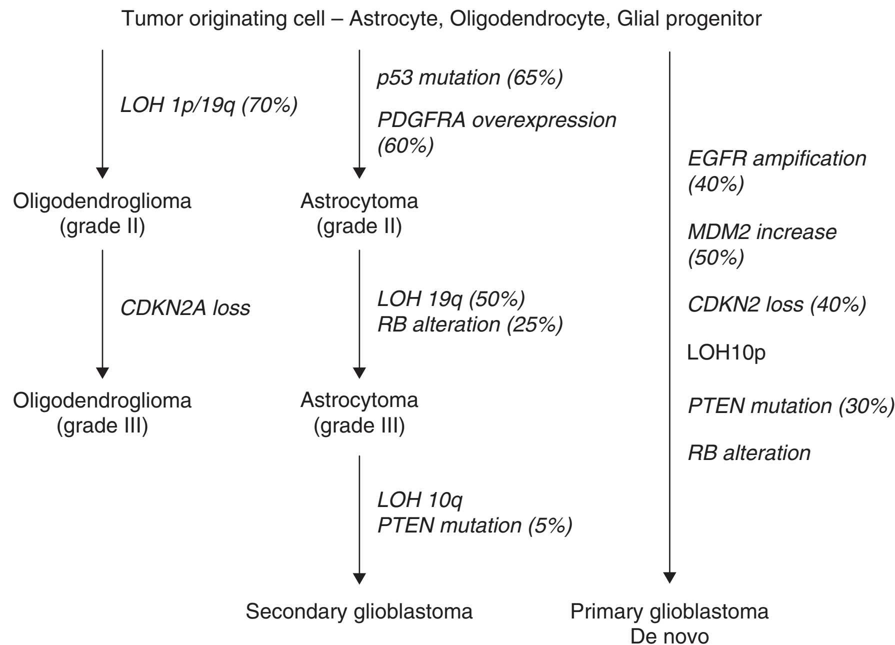


Figure 41.1 Pathways to glioma formation. Three pathways of genetic alteration are currently known that define Oligodendroglioma, primary glioblastoma and secondary glioblastoma. The major genetic alterations associated with tumor progression are shown, with their approximate frequency.

in large sample groups to confirm many preliminary findings, that have been reported in small studies. A brain tumor epidemiology consortium has recently been established in order to further understanding in this area (Schwartzbaum et al., 2006).

DIAGNOSIS AND PROGNOSIS

Brain tumor diagnosis is currently based largely on histologic analysis of tumor biopsies, according to WHO guidelines. Major prognostic factors in gliomas are tumor type and grade, patient age, symptom duration, degree of surgery and neurological deficit. The few long-term survivors of glioblastoma are young, in otherwise good health, and able to undergo gross total resection followed by chemo and radiotherapy. The current WHO tumor classification system suffers from well recognized problems due to interobserver variability, tumor heterogeneity and ambiguous tumor types (Louis et al., 2001). Very few further molecular markers are in common use at present, although codeletion of 1p and 19q is useful in predicting clinical responsiveness in oligodendroglioma, as described below. The incorporation of molecular information would allow less subjective diagnosis, and more precise prognostic information.

Chromosomal Alterations in Brain Tumors

A range of characteristic genetic changes have been observed, which in high-grade gliomas are variable and complex (reviewed by Bayani et al., 2005). Benign tumor types are characterized by fewer changes, the most prominent being alteration of chromosome 22. Cytogenetic analysis often reveals a normal karyotype for grade I astrocytomas, whereas grade II gliomas often have mutations in the *TP53* gene and overexpression of *PDGFRA*. The most commonly observed genetic alteration in glioblastoma is loss of chromosome 10 (60–80%), resulting in reduced *PTEN* expression, although it is possible that other genes on chromosome 10 are important. Gain of chromosome 7 (EGFR), and loss of 9p (p16 and $P14^{ARF}$), are also consistently observed (shown in Figure 41.1). Other consistently observed chromosomal alterations are shown in Table 41.2.

Recent, detailed microarray-based high-resolution genomic mapping studies have confirmed the presence of the commonly observed genetic changes in brain tumors and are revealing many more alterations whose significance is not yet understood, which may also eventually become useful diagnostic tools. Kotliarov et al. (2006) reported the use of Affymetrix high density 100 K SNP arrays to analyze genomic alterations in 178 glioma samples at an unprecedented resolution of 25 Kb. Array-CGH analysis of glioblastoma has uncovered numerous novel changes and can easily distinguish primary from secondary glioblastoma, and also showed that secondary glioblastoma falls into two distinct groups (Maher et al., 2006). Each tumor group had many more unique events than shared and therefore may benefit from quite different therapeutic approaches. cDNA microarray-based CGH was able to predict tumor type successfully based on common genetic alterations (Bredel et al., 2005). All these studies revealed large numbers of previously unidentified alterations. Other studies have reported deletions in chromosomes 6, 21, and 22 (Lassman et al., 2005). Two regions in 6q26 were found to be commonly

Table 41.2 Common genetic alterations seen in glial tumors

Region	Alterations	Candidate glioma genes
1p (34-pter) (various)	Gains and deletions	Unknown
1q32	Gains	*RIPK5, MDM4, PIK3C2B*
4q	Deletions	*NEK1, NIMA*
7p11.2 – p12	Amplification/gain	*EGFR*
9p21 – p24	Deletions	*CDKN2*
10q23	Deletions	*PTEN*
10q25 – q26	Deletions	*MGMT*
12q13.3 – q15	Amplification	*MDM2, CDK4*
13p11 – p13	Loss	*RB*
19q13	Loss	*GLTSCR1, GLTSCR2, LIG1, PSCD2*
22q11.2 – q12.2	Loss	28 genes including *INI1*
22q13.1 – 13.3	Loss	Not known

Adapted from Schwartzbaum et al. (2006).

deleted, containing three novel genes, which may have tumor suppressor functions (Ichimura et al., 2006). Furthermore, the NIH Cancer Genome Atlas Project (CGA) promises to carry out large-scale genomic sequencing to fully understand genetic alterations in cancer. Glioblastoma is one of the initial three cancer types that will be analyzed in this project (http://cancergenome.nih.gov). This analysis promises to revolutionize the level of understanding of this cancer.

The most commonly used molecular diagnostic tool based on chromosomal alterations at present is in oligodendroglioma. Oligodendrogliomas frequently show a remarkable sensitivity to PCV chemotherapy, (procarbazine, CCNU, and vincristine), with a response rate of approximately 75%. Thus, within a histologically indistinguishable entity, there exist subgroups which show different biological behavior. Molecular genetic studies revealed that LOH of chromosome 1p which is found in 60–80% of oligodendrogliomas, and often accompanied by LOH of 19q (Cairncross et al., 1998), was the underlying reason for the increase in sensitivity. This occurs in both high and low-grade oligodendrogliomas and also applies to radiotherapy and temozolomide (Hoang-Xuan et al., 2004), the current chemotherapeutic drug of choice in glioma. Combined loss of 1p and 19q are strong correlates of longer survival in oligodendroglioma (Cairncross et al., 1998) and has been associated with increased survival in glioblastoma patients (Ino et al., 2000). Therefore genetic analysis for 1p and 19q is now commonly performed for oligodendroglioma and can be used to determine the most suitable therapeutic strategy. The prognostic significance of 1p/19q was confirmed in large recent trials (Van den Bent et al., 2006).

Identification of the key genes involved at 1p and 19q may lead to novel treatment strategies broadly applicable to gliomas in general. It has recently been shown that 1p/19q co-deletion is a result of an imbalanced translocation. This finding may lead to a better understanding of the functional importance of 1p/19q co-deletion (Jenkins et al., 2006). Studies also show that oligodendrogliomas and oligoastrocytomas may be further subgrouped on the basis of distinct genetic changes in addition to 1p and 19q alterations. CGH revealed a hemizygous deletion in 500 kb region in 11q13 and 300 kb region in 13q12 in virtually all low-grade oligodendrogliomas, regardless of 1p/19q status (Rossi et al., 2005), suggesting that further markers exist.

Other molecular diagnostic tools include analysis of EGFR amplification, seen commonly in gliomas. Glioblastomas with amplified EGFR often occur in older patients and can be difficult to distinguish from anaplastic oligodendroglioma due to their small cell appearance. Assessment of EGFR along with 1p and 19q can therefore be helpful in this situation (Burger et al., 2001). In summary, at present, to distinguish between different glioma groups, the following changes are the most typical and some are used diagnostically. Loss of chromosomes 1p/19q is typical of oligodendrogliomas, whereas gains of chromosome 7 in the setting of intact 1p/19q are more typical of astrocytomas. The detection of amplified EGFR favors the diagnosis of high-grade astrocytomas over anaplastic oligodendroglioma, which is especially relevant for small cell astrocytomas.

Transcriptional Alterations in Brain Tumors

Global transcriptional profiling using microarrays has revealed many novel changes in brain tumors (reviewed by Mischel et al., 2004). This initial burst of studies has revealed many alterations of potential utility. However, the vast majority of this data has not been replicated in independent patient cohorts. Expression profiles can readily distinguish different histologically classified tumor types. For example, microarray data could separate glioblastoma from oligodendroglioma on the basis of a 170 (Shai et al., 2003) or a 70 gene signature (Nutt et al., 2003). EGFR amplified glioblastoma was distinguished from non-amplified EGFR tumors using a 90 gene signature (Mischel et al., 2003), grade II and grade III oligodendrogliomas by a 200 gene subset (Watson et al., 2001) and also oligodendrogliomas with and without 1p loss (Mukasa et al., 2002). Gene expression analysis has also identified tumor subgroups that are indistinguishable histologically. For example, clustering analysis separated oligodendrogliomas into two survival groups based on gene expression (Huang et al., 2004). Two microarray studies of glioblastoma identified gene signatures corresponding to different survival groups, and suggest that increased expression of genes that promote infiltration can lead to a poor prognosis (Liang et al., 2005, Rich et al., 2005). Furthermore, microarrays were better

predictors of survival than histological grading for oligodendrogliomas (Nutt et al., 2003). Gene expression profiling therefore may provide useful diagnostic and prognostic information.

In a comprehensive study, gene expression profiles were obtained from 76 resected astrocytomas with known survival data (Phillips et al., 2006). Tumors were separated into either short or relatively long survival groups, and cluster analysis then segregated tumors into 3 distinct discrete sample sets based on 35 signature genes. These were defined by the predominant class of gene expressed, as either proneural (*PN*), proliferative (*Prolif*), or mesenchymal (*Mes*). The *PN* subset was associated with substantially longer survival, was more prominent in younger patients, and showed activated Notch signaling as seen by elevated delta-like protein 3 (DLL3) expression. Poor prognosis was associated with upregulation of proliferative markers such as proliferating cell nuclear antigen (PCNA), topoisomerase 2a (TOP2A), and angiogenesis markers such as vascular endothelial growth factor (VEGF) and its receptors, and platelet endothelial cell adhesion molecule 1 (PECAM1). Markers of neural stem cells were also associated with poor prognosis. All WHO classified grade II tumors were *PN*, whereas for grade IV tumors 31% were *PN*, 20% were *Prolif* and 49% were *Mes*, with expression profiles shifting toward the *Mes* class after recurrence. This suggests that the identified subtypes represent stages of tumorigenesis rather than different tumors. This study also correlated gene expression profiles with genetic abnormalities demonstrating that losses of chromosome 10 and gains on chromosome 7 are associated with *Prolif* and *Mes* phenotypes (around 80%), compared with 20% for the *PN* class. The authors therefore propose a model in which *PN* phenotype progresses to a *Prolif* phenotype, and ultimately the *Mes* phenotype, with the poorest prognosis. This kind of study demonstrates that useful information could be obtained from a small subset of signature genes. A similar study identified a 44 gene expression signature to classify gliomas into previously unrecognized biological and prognostic groups which outperformed histology-based classification in survival prediction (Freije et al., 2004). In agreement with Phillips et al. key genes identified include DLL3 (good prognosis), TOP2A (poor) leukemia inhibitory factor (LIF) (poor), S100A4 (poor), and VEGF (poor).

The most common alterations seen in gene expression profiling studies of gliomas are in genes involved in immune system regulation, hypoxia, cell proliferation, angiogenesis, neurogenesis, and cell motility. Ten of the most commonly reported upregulated genes in glioblastoma are summarized in Table 41.3. Many of these changes were not predicted by analysis of chromosomal alterations. Of these highly upregulated genes, the transmembrane glycoprotein GPNMB (Kuan et al., 2006) and YKL40 (also known as chitinase 3 like-1 (CHI3L1)) (Hormigo et al., 2006), have been proposed as prognostic markers. YKL-40 has emerged from gene screening studies as one of the most robust and consistently observed markers in glioblastoma (Freije et al., 2004; Phillips et al., 2006). YKL-40 mRNA was on average 82-fold higher in glioblastoma than anaplastic oligodendrogliomas (Nutt et al., 2003). A further study showed that that YKL-40 may be a better marker than glial fibrillary acidic protein (GFAP) which is currently the standard in distinguishing diagnostically challenging gliomas, and suggests that combining YKL-40 and GFAP staining is the best approach at present (Nutt et al., 2005). YKL-40 expression is also associated with radioresistance (Pelloski et al., 2005). Serial analysis of gene expression (SAGE) analysis of glioma samples revealed that it is the most upregulated gene compared with normal brain (Boon et al., 2004). Taken together the data described shows that gene expression profiling may predict tumor type, patient outcome, and is providing the data to identify new therapeutic targets (Horvath et al., 2006). However, there are a large number of potential markers, and it is not yet clear which of these will finally be used clinically.

A potential drawback of transcriptional profiling is that mRNA levels in a given sample may not be representative of the whole tumor, and also that they do not provide an accurate indication of the translation into protein. mRNA must be bound to polysomes in order to be translated, and a study of polysomal RNA in glioma cell lines revealed marked differences to global RNA in the same sample (Rajasekhar et al., 2003). In addition, the recently discovered non-protein coding RNA

TABLE 41.3 The "top ten" overexpressed genes in glioma as judged by their reported appearance in microarray studies cited in this chapter

Gene	Symbol(s)	Function
Fibronectin 1	*FN1*	ECM – angiogenesis, invasion
Insulin like growth factor binding protein 2	*IGFBP2*	Promotes invasion
Collagens (IV and VI)	*COL6A/ COL4A*	ECM, promotes invasion
Topoisomerase 2A	*TOP2A*	DNA replication and transcription
Biglycan	*BGN*	ECM proteoglycan. Role in glioma unknown
Chitinase 3-like-1	*CHI3LI*	ECM, invasion, angiogenesis, survival
Vascular endothelial growth factor	*VEGF*	Angiogenesis, invasion
Vimentin	*VIM*	Intermediate filament, invasion
Epidermal growth factor receptor	*EGFR*	Survival, growth, invasion
TGF-β1	*TGFB1*	Proliferation, differentiation, invasion

molecules such as microRNAs can have a profound impact on gene expression by either causing degradation of specific transcripts, or by preventing translation of the mRNA. MicroRNA expression is altered in many cancers (Volinia et al., 2006), including brain tumors, with mir-21 being particularly over-represented in glioblastoma (Chan et al., 2005; Ciafre et al., 2005). This also has functional importance because knockdown of mir-21 leads to apoptosis in glioma cells (Chan et al., 2005).

Proteomic Studies in Brain Tumors

Proteomic studies on brain tumors typically have used the 2-dimensional gel electrophoresis and mass spectrometry approach. Comparison of glioblastoma with non-tumor tissue identified 11 upregulated and 4 down-regulated proteins including the fatty acid binding protein FABP7 – also identified as a prognostic marker in microarrays (Hiratsuka et al., 2003, Liang et al., 2005). The study closely examined SIRT2, a tubulin deacetylase, downregulated in 12 out of 17 gliomas, and suggested it may act as a tumor suppressor. Another study used 2d-gel-based proteomics to molecularly classify and predict survival in 85 samples including various tumors and normal brain. Expression patterns were identified that could distinguish tumor types and predict survival (Iwadate et al., 2004). Proteomics may find utility in the identification of serum biomarkers that could lead to much simpler diagnosis and monitoring. Few studies have yet been performed in this area, although cathepsin D was found to be elevated in the serum of glioblastoma patients, and associated with poor prognosis (Fukuda et al., 2005). Matrix metalloprotease-9 (MMP-9) and YKL-40 have also been detected in the serum of glioblastoma patients (Hormigo et al., 2006). However, like other studies in this area, independent analysis on further patient groups is not yet available.

Direct mass spectrometry of tissue samples is emerging as a powerful tool and may become increasingly influential in the diagnostic arena (Caprioli, 2005). Analysis of 162 patient biopsies from glioma patients using this approach, known as direct tissue matrix-assisted laser desorption ionization mass spectroscopy (MALDI-MS), was used to generate a novel classification scheme based on protein expression profiles, which segregated long-term and short-term survivors (Schwartz et al., 2005). The authors were able to distinguish normal brain from glioma biopsies and also to distinguish various grades of tumor by protein expression in these samples. MALDI-MS technology is also amenable to high-throughput tissue screening in a clinical setting. Significant developments can be expected in this area in the near future.

PHARMACOGENOMICS

At present, many brain tumors are treated in a similar way regardless of their classification, but this situation is changing as patient responsiveness to various therapies is correlated with genetic changes in the tumor. For example patients with combined 1p and 19q loss may opt for chemotherapy but withhold radiation until recurrence, therefore avoiding the problematic effects of extensive radiation to the brain in long surviving patients. On the other hand, a patient with intact chromosome 1p and EGFR amplification may opt to forego PCV chemotherapy, and have radiation and a novel therapy, thereby avoiding the myelotoxic effects of PCV (Louis, 2006).

The benefit of chemotherapy in glioblastoma and anaplastic astrocytoma is small, with response plus stabilization in 20–50% of patients, because these tumors have intrinsic chemoresistance. Esteller et al. (2000) first described the mechanism by which some gliomas are resistant to nitrosourea alkylating agents such as carmustine (BCNU). These agents kill by alkylation of the O6 position of guanine, thereby crosslinking adjacent DNA strands. These crosslinks can be repaired by the DNA repair enzyme O6-methylguanine-DNA methyltransferase (MGMT), which rapidly reverses alkylation. Around 30% of gliomas lack MGMT, and this correlates closely with chemosensitivity. However, mutations in MGMT are rare, and it was found that methylation of the promoter region of the gene, leading to transcriptional silencing, accounts for the variation. 12 of 19 high-grade glioma patients with methylated MGMT promoters had a partial or complete response to treatment, whereas only one of 28 patients with an unmethylated promoter had a response. MGMT promoter methylation also is linked to responsiveness to temozolomide (Hegi et al., 2005).

A second example of pharmacogenomics is in the use of EGFR kinase inhibitors for glioma treatment. Molecularly targeted therapeutics such as EGFR inhibitors are an attractive option for glioma because of the frequently observed amplification of EGFR in these tumors. However, trials so far have shown only a 20% response rate (Mellinghoff et al., 2005). Molecular determinants of responsiveness were analyzed, and it was found that combined EGFRVIII mutation and the presence of PTEN sensitizes tumor, suggesting that secondarily targeting PI3K signaling may improve this kind of therapy, and that patients with the appropriate profile should be enlisted in further trials for these drugs. As glioma treatment becomes increasingly sophisticated further details and examples of chemoresistance mechanisms, which can affect therapeutic decision making are sure to emerge.

SUMMARY

The prognosis for many brain tumors remains dismal. However, the wealth of data currently being generated by microarray-based studies in particular should have a profound effect on brain tumor management. Headway is currently being made in four main areas:

1. *Identification of robust molecular markers and signatures that define tumor types*: Recent studies have revealed that molecular signatures readily distinguish tumor types, and can also further subdivide histologically similar tumors into previously unknown groups. This opens up the possibility that molecular information will be increasingly useful in brain tumor diagnosis.
2. *Correlation of molecular signatures with patient prognosis*: Molecular information may provide the ability to predict

patient survival more reliably than at present. Recent studies show that molecular classification can improve prediction of patient outcome compared with histology in some cases.

3. *Correlation of molecular data with therapeutic response*: Molecular alterations play a role in drug responsiveness. The identification of these alterations is beginning to impact clinical decision making; 1p/19q is now widely used, and MGMT and EGFR status may be used in the near future. These represent the first steps towards the development of patient-tailored therapies.
4. *Identification of novel therapeutic targets*: Improved understanding of brain tumor biology is allowing the development of new therapeutic strategies based on novel molecular targets revealed by global genetic analyses.

The most promising approaches so far may involve tumor classification by gene expression profiling using a panel of relevant genes. The identification of the *PN*, *Mes* and *Prolif* tumor subgroups (Phillips et al., 2006) provides a strong foundation for such an approach. Microarray studies have identified many candi dates that could be of use in prognosis and monitoring – the most promising of these so far may be YKL-40, which can differentiate astrocytic from oligodendroglial tumors and is associated with poor prognosis. The association between MGMT promoter methylation and improved treatment response in glioblastoma patients has provoked much interest in this area. The incorporation of MGMT expression analysis may therefore soon emerge as a useful predictive clinical tool.

The list of molecular alterations is growing, but is still far from complete. In addition the role of epigenetic changes has barely been studied as yet in gliomas. The major challenge at this point is to translate this increasingly large amount of data into clinically useful tools. This should lead to the development of a new genomic approach to glioma treatment, using molecular signatures to provide accurate diagnosis, and to stratify patients for the most effective therapy, or for targeted therapies. This should finally lead to a better outlook for these cancers.

2009 UPDATE

The most significant recent development in brain tumor genomics is the development of TCGA (Cancer Genome Atlas Research Network, 2008). This freely available resource contains DNA copy number, gene expression, microRNA, and methylation data on 206 glioblastomas so far. Some 600 candidate genes are also being sequenced. In another study, next-generation sequencing as well as profiling was performed on 20,661 genes in 22 glioblastomas (Parsons et al., 2008). These studies identified mutations and alterations in core pathways and mark the beginning of a new era in glioblastoma research and discovery.

Little is known about genetic factors predisposing to glioma. Previous data established a potential inverse relationship between allergy and glioma incidence. This is supported by the demonstration that polymorphisms in the IL4-receptor gene are significantly associated with improved survival (Scheurer et al., 2008a). Effects of combined SNPs in multiple DNA repair genes increases glioma risk (Liu et al., 2008). Evidence is beginning to accumulate suggesting a role of CMV in glioma progression and awaits further investigation (Cobbs et al., 2008; Prins et al., 2008; Scheurer et al., 2008b).

Several groups have reported that high levels of the stem cell marker CD133 are associated with poor prognosis in gliomas (Beier et al., 2008; Rebetz et al., 2008; Zeppernick et al., 2008). Resistance to temozolomide chemotherapy also appears to be linked to a stem cell signature (Murat et al., 2008). Co-deletion of chromosomes 1p and 19q is the major molecular marker currently in wide use. MGMT is still under close study, but some difficulties have been reported in relating immunostaining to patient response (Preusser et al., 2008).

Efforts to classify gliomas based on transcriptional profiling have continued (Petalidis et al., 2008). As described in earlier studies, this data reveals three distinct molecular classes of glioma and suggests that grading based on molecular features can outperform traditional histologic methods. It is increasingly recognized that personalized therapies may be necessary for the effective treatment of glioblastoma. An example of this was outlined in a study describing an oncogenic role for STAT3 by association with EGFRVIII, and a tumor suppressor role for STAT3 in response to PTEN deletion (de la Iglesia, 2008).

An increasing number of microRNAs have been characterized in glioblastoma. MiR-21 antagonism and re-expression of weakly expressed microRNAs, such as miR-7, miR124, and miR-128 all have effects on glioma cell proliferation and/or invasion, and target important genes, such as EGFR (miR-7) (Kefas et al., 2008), CDK6 (miR-124) (Silber et al., 2008), and Bmi1 (miR-128) (Godlewski et al., 2008). Also, miR-296 was identified in endothelial cells as a pro-angiogenic microRNA (Wurdinger et al., 2008).

Anti-angiogenic approaches show promise in early trials. These include the VEGFR inhibitor Cediranib (AZD2171) (Batchelor et al., 2007) and the integrin inhibitor cilengitide (Reardon et al., 2008). The anti-VEGF monoclonal antibody Avastin showed excellent radiographic responses in trials for recurrent glioblastoma in combination with CPT11 (Vredenburgh et al., 2007). It is not yet clear whether these impressive responses will translate into improved survival (Norden et al., 2008). Avastin will now be investigated in newly diagnosed glioblastoma in combination with temozolomide and radiation.

2009 UPDATE REFERENCES

Batchelor, T.T., Sorensen, A.G., di Tomaso, E., Zhang, W.T., Duda, D.G., Cohen, K.S. et al. (2007). AZD2171, a pan-VEGF receptor tyrosine kinase inhibitor, normalizes tumor vasculature and alleviates edema in glioblastoma patients. *Cancer Cell* 11, 83-95.

Beier, D., Wischhusen, J., Dietmaier, W., Hau, P., Proescholdt, M., Brawanski, A., Bogdahn, U. and Beier, C.P. (2008). CD133 expression and cancer stem cells predict prognosis in high-grade oligodendroglial tumors. *Brain Pathol* 18, 370-377.

Cancer Genome Atlas Research Network (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455, 1061-1068.

Cobbs, C.S., Soroceanu, L., Denham, S., Zhang, W. and Kraus, M.H. (2008). Modulation of oncogenic phenotype in human glioma cells by cytomegalovirus IE1-mediated mitogenicity. *Cancer Res* 68, 724-730.

Godlewski, J., Nowicki, M.O., Bronisz, A., Williams, S., Otsuki, A., Nuovo, G. et al. (2008). Targeting of the Bmi-1 oncogene/stem cell renewal factor by microRNA-128 inhibits glioma proliferation and self-renewal. *Cancer Res* 68, 9125-9130.

Kefas, B., Godlewski, J., Comeau, L., Li, Y., Abounader, R., Hawkinson, M. et al. (2008). microRNA-7 inhibits the epidermal growth factor receptor and the Akt pathway and is down-regulated in glioblastoma. *Cancer Res* 68, 3566-3572.

Liu, Y., Zhou, K., Zhang, H., Shugart, Y.Y., Chen, L., Xu, Z., Zhong, Y., Liu, H., Jin, L., Wei, Q. et al. (2008). Polymorphisms of LIG4 and XRCC4 involved in the NHEJ pathway interact to modify risk of glioma. *Hum Mutat* 29, 381-389.

Murat, A., Migliavacca, E., Gorlia, T., Lambiv, W.L., Shay, T., Hamou, M.F. et al. (2008). Stem cell-related "self-renewal" signature and high epidermal growth factor receptor expression associated with resistance to concomitant chemoradiotherapy in glioblastoma. *J Clin Oncol* 26, 3015-3024.

Norden, A.D., Drappatz, J., Muzikansky, A., David, K., Gerard, M., McNamara, M.B. et al. An exploratory survival analysis of anti-angiogenic therapy for recurrent malignant glioma. *J Neurooncol* 92(2), 149-155.

Parsons, D.W., Jones, S., Zhang, X., Lin, J.C., Leary, R.J., Angenendt, P. et al. (2008). An integrated genomic analysis of human glioblastoma multiforme. *Science* 321, 1807-1812.

Petalidis, L.P., Oulas, A., Backlund, M., Wayland, M.T., Liu, L., Plant, K. et al. (2008). Improved grading and survival prediction of human astrocytic brain tumors by artificial neural network analysis of gene expression microarray data. *Mol Cancer Ther* 7, 1013-1024.

Preusser, M., Charles, J.R., Felsberg, J., Reifenberger, G., Hamou, M.F., Diserens, A.C. et al. (2008). Anti-O6-methylguanine-methyltransferase (MGMT) immunohistochemistry in glioblastoma multiforme: Observer variability and lack of association with patient survival impede its use as clinical biomarker. *Brain Pathol* 18, 520-532.

Prins, R.M., Cloughesy, T.F. and Liau, L.M. (2008). Cytomegalovirus immunity after vaccination with autologous glioblastoma lysate. *N Engl J Med* 359, 539-541.

Reardon, D.A., Nabors, L.B., Stupp, R. and Mikkelsen, T. (2008). Cilengitide: An integrin-targeting arginine-glycine-aspartic acid peptide with promising activity for glioblastoma multiforme. Expert Opin Investig Drugs 17, 1225-1235.

Rebetz, J., Tian, D., Persson, A., Widegren, B., Salford, L.G., Englund, E. et al. (2008). Glial progenitor-like phenotype in low-grade glioma and enhanced CD133-expression and neuronal lineage differentiation potential in high-grade glioma. *PLoS ONE* 3, e1936.

Scheurer, M.E., Amirian, E., Cao, Y., Gilbert, M.R., Aldape, K.D., Kornguth, D.G. et al. (2008a). Polymorphisms in the interleukin-4 receptor gene are associated with better survival in patients with glioblastoma. *Clin Cancer Res* 14, 6640-6646.

Scheurer, M.E., Bondy, M.L., Aldape, K.D., Albrecht, T. and El-Zein, R. (2008b). Detection of human cytomegalovirus in different histological types of gliomas. *Acta Neuropathol* 116, 79-86.

Silber, J., Lim, D.A., Petritsch, C., Persson, A.I., Maunakea, A.K., Yu, M. et al. (2008). miR-124 and miR-137 inhibit proliferation of glioblastoma multiforme cells and induce differentiation of brain tumor stem cells. *BMC Med* 6, 14.

Vredenburgh, J.J., Desjardins, A., Herndon, J.E., 2nd, Marcello, J., Reardon, D.A., Quinn, J.A. et al. (2007). Bevacizumab plus irinotecan in recurrent glioblastoma multiforme. *J Clin Oncol* 25, 4722-4729.

Würdinger, T., Tannous, B.A., Saydam, O., Skog, J., Grau, S., Soutschek, J., et al. (2008). miR-296 regulates growth factor receptor overexpression in angiogenic endothelial cells. *Cancer Cell* 14, 382-393.

Zeppernick, F., Ahmadi, R., Campos, B., Dictus, C., Helmke, B.M., Becker, N. et al. (2008). Stem cell marker CD133 affects clinical outcome in glioma patients. *Clin Cancer Res* 14, 123-129.

REFERENCES

Batchelor, T.T., Sorensen, A.G., di Tomaso, E., Zhang, W.T., Duda, D.G., Cohen, K.S. et al. (2007). AZD2171, a pan-VEGF receptor tyrosine kinase inhibitor, normalizes tumor vasculature and alleviates edema in glioblastoma patients. *Cancer Cell* 11, 83–95.

Bayani, J., Pandita, A. and Squire, J.A. (2005). Molecular cytogenetic analysis in the study of brain tumors: Findings and applications. *Neurosurg Focus* 19, E1.

Bondy, M.L., Wang, L.E., El-Zein, R., de Andrade, M., Selvan, M.S., Bruner, J.M., Levin, V.A., Alfred Yung, W.K., Adatto, P. and Wei, Q (2001). Gamma-radiation sensitivity and risk of glioma. *J Natl Cancer Inst* 93, 1553–1557.

Boon, K., Edwards, J.B., Eberhart, C.G. and Riggins, G.J. (2004). Identification of astrocytoma associated genes including cell surface markers. *BMC Cancer* 4, 39–47.

Bredel, M., Bredel, C., Juric, D., Harsh, G.R., Vogel, H., Recht, L.D. and Sikic, B.I. (2005). Functional network analysis reveals extended gliomagenesis pathway maps and three novel MYC-interacting genes in human gliomas. *Cancer Res* 65, 8679–8689.

Burger, P.C., Pearl, D.K., Aldape, K., Yates, A.J., Scheithauer, B.W., Passe, S.M., Jenkins, R.B. and James, C.D. (2001). Small cell architecture – a histological equivalent of EGFR amplification in glioblastoma multiforme? *J Neuropathol Exp Neurol* 60, 1099–1104.

Cairncross, J.G., Ueki, K., Zlatescu, M.C., Lisle, D.K., Finkelstein, D.M., Hammond, R.R., Silver, J.S., Stark, P.C., Macdonald, D.R., Ino, Y. et al. (1998). Specific genetic predictors of chemotherapeutic response and survival in patients with anaplastic oligodendrogliomas. *J Natl Cancer Inst* 90, 1473–1479.

Caprioli, R.M. (2005). Deciphering protein molecular signatures in cancer tissues to aid in diagnosis, prognosis, and therapy. *Cancer Res* 65, 10642–10645.

Chan, J.A., Krichevsky, A.M. and Kosik, K.S. (2005). MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. *Cancer Res* 65, 6029–6033.

Ciafre, S.A., Galardi, S., Mangiola, A., Ferracin, M., Liu, C.G., Sabatino, G., Negrini, M., Maira, G., Croce, C.M. and Farace, M.G. (2005). Extensive modulation of a set of microRNAs in primary glioblastoma. *Biochem Biophys Res Commun* 334, 1351–1358.

De Angelis, L.M. (2001). Brain tumors. *N Engl J Med* 344, 114–123.

De la Iglesia, N., Konopka, G., Puram, S.V., Chan, J.A., Bachoo, R.M., You, M.J. et al. (2008). Identification of a PTEN-regulated STAT3 brain tumor suppressor pathway. *Genes Dev* 22, 449–462.

Esteller, M., Garcia-Foncillas, J., Andion, E., Goodman, S.N., Hidalgo, O.F., Vanaclocha, V., Baylin, S.B. and Herman, J.G. (2000). Inactivation of the DNA-repair gene MGMT and the clinical response of gliomas to alkylating agents. *N Engl J Med* 343, 1350–1354.

Freije, W.A., Castro-Vargas, F.E., Fang, Z., Horvath, S., Cloughesy, T., Liau, L.M., Mischel, P.S. and Nelson, S.F. (2004). Gene expression profiling of gliomas strongly predicts survival. *Cancer Res* 64, 6503–6510.

Fukuda, M.E., Iwadate, Y., Machida, T., Hiwasa, T., Nimura, Y., Nagai, Y., Takiguchi, M., Tanzawa, H., Yamaura, A. and Seki, N. (2005). Cathepsin D is a potential serum marker for poor prognosis in glioma patients. *Cancer Res* 65, 5190–5194.

Hegi, M.E., Diserens, A.C., Gorlia, T., Hamou, M.F., de Tribolet, N., Weller, M., Kros, J.M., Hainfellner, J.A., Mason, W., Mariani, L. et al. (2005). MGMT gene silencing and benefit from temozolomide in glioblastoma. *N Engl J Med* 352, 997–1003.

Hiratsuka, M., Inoue, T., Toda, T., Kimura, N., Shirayoshi, Y., Kamitani, H., Watanabe, T., Ohama, E., Tahimic, C.G., Kurimasa, A. et al. (2003). Proteomics-based identification of differentially expressed genes in human gliomas: Down-regulation of SIRT2 gene. *Biochem Biophys Res Commun* 309, 558–566.

Hoang-Xuan, K., Capelle, L., Kujas, M., Taillibert, S., Duffau, H., Lejeune, J., Polivka, M., Criniere, E., Marie, Y., Mokhtari, K. et al. (2004). Temozolomide as initial treatment for adults with low-grade oligodendrogliomas or oligoastrocytomas and correlation with chromosome 1p deletions. *J Clin Oncol* 22, 133–138.

Hormigo, A., Gu, B., Karimi, S., Riedel, E., Panageas, K.S., Edgar, M.A., Tanwar, M.K., Rao, J.S., Fleisher, M., DeAngelis, L.M. et al. (2006). YKL-40 and matrix metalloprotease-9 as potential serum biomarkers for patients with high-grade gliomas. *Clin Cancer Res* 12, 5698–5704.

Horvath, S., Zhang, B., Carlson, M., Lu, K.V., Zhu, S., Felciano, R.M., Laurance, M.F., Zhao, W., Qi, S., Chen, Z. et al. (2006). Analysis of oncogenic signling networks in glioblastoma identifies ASPM as a molecular target. *Proc Natl Acad Sci U S A* 103, 17402–17407.

Huang, H., Okamoto, Y., Yokoo, H., Heppner, F.L., Vital, A., Fevre-Montange, M., Jouvet, A., Yonekawa, Y., Lazaridis, E.N., Kleihues, P. et al. (2004). Gene expression profiling and subgroup identification of oligodendrogliomas. *Oncogene* 23, 6012–6022.

Ichimura, K., Mungall, A.J., Fiegler, H., Pearson, D.M., Dunham, I., Carter, N.P. and Collins, V.P. (2006). Small regions of overlapping deletions on 6q26 in human astrocytic tumours identified using chromosome 6 tile path array-CGH. *Oncogene* 25, 1261–1271.

Ino, Y., Zlatescu, M.C., Sasaki, H., Macdonald, D.R., Stemmer-Rachamimov, A.O., Jhung, S., Ramsay, D.A., von Deimling, A., Louis, D.N. and Cairncross, J.G. (2000). Long survival and therapeutic responses in patients with histologically disparate high-grade gliomas demonstrating chromosome 1p loss. *J Neurosurg* 92, 983–990.

Iwadate, Y., Sakaida, T., Hiwasa, T., Nagai, Y., Ishikura, H., Takiguchi, M. and Yamaura, A. (2004). Molecular classification and survival prediction in human gliomas based on proteome analysis. *Cancer Res* 64, 2496–2501.

Jenkins, R.B., Blair, H., Ballman, K.V., Giannini, C., Arusell, R.M., Law, M., Flynn, H., Passe, S., Felten, S., Brown, P.D. et al. (2006). A t(1;19)(q10;p10) mediates the combined deletions of 1p and 19q and predicts a better prognosis of patients with oligodendroglioma. *Cancer Res* 66, 9852–9861.

Kefas, B., Godlewski, J., Comeau, L., Li, Y., Abounader, R., Hawkinson, M. et al. (2008). microRNA-7 inhibits the epidermal growth factor receptor and the Akt pathway and is down-regulated in glioblastoma. *Cancer Res* 68, 3566–3572.

Kleihues, P. and Cavenee, W.K. (2000). *World Health Organization Classification of Tumours: Pathology and Genetics: Tumours of the Nervous System*. IARC press.

Kotliarov, Y., Steed, M.E., Christopher, N., Walling, J., Su, Q., Center, A., Heiss, J., Rosenblum, M., Mikkelsen, T., Zenklusen, J.C. et al. (2006). High-resolution global genomic survey of 178 gliomas reveals novel regions of copy number alterations and allelic imbalances. *Cancer Res* 66, 9428–9436.

Kuan, C.T., Wakiya, K., Dowell, J.M., Herndon, J.E., II, Reardon, D.A., Graner, M.W., Riggins, G.J., Wikstrand, C.J. and Bigner, D.D. (2006). Glycoprotein nonmetastatic melanoma protein B, a potential molecular therapeutic target in patients with glioblastoma multiforme. *Clin Cancer Res* 12, 1970–1982.

Lassman, A.B., Rossi, M.R., Raizer, J.J., Abrey, L.E., Lieberman, F.S., Grefe, C.N., Lamborn, K., Pao, W., Shih, A.H., Kuhn, J.G. et al. (2005). Molecular study of malignant gliomas treated with epidermal growth factor receptor inhibitors: Tissue analysis from North American Brain Tumor Consortium Trials 01–03 and 00-01. *Clin Cancer Res* 11, 7841–7850.

Liang, Y., Diehn, M., Watson, N., Bollen, A.W., Aldape, K.D., Nicholas, M.K., Lamborn, K.R., Berger, M.S., Botstein, D., Brown, P.O. et al. (2005). Gene expression profiling reveals molecularly and clinically distinct subtypes of glioblastoma multiforme. *Proc Natl Acad Sci USA* 102, 5814–5819.

Louis, D.N. (2006). Molecular pathology of malignant gliomas. *Annu Rev Pathol Mech Dis* 1, 97–117.

Louis, D.N., Holland, E.C. and Cairncross, J.G. (2001). Glioma classification: A molecular reappraisal. *Am J Pathol* 159, 779–786.

Louis, D.N., Ohgaki, H., Wiestler, O.D., Cavenee, W.K., Burger, P.C., Jouvet, A., Scheithauer, B.W. and Kleihues, P. (2007). The 2007 WHO classification of tumors of the central nervous system. *Acta Neuropathol (Berl)* 114, 97–109.

Maher, E.A., Brennan, C., Wen, P.Y., Durso, L., Ligon, K.L., Richardson, A., Khatry, D., Feng, B., Sinha, R., Louis, D.N. et al. (2006). Marked genomic differences characterize primary and secondary glioblastoma subtypes and identify two distinct molecular and clinical secondary glioblastoma entities. *Cancer Res* 66, 11502–11513.

Mellinghoff, I.K., Wang, M.Y., Vivanco, I., Haas-Kogan, D.A., Zhu, S., Dia, E.Q., Lu, K.V., Yoshimoto, K., Huang, J.H., Chute, D.J. et al. (2005). Molecular determinants of the response of glioblastomas to EGFR kinase inhibitors. *N Engl J Med* 353, 2012–2024.

Mischel, P.S., Shai, R., Shi, T., Horvath, S., Lu, K.V., Choe, G., Seligson, D., Kremen, T.J., Palotie, A., Liau, L.M. et al. (2003). Identification of molecular subtypes of glioblastoma by gene expression profiling. *Oncogene* 22, 2361–2373.

Mischel, P.S., Cloughesy, T.F. and Nelson, S.F. (2004). DNA-microarray analysis of brain cancer: Molecular classification for therapy. *Nat Rev Neurosci* 5, 782–792.

Mukasa, A., Ueki, K., Matsumoto, S., Tsutsumi, S., Nishikawa, R., Fujimaki, T., Asai, A., Kirino, T. and Aburatani, H. (2002). Distinction in gene expression profiles of oligodendrogliomas with and without allelic loss of 1p. *Oncogene* 21, 3961–3968.

Nutt, C.L., Mani, D.R., Betensky, R.A., Tamayo, P., Cairncross, J.G., Ladd, C., Pohl, U., Hartmann, C., McLaughlin, M.E., Batchelor, T.T. et al. (2003). Gene expression-based classification of malignant gliomas correlates better with survival than histological classification. *Cancer Res* 63, 1602–1607.

Nutt, C.L., Betensky, R.A., Brower, M.A., Batchelor, T.T., Louis, D.N. and Stemmer-Rachamimov, A.O. (2005). YKL-40 is a differential diagnostic marker for histologic subtypes of high-grade gliomas. *Clin Cancer Res* 11, 2258–2264.

Offit, K., Levran, O., Mullaney, B., Mah, K., Nafa, K., Batish, S.D., Diotti, R., Schneider, H., Deffenbaugh, A., Scholl, T. et al. (2003). Shared genetic susceptibility to breast cancer, brain tumors, and Fanconi anemia. *J Natl Cancer Inst* 95, 1548–1551.

Pelloski, C.E., Mahajan, A., Maor, M., Chang, E.L., Woo, S., Gilbert, M., Colman, H., Yang, H., Ledoux, A., Blair, H. et al. (2005). YKL-40 expression is associated with poorer response to radiation and shorter overall survival in glioblastoma. *Clin Cancer Res* 11, 3326–3334.

Petalidis, L.P., Oulas, A., Backlund, M., Wayland, M.T., Liu, L., Plant, K. et al. (2008). Improved grading and survival prediction of human astrocytic brain tumors by artificial neural network analysis of gene expression microarray data. *Mol Cancer Ther* 7, 1013–1024.

Phillips, H.S., Kharbanda, S., Chen, R., Forrest, W.F., Soriano, R.H., Wu, T.D., Misra, A., Nigro, J.M., Colman, H., Soroceanu, L. et al. (2006). Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* 9, 157–173.

Rajasekhar, V.K., Viale, A., Socci, N.D., Wiedmann, M., Hu, X. and Holland, E.C. (2003). Oncogenic Ras and Akt signaling contribute to glioblastoma formation by differential recruitment of existing mRNAs to polysomes. *Mol Cell* 12, 889–901.

Rich, J.N., Hans, C., Jones, B., Iversen, E.S., McLendon, R.E., Rasheed, B.K., Dobra, A., Dressman, H.K., Bigner, D.D., Nevins, J.R. et al. (2005). Gene expression profiling and genetic markers in glioblastoma survival. *Cancer Res* 65, 4051–4058.

Rossi, M.R., Gaile, D., Laduca, J., Matsui, S., Conroy, J., McQuaid, D., Chervinsky, D., Eddy, R., Chen, H.S., Barnett, G.H. et al. (2005). Identification of consistent novel submegabase deletions in low-grade oligodendrogliomas using array-based comparative genomic hybridization. *Genes Chromosomes Cancer* 44, 85–96.

Schwartz, S.A., Weil, R.J., Thompson, R.C., Shyr, Y., Moore, J.H., Toms, S.A., Johnson, M.D. and Caprioli, R.M. (2005). Proteomic-based prognosis of brain tumor patients using direct-tissue matrix-assisted laser desorption ionization mass spectrometry. *Cancer Res* 65, 7674–7681.

Schwartzbaum, J., Ahlbom, A., Malmer, B., Lonn, S., Brookes, A.J., Doss, H., Debinski, W., Henriksson, R. and Feychting, M. (2005). Polymorphisms associated with asthma are inversely related to glioblastoma multiforme. *Cancer Res* 65, 6459–6465.

Schwartzbaum, J.A., Fisher, J.L., Aldape, K.D. and Wrensch, M. (2006). Epidemiology and molecular pathology of glioma. *Nat Clin Pract Neurol* 2, 494–503.

Shai, R., Shi, T., Kremen, T.J., Horvath, S., Liau, L.M., Cloughesy, T.F., Mischel, P.S. and Nelson, S.F. (2003). Gene expression profiling identifies molecular subtypes of gliomas. *Oncogene* 22, 4918–4923.

Silber, J., Lim, D.A., Petritsch, C., Persson, A.I., Maunakea, A.K., Yu, M. et al. (2008). miR-124 and miR-137 inhibit proliferation of glioblastoma multiforme cells and induce differentiation of brain tumor stem cells. *BMC Med* 6, 14.

Van den bent, M.J., Carpentier, A.F., Brandes, A.A., Sanson, M., Taphoorn, M.J., Bernsen, H.J., Frenay, M., Tijssen, C.C., Grisold, W., Sipos, L. et al. (2006). Adjuvant procarbazine, lomustine, and vincristine improves progression-free survival but not overall survival in newly diagnosed anaplastic oligodendrogliomas and oligoastrocytomas: A randomized European Organisation for Research and Treatment of Cancer phase III trial. *J Clin Oncol* 24, 2715–2722.

Volinia, S., Calin, G.A., Liu, C.G., Ambs, S., Cimmino, A., Petrocca, F., Visone, R., Iorio, M., Roldo, C., Ferracin, M. et al. (2006). A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc Natl Acad Sci USA* 103, 2257–2261.

Wang, L.E., Bondy, M.L., Shen, H., El-Zein, R., Aldape, K., Cao, Y., Pudavalli, V., Levin, V.A., Yung, W.K. and Wei, Q. (2004). Polymorphisms of DNA repair genes and risk of glioma. *Cancer Res* 64, 5560–5563.

Wang, L., Wei, Q., Wang, L.E., Aldape, K.D., Cao, Y., Okcu, M.F., Hess, K.R., El-Zein, R., Gilbert, M.R., Woo, S.Y. et al. (2006). Survival prediction in patients with glioblastoma multiforme by human telomerase genetic variation. *J Clin Oncol* 24, 1627–1632.

Watson, M.A., Perry, A., Budhraja, V., Hicks, C., Shannon, W.D. and Rich, K.M. (2001). Gene expression profiling with oligonucleotide microarrays distinguishes World Health Organization grade of oligodendrogliomas. *Cancer Res* 61, 1825–1829.

CHAPTER

Targeted Therapies for Cancer

Jeffrey S. Ross

INTRODUCTION

The regulatory approvals in the United States and Europe of trastuzumab (Herceptin®) for the treatment of HER2 overexpressing metastatic breast cancer and imatinib mesylate (Gleevec®) for the treatment of patients with *bcr/abl* translocation positive chronic myelogenous leukemia and gastrointestinal stromal tumors (GISTs) featuring an activating c-*kit* growth factor receptor mutation has created enthusiasm for anticancer targeted therapy in both the scientific and public communities (Mauro and Druker, 2001; O'Dwyer and Druker, 2001). Recent major news magazines and other public media have highlighted interest in new anticancer drugs that exploit disease-specific genetic defects as the target of their mechanism of action (Brown, et al., 2001; Lemonick and Parl, 2001). It is now widely held that the integration of molecular oncology and molecular diagnostics will further revolutionize oncology drug discovery and development; customize the selection, dosing, and route of administration of both previously approved traditional agents and new therapeutics in clinical trials, and individualize medical care for the cancer patient (Amos and Patnaik, 2002; Bottles, 2001; Evans and McLeod, 2003; Weinshilboum, 2003).

TARGETED THERAPIES FOR CANCER

From the regulatory perspective, "targeted therapy" has been defined as a drug in whose approval label there is a specific reference to a simultaneously or previously approved diagnostic test that must be performed before the patient can be considered eligible to receive that specific drug. The co-approvals of the anti-breast cancer antibody trastuzumab (Herceptin®) and the required tissue-based tests for patient eligibility (Herceptest®, Pathway®, InSite®, and Pathvysion®) are examples of this strict definition of targeted therapy. However, for many scientists and oncologists, anticancer drugs are considered to "targeted" when they feature a focused mechanism that specifically acts on a well-defined target or biologic pathway that, when inactivated, causes regression or destruction of the malignant process. Examples of this less rigorous definition of targeted therapy include hormonal-based therapies for breast cancer, small-molecule inhibitors of the epidermal growth factor receptor (EGFR), blockers of invasion and metastasis enabling proteins and enzymes, antiangiogenesis agents, proapoptotic drugs and proteasome inhibitors. Finally, another definition of targeted therapy involves anticancer antibody therapeutics that seek out and kill malignant cells bearing the target antigen.

THE IDEAL TARGET

The ideal cancer target (Table 42.1) can be defined as a macromolecule that is crucial to the malignant phenotype and is not significantly expressed in vital organs and tissues; that has biologic relevance that can be reproducibly measured in readily obtained clinical samples; that is definably correlated with clinical outcome; and that interruption, interference, or inhibition of such a macromolecule yields a clinical response in a significant proportion of patients whose tumors express the target with minimal to absent responses in patients whose tumors do not express the target. For antibody therapeutics, additional important criteria include the use of cell surface targets that when

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

TABLE 42.1 Features of the ideal anti-cancer target
■ Crucial to the malignant phenotype
■ Not significantly expressed in vital organs and tissues
■ A biologically-relevant molecular feature
■ Reproducibly measurable in readily obtained clinical samples
■ Correlated with clinical outcome
■ When interrupted, interfered with or inhibited, the result is a clinical response in a significant proportion of patients whose tumors express the target
■ Responses in patients whose tumors do not express the target are minimal

Reprinted from Ross, J.S. and Hortobagyi, G.H. eds. *The Molecular Oncology of Breast Cancer*. Jones and Bartlett, Inc., Sudbury, MA with permission by the publisher.

complexed with the therapeutic naked or conjugated antibody, internalize the antigen–antibody complex by reverse pinocytosis, thus facilitating tumor cell killing.

THE FIRST DIAGNOSTIC-THERAPEUTIC COMBINATION IN CANCER THERAPY: HORMONAL THERAPY FOR BREAST CANCER

Targeted therapy for cancer began in the early 1970s with the introduction of the estrogen receptor (ER) biochemical assay to select patients with painful metastatic breast cancer for surgical ablation of estrogen producing organs (ovaries, adrenals) (Figure 42.1) (Osborne, 1998). The ER assay was followed by a similar dextran coated charcoal biochemical assay for the progesterone receptor (PR) and subsequently converted to an immunohistochemistry (IHC) platform when the decreased size of primary tumors associated with self-examination and mammography-based screening programs prevented the use of the biochemical test (Wilbur et al., 1992). The drug tamoxifen (Nolvadex®), which has both hormonal and nonhormonal mechanisms of action, has been the most widely prescribed antiestrogen for the treatment of metastatic breast cancer and chemoprevention of the disease in high risk women (Ciocca and Elledge, 2000; Jordan, 2003a, b, c). Although, ER and PR testing is the front line for predicting tamoxifen response, additional biomarkers, including HER-2/neu (HER-2) and cathepsin D testing, have been used to further refine therapy selection (Locker, 1998). The introductions of specific estrogen response modulators and aromatase inhibitors such as anastrozole (Arimidex®), letrozole (Femara®), and the combination chemotherapeutic, estramustine (Emcyt®) (Buzdar et al., 2006; Ibrahim and Hortobagyi, 1999; Jordan, 2003a, b, c; Miller et al., 2002) have added new strategies for evaluating tumors for hormonal therapy.

Most recently, the Oncotype Dx® (Genomic Health, Redwood City, CA) multigene RT-PCR multiplex assay using a 21-gene probe set and mRNA extracted from paraffin blocks of stored breast cancer tissues was introduced as a new guide to the use of tamoxifen in ER positive node negative breast cancer patients (Paik et al., 2004). The assay features 16 cancer-related genes and 5 reference genes that were selected based on a series of transcriptional profiling experiments. The cancer-related genes include markers of proliferation including Ki-67, markers of apoptosis including survivin, invasion-associated protease genes including MMP11 and cathepsin L2, ER and HER2/*neu* gene family members, the glutathione-S-transferase genotype M1, CD68, a lysosomal monocyte/macrophage marker and BAG1, a co-chaperone glucocorticoid receptor associated with bcl-2 and apoptosis. Using a cohort of 688 lymph node negative, ER+ tumors obtained from patients enrolled in the NSABP B-14 clinical trial treated with tamoxifen alone, the 21-gene assay produced three prognosis scores of low, intermediate and high risk. The recurrence rates for these patients at 10 years follow-up was 7% for the low risk, 14% for the intermediate risk and 31% for the high risk groups. The difference in relapse rates between the low risk and high risk patients was highly significant ($p < 0.001$). On multivariate analysis this assay predicted adverse outcome independent of tumor size and also predicted overall survival (Paik et al., 2004). Although not currently approved by the FDA, the interest in this new assay has been intense and it has become commercially available in a centralized format for new patients. Recent data presented at the 2005 ASCO Meeting showed that the Oncotype Dx® is also capable of performing as a stand-alone prognostic test based on the test results in an untreated patient population (Paik et al., 2005). Detailed evaluation of the gene set in the Oncotype Dx® assay indicates that the mRNA levels of ER appear to be the most significant predictors in the node-negative ER-positive population (by IHC). Further studies are needed to validate the assay, learn its best uses and limitations given the evolving approach to hormonal therapy with non-tamoxifen drugs, the wide use of cytotoxic agents in the adjuvant setting for node-negative patients and the availability of both RT-PCR based and non-RT-PCR approaches to predicting breast cancer response to anti-estrogen and other anti-neoplastic agents used for treatment of the disease (Bast and Hortobagyi, 2004). In current practice the routine testing of tumor samples for androgen receptor status has not been incorporated into the selection of hormonal therapy for the disease.

DIAGNOSTIC-THERAPEUTIC COMBINATIONS FOR LEUKEMIA AND LYMPHOMA

The introduction of immunophenotyping for leukemia and lymphoma was followed by the first applications of DNA-based assays, the polymerase chain reaction, and RNA-based molecular technologies in these diseases that complemented continuing advances in tumor cytogenetics (Gleissner and Thiel, 2001; Rubnitz et al.,

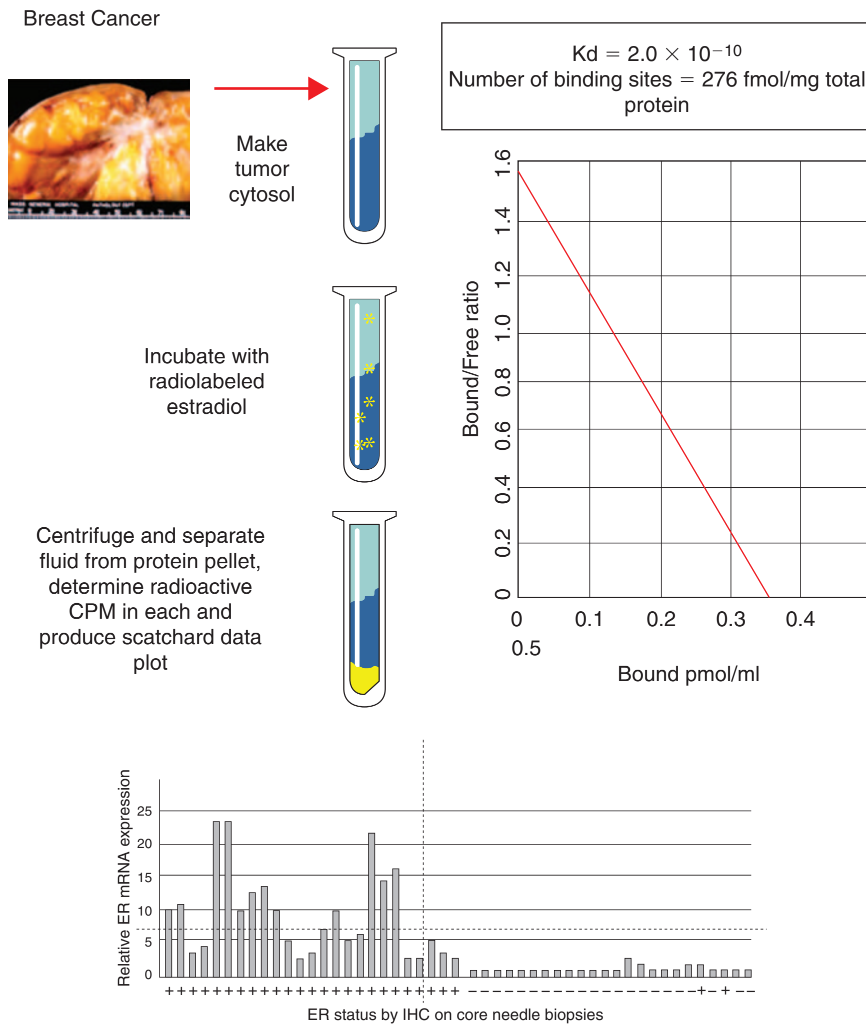


Figure 42.1 ER status determination. (a) Biochemical competitive binding assay for ER status determination. (b) Comparison of ER messenger RNA expression detected by microarray profiling and corresponding ER protein expression measured by IHC. The concordance between ER levels determined by IHC and ER levels determined by gene expression profiling was about 95%. Reprinted from Ross, J.S, Hortobagyi, G.H. eds. *The Molecular Oncology of Breast Cancer*. Jones and Bartlett, Inc., Sudbury, MA with permission by the publisher.

1999). In addition to the imatinib (Gleevec®) targeted therapy for chronic myelogenous leukemia, other molecular targeted therapies in hematologic malignancies includes the use of all-*trans*-retinoic acid (ATRA) for the treatment of acute promyelocytic leukemia (Grimwade and Lo Coco, 2002); anti-CD20 antibody therapeutics targeting non-Hodgkin lymphomas, including rituximab (Rituxan®) (Kiyoi and Naoe, 2005); and the emerging Flt-3 target for a subset of acute myelogenous leukemia patients (see below) (Ross et al., 2003a).

HER-2 POSITIVE BREAST CANCER AND TRASTUZUMAB (HERCEPTIN®)

After the introduction of hormone receptor testing, some 30 years then elapsed before the next major targeted cancer chemotherapy program for a solid tumor was developed. In the mid-1980s, the discovery of the *HER-2* (c-*erb*B2) gene and protein and subsequent association with an adverse outcome in breast cancer provided clinicians with a new biomarker that could be used to guide adjuvant chemotherapy (Slamon et al., 2001). The development of trastuzumab (Herceptin®), a humanized monoclonal antibody designed to treat advanced metastatic breast cancer that had failed first- and second-line chemotherapy, caused a rapid wide adoption of HER-2 testing of the patients' primary tumors (Schnitt and Jacobs, 2001). However, soon after its approval, widespread confusion concerning the most appropriate diagnostic test to determine HER-2 status in formalin-fixed paraffin-embedded breast cancer tissues substantially impacted trastuzumab use (Hayes and Thor, 2002; Hortobagyi, 2001; Masood and Bui, 2002, Paik et al., 2002;

Tanner et al., 2000; Wang et al., 2000, 2001; Zhao et al., 2002). Since its launch in 1998, trastuzumab has become an important therapeutic option for patients with HER-2-positive breast cancer (Bast et al., 2000; Ligibel and Winer, 2002; McKeage and Perry, 2002, Shawver et al., 2002).

In general, when specimens have been carefully fixed, processed and embedded, there has been excellent correlation between *HER-2* gene copy status determined by Fluorescence *In Situ* Hybridization (FISH) and HER-2 protein expression levels determined by IHC (Slamon et al., 2001). The main use of either method in current clinical practice is focused on the negative prediction of response to trastuzumab. Currently, both the American Society of Clinical Oncology and the College of American Pathologists consider HER-2 testing to be part of the standard work-up and management of breast cancer (Hammond et al., 2000; Pawlowski et al., 2000). Recently, the chromogenic (non-fluorescent) *in situ* hybridization technique has been used to determine the *HER-2* gene amplification status with promising results (Hortobagyi, 2001; Zhao et al., 2002). Non-morphologic approaches for determining HER-2 status have also been developed. The RT-PCR technique which has been predominantly used to detect HER-2 mRNA in peripheral blood and bone marrow samples, has correlated more with gene amplification status than IHC levels of primary tumors, but has been less successful as a predictor of survival (Bieche et al., 1999; Dressman et al., 2003; Tubbs et al., 2001). With the advent of laser capture microscopy and the acceptance of RT-PCR as a routine and reproducible laboratory technique, the use of RT-PCR for the determination of HER-2 status may increase in the future. The cDNA microarray-based method of detecting HER-2 mRNA expression levels has recently received interest as an alternative method for measuring HER-2/neu status in breast cancer (Fornier et al., 2005; Pusztai et al., 2003). Finally, the serum HER-2/ELISA test measuring circulating HER-2 (p185neu) protein is an FDA-approved test that has seen increased clinical use as a method for monitoring the response to trastuzumab (Ross et al., 2005). A summary of HER-2 testing methods in breast cancer is shown in Table 42.2.

TABLE 42.2 Methods of detection of HER-2/neu status in breast cancer

Method	Target	FDA-approved	Slide-based
IHC	Protein	Yes[a]	Yes
FISH	Gene	Yes[a]	Yes
CISH	Gene	No	Yes
Southern blot	Gene	No	No
RT-PCR	mRNA	No	No
Microarray TP	mRNA	No	No
Tumor ELISA	Protein	No	No
Serum ELISA	Protein	Yes[b]	No

Reprinted from Ross, J.S. and Hortobagyi, G.H. eds. *The Molecular Oncology of Breast Cancer*. Jones and Bartlett, Inc., Sudbury, MA with permission by the publisher.
IHC = immunohistochemistry
FISH = fluorescence *in situ* hybridization
CISH = chromogenic *in situ* hybridization
RT-PCR = reverse transcriptase polymerase chain reaction
TP = transcriptional profiling
ELISA = enzyme-linked immunosorbent assay
[a]For prognosis and prediction of response and eligibility to receive trastuzumab therapy
[b]For monitoring response of breast cancer to treatment

OTHER TARGETED ANTICANCER THERAPIES USING ANTIBODIES

An unprecedented number and variety of targeted small molecule and antibody-based therapeutics are currently in early development and clinical trials for the treatment of cancer. Therapeutic antibodies have become a major strategy in clinical oncology because of their ability to specifically bind to primary and metastatic cancer cells with high affinity and create antitumor effects by complement-mediated cytolysis and antibody-dependent cell-mediated cytotoxicity (naked antibodies) or by the focused delivery of radiation or cellular toxins (conjugated antibodies) (Carter, 2001; Goldenberg, 2002; Hemminki, 2002; Milenic, 2002; Reichert, 2002; Ross et al., 2003b,). Currently, there are eight anticancer therapeutic antibodies approved by the US Food and Drug Administration (FDA) for sale in the United States (Table 42.3). Therapeutic monoclonal antibodies are typically of the IgG class containing two heavy and two light chains. The heavy chains form a fused "Y" structure with two light chains running in parallel to the open portion of the heavy chain. The tips of the heavy-light chain pairs form the antigen binding sites, with the primary antigen recognition regions known as the complementarity determining regions.

The early promise of mouse monoclonal antibodies for the treatment of human cancers was not realized because (1) unfocused target selection led to the identification of target antigens that were not critical for cancer cell survival and progression, (2) there was a low overall potency of naked mouse antibodies as anticancer drugs, (3) antibodies penetrated tumor cells poorly, (4) there was limited success in producing radioisotope and toxin conjugates, and (5) the development of human antimouse antibodies (HAMA) prevented the use of multiple dosing schedules (Reilly et al., 1995).

The next advance in antibody therapeutics began in the early 1980s when recombinant DNA technology was applied to antibody design to reduce the antigenicity of murine and other rodent-derived monoclonal antibodies. Chimeric antibodies were developed where the constant domains of the human IgG molecule were combined with the murine variable regions by transgenic fusion of the immunoglobulin genes; the chimeric monoclonal antibodies were produced from engineered hybridomas and Chinese Hamster Ovary (CHO) cells (Merluzzi et al., 2000; Winter and

Table 42.3 Antibody therapeutics for cancer

Name	FDA approval date	Source *Partners*	Type	Target	Indication(s) (both approved and investigational)
Alemtuzumab *Campath®*	05/01	BTG *ILEX Oncology Schering AG*	Monoclonal antibody, humanized Anticancer, immunological Multiple sclerosis treatment Immunosuppressant	CD52	Cancer, leukemia, chronic lymphocytic Cancer, leukemia, chronic myelogenous Multiple sclerosis, chronic progressive
Daclizumab *Zenapax®*	03/02	Protein Design Labs *Hoffmann-La Roche*	Monoclonal IgG_1 Chimeric Immunosuppressant Antipsoriasis Antidiabetic Ophthalmological Multiple sclerosis treatment	IL2R	Transplant rejection, general Transplant rejection, bone marrow Uveitis Multiple sclerosis, relapsing-remitting Multiple sclerosis, chronic progressive Cancer, leukemia, general Psoriasis Diabetes, type I Asthma Colitis, ulcerative
Rituximab *Rituxan®*	11/97	IDEC *Genentech Hoffmann-La Roche Zenyaku Kogyo*	Monoclonal IgG_1 Chimeric Anticancer, immunological Antiarthritic, immunological Immunosuppressant	CD20	Cancer, lymphoma, non-Hodgkin's Cancer, lymphoma, B-cell Arthritis, rheumatoid Cancer, leukaemia, chronic lymphocytic Thrombocytopenic purpura
Trastuzumab *Herceptin®*	09/98	Genentech *Hoffmann-La Roche ImmunoGen*	Monoclonal IgG_1 Humanized Anticancer, immunological	p185neu	Cancer, breast Cancer, lung, non-small cell Cancer, pancreatic
Gemtuzumab *Mylotarg®*	05/00	Wyeth/AHP	Monoclonal IgG_4 Humanized	CD33 / coleacheamycin	Cancer, leukemia, AML (patients older than 60 years)
Ibritumomab *Zevalin®*	02/02	IDEC	Monoclonal IgG_1 Murine Anticancer	CD20 / ^{90}Yttrium	Cancer, lymphoma, low grade, follicular, Transformed non-Hodgkin's (relapsed or refractory)
Tositomumab *Bexxar®*	06/03	Corixa	Anti-CD 20 Murine Monocolonal antibody with ^{131}I conjugation	CD20	Cancer, lymphoma, non-Hodgkin's
Cetuximab *Erbitux®*	02/04	Imclone Bristol Myers Squibb	Anti-EGFR monoclonal antibody	EGFR	Approved for third line treatment of metatstatic colorectal cancer that has failed primary chemotherapy
Bevacizumab *Avastin®*	02/04	Genentech	Anti-VEGF (ligand)	VEGF	Avastin is approved for use in combination with intravenous 5-fluorouracil-based chemotherapy as a treatment for patients with first-line – or previously untreated – metastatic colorectal cancer.
Edrecolomab *Panorex™*	01/95 (Europe Only, nut FDA-approved)	Glaxo-Smith-Kline	Monoclonal IgG_{2A} Murine Anticancer	Epithelial cell adhesion molecule (EpCAM)	Cancer, colorectal

Adapted from Ross, J.S. and Hortobagyi, G.H. eds. *The Molecular Oncology of Breast Cancer*. Jones and Bartlett, Inc., Sudbury, MA with permission by the publisher.

Harris, 1993). The use of chimeric antibodies significantly reduced the HAMA responses but did notcompletely eliminate them (Kuus-Reichel et al., 1994; Merluzzi et al., 2000). Although several chimeric antibodies achieved regulatory approval, certain targets required humanized antibodies to achieve appropriate dosing. Partially humanized antibodies were then developed where the six complementarity determining regions of the heavy and light chains and a limited number of structural amino acids of the murine monoclonal antibody were grafted by recombinant technology to the complementarity determining region depleted human IgG scaffold (Milenic, 2002). Although this process further reduced or eliminated the HAMA responses, in many cases significant further antibody design procedures were needed to reestablish the required specificity and affinity of the original murine antibody (Jones et al., 1986, Isaacs, 2001; Pimm, 1994; Watkins and Ouwehand, 2000).

A second approach to reducing the immunogenicity of monoclonal antibodies has been to replace immunogenic epitopes in the murine variable domains with benign amino acid sequences, resulting in a deimmunized variable domain. The deimmunized variable domains are genetically linked to human IgG constant domains to yield a deimmunized antibody (Biovation, Aberdeen, Scotland). Additionally, primatized antibodies were subsequently developed that featured a chimeric antibody structure of human and monkey that, as a near exact copy of a human antibody, further reduced immunogenicity and enabled the capability for continuous repeat dosing and chronic therapy (Reff et al., 2002). Finally, fully human antibodies have now been developed using murine sources and transgenic techniques (Reff et al., 2002).

Using modern antibody design and deimmunization technologies, scientists and clinicians have attempted to improve the efficacy and reduce the toxicity of anticancer antibody therapeutics (Carter, 2001; Chester and Hawkins, 1995; Reff Heard, 2001; Reilly et al., 1995; Nielsen and Marks, 2000). The bacteriophage antibody design system has facilitated the development of high affinity antibodies by increasing antigen binding rates and reducing corresponding detachment rates (Nielsen and Marks, 2000). Increased antigen binding is also achieved in bivalent antibodies with multiple attachment sites, a feature known as avidity. Modern antibody design has endeavored to create small antibodies that can penetrate to cancerous sites but maintain their affinity and avidity. A variety of approaches has been used to increase antibody efficacy (Carter, 2001). Clinical trials have recently combined anticancer antibodies with conventional cytotoxic drugs, yielding promising results (Carter, 2001; Goldenberg, 2002; Hemminki, 2002; Milenic, 2002; Reichert, 2002; Ross et al., 2003b). The applications of radioisotope, small molecule cytotoxic drug, and protein toxin conjugation have resulted in promising results in clinical trials and achieved regulatory approval for several drugs now on the market (see below). Antibodies have also been designed to increase their enhancement of effector functions of antibody-dependent cellular cytotoxicity. Another cause of toxicity of conjugated antibodies has been the limitations of the conjugation technology, which can restrict the ratio of the number of toxin molecules per antibody molecule (Carter, 2001; Goldenberg, 2002; Watkins and Ouwehand, 2000). Methods designed to overcome the toxicity of conjugated antibodies include the use of antibody targeted liposomal small molecule drug conjugates and the use antibody conjugates with drugs in nanoparticle formats to enhance bonding strength that enable controlled release of the cytotoxic agent. Another technique that uses site selective prodrug activation to reduce bystander tissue toxicity is the antibody directed enzyme prodrug therapy. An antibody-bound enzyme is targeted to tumor cells. This allows for selective activation of a nontoxic prodrug to a cytotoxic agent at the tumor site for cancer therapy.

A variety of factors can reduce antibody efficacy (Reilly et al., 1995): (1) limited penetration of the antibody into a large solid tumor or into vital regions such as the brain, (2) reduced extravasations of antibodies into target sites due to decreased vascular permeability, (3) cross-reactivity and nonspecific binding of antibody to normal tissues reduces targeting effect, (4) heterogeneous tumor uptake results in untreated zones, (5) increased metabolism of injected antibodies reduces therapeutic effects, and (6) HAMA and human antihuman antibodies form rapidly and inactivate the therapeutic antibody.

Toxicity has been a major obstacle in the development of therapeutic antibodies for cancer [Carter, 2001; Goldenberg, 2002; Ross et al., 2003b; Watkins and Ouwehand, (2000)]. Cross-reactivity with normal tissues can cause significant side effects for unconjugated (naked) antibodies, which can be enhanced when the antibodies are conjugated with toxins or radioisotopes. Immune-mediated complications can include dyspnea from pulmonary toxicity, occasional central and peripheral nervous system complications, and decreased liver and renal function. On occasion, unexpected toxic complications can be seen, such as the cardiotoxicity associated with the HER-2 targeting antibody trastuzumab. Radioimmunotherapy with isotopic-conjugated antibodies can also cause bone marrow suppression (see below).

Unconjugated or naked antibodies include a variety of targeting molecules both on the market and in early and late clinical development. A variety of mechanisms have been cited to explain the therapeutic benefit of these drugs, including enhanced immune effector functions and direct inactivation of the targeted pathways as seen in the antibodies directed at surface receptors such as HER-1 (EGFR) and HER-2 (Amos and Patnaik, 2002; Brown et al., 2001; Lemonick and Parl, 2001; Mauro and Druker, 2001). Surface receptor targeting can reduce intracellular signaling, resulting in decreased cell growth and increased apoptosis (Reff et al., 2002).

As seen in Table 42.3, of the nine anticancer antibodies on the market in the United States, two are conjugated with a radioisotope ^{90}Y-ibritumomab tiuxetan (Zevalin®) and ^{131}I-tositomumab (Bexxar®) and one is conjugated to a complex natural product toxin gemtuzumab ozogamicin (Mylotarg®). Conjugation procedures have been designed to improve antibody therapy efficacy and have used a variety of methods to complex the isotope, toxin, or cytotoxic agent to the antibody (Carter, 2001; Goldenberg, 2002). Cytotoxic small molecule drug conjugates have been widely tested, but enthusiasm for this

approach has been limited by the relatively low potency of these compounds (Carter, 2001). Fungal derived potent toxins have yielded greater success with the calicheamicin conjugated anti-CD33 antibody gemtuzumab ozogamicin approved for the treatment of acute myelogenous leukemia and a variety of antibodies conjugated with the fungal toxin maytansanoid (DM-1) in preclinical development and early clinical trials. The interest in radioimmunotherapy increased significantly in 2001 with the FDA approvals of the ^{90}Y-conjugated anti-CD20 antibody ^{90}Y-ibritumomab tiuxetan and the ^{131}I-conjugated anti-CD20 antibody ^{131}I-tositumomab. A variety of isotopes are under investigation in addition to ^{90}Y as potential conjugates for anticancer antibodies (Goldenberg, 2002). Radioimmunotherapy features the phenomenon of the bystander effect, in which if antigen expression is heterogeneous, extensive tumor cell killing can still take place, even on nonexpressing cells, but can also lead to significant toxicity when the neighboring cells are vital non-neoplastic tissues such as the bone marrow and liver.

Antibody Therapeutics for Hematologic Malignancies

The earliest and most successful clinical use of antibodies in oncology has been for the treatment of hematologic malignancies (Burke et al., 2002; Carter, 2001; Goldenberg, 2002; Linenberger et al., 2002; Reff et al., 2002; Ross et al., 2003b; Stevenson et al., 2002; Watkins and Ouwehand, 2000; Wiseman et al., 2002). By taking advantage of improved recombinant technologies generating more specific and higher affinity monoclonal antibodies with reduced immunogenicity after humanization or deimmunization and the emerging conjugation capabilities, antibody therapeutics have become a major weapon in the treatment of leukemias and lymphomas (Burke et al., 2002; Linenberger et al., 2002; Stevenson et al., 2002; Wiseman et al., 2002).

Rituximab (Rituxan®)

Approved in 1997, rituximab (Rituxan®) is arguably the most commercially successful anticancer drug of any type since the introduction of taxanes. Rituximab sales exceeded $ 700 million in sales in the United States in 2001 (Reichert, 2002). Targeting the CD20 surface receptor common to many B cell non-Hodgkin lymphoma subtypes, rituximab is a chimeric monoclonal IgG_1 antibody that induces apoptosis, antibody-dependent cell cytotoxicity, and complement-mediated cytotoxicity (Reff et al., 2002) and has achieved significantly improved disease-free survival rates compared with patients receiving cytotoxic agents alone (Dillman, 2001; Coiffier, 2002; Grillo-Lopez, 2002; Grillo-Lopez et al., 2002).

^{90}Y-ibritumomab Tiuxetan (Zevalin®)

^{90}Y-ibritumomab tiuxetan (Zevalin®) consists of the murine version of the anti-CD20 chimeric monoclonal antibody, rituximab, which has been covalently linked to the metal chelator, MD-DTPA, permitting stable binding of ^{111}In when used for radionucleotide tumor imaging and ^{90}Y when used to produce enhanced targeted cytotoxicity (Dillman, 2002; Gordon et al., 2002; Krasner and Joyce, 2001; Wagner et al., 2002). In early 2002, ^{90}Y-ibritumomab tiuxetan became the first radioconjugated antibody therapeutic for cancer approved by the FDA. Since its FDA approval, numerous patients who have received ^{90}Y-ibritumomab tiuxetan after becoming refractory to a rituximab-based regimen have achieved significant responses (Dillman, 2001; Gordon et al., 2002).

Gemtuzumab ozogamicin (Mylotarg®)

The approval of gemtuzumab ozogamicin (Mylotarg®) by the FDA in 2000 marked the first introduction of a plant toxin conjugated antibody therapeutic (Bross et al., 2001; Larson et al., 2002; Nabhan and Tallman, 2002; Sievers and Linenberger, 2001; Stadtmauer, 2002). Gemtuzumab ozogamicin is targeted against CD33, a surface marker expressed by 90% of myeloid leukemic blasts but absent from stem cells, armed with calicheamicin, a potent cytotoxic antibiotic that inhibits DNA synthesis and induces apoptosis (Stadtmauer, 2002). The current indication for use of gemtuzumab ozogamicin is in acute myelogenous leukemia patients older than 60 years with the recommendation that before the initiation of therapy, the leukemic blast count is below 30,000/mL (Bross et al., 2001; Nabhan and Tallman, 2002; Sievers and Linenberger, 2001).

Alemtuzumab (Campath®)

Alemtuzumab (Campath®), a humanized monoclonal antibody, was approved in mid-2001 for the treatment of B-cell chronic lymphocytic leukemia in patients who have been treated with alkylating agents and who have failed fludarabine therapy (Dumont, 2002; Pangalis et al., 2001).

Daclizumab (Zenapax®)

Daclizumab (Zenapax®) is a chimeric monoclonal antibody that targets the interleukin-2 receptor. This antibody is primarily used to prevent and treat patients with organ transplant rejection but has also been used in a wide variety of chronic inflammatory conditions, including psoriasis, multiple sclerosis, ulcerative colitis, asthma, type I diabetes mellitus, uveitis, and also in a variety of leukemias (Carswell et al., 2001, Kreitman et al., 1992).

^{131}I-tositumomab (Bexxar®)

^{131}I-tositumomab (Bexxar®) is a radiolabeled anti-CD20 murine monoclonal antibody approved in 2003 for the treatment of relapsed and refractory follicular/low grade and transformed non-Hodgkin lymphoma (Cheson, 2002; Zelenetz, 2003).

Antibody Therapeutics for Solid Tumors

Interest in the development of antibody therapeutics for solid tumors among many commercial organizations and universities has been significantly impacted by the technologic advances in antibody engineering and the approval and recent clinical and commercial success of trastuzumab, the only therapeutic antibody approved by the FDA for the treatment of solid tumors (edrecolomab is approved in Germany, but not in the United States).

Trastuzumab (Herceptin®)

Trastuzumab (Herceptin®) has been described above. During the 6 years since the FDA approval of trastuzumab, two additional antibodies have been approved for the treatment of solid tumors (cetuximab and bevacizumab). In addition, continuing progress has been made in this field, and there are a number of both late stage and early stage products in development which show substantial promise.

Cetuximab (Erbitux®)

The epidermal growth factor receptor (EGFR), also known as HER-1, is the target of two FDA-approved small molecule drugs (see below) and one FDA-approved antibody (Mendelsohn and Baselga, 2000). Cetuximab (Erbitux®) is a chimeric monoclonal antibody that binds to the EGFR with high affinity, blocking growth factor binding, receptor activation, and subsequent signal transduction events and leading to cell proliferation (Baselga, 2001). Cetuximab enhanced the antitumor effects of chemotherapy and radiotherapy in preclinical models by inhibiting cell proliferation, angiogenesis, and metastasis and by promoting apoptosis (Baselga, 2001). Cetuximab has been evaluated both alone and in combination with radiotherapy and various cytotoxic chemotherapeutic agents in a series of phase II/III studies that primarily treated patients with either head and neck or colorectal cancer (Baselga, 2001, Herbst and Langer, 2002). Breast cancer trials are also underway (Leonard et al., 2002). Although the FDA approval process for cetuximab was initially slowed because of concerns over clinical trial design and outcome data management (Reynolds, 2002), the antibody was approved for use in the treatment of advanced metastatic colorectal cancer in February 2004. Similar to trastuzumab, the development of cetuximab also included an immunohistochemical test for determining EGFR overexpression to define patient eligibility to receive the antibody (Wong, 2005). Thus, cetuximab has joined trastuzumab as an FDA-approved targeted therapy featuring an unconjugated antibody. However, there have been conflicting reports suggesting that the use of a pharmacodiagnostic test (EGFR immunostaining) is unnecessary for the selection of cetuximab in colorectal cancer therapy (Saltz, 2005). Recent clinical trials have found significant efficacy for cetuximab in the treatment of head and neck squamous cell cancers often in combination with radiation treatment (Harari and Huang, 2006).

Bevacizumab (Avastin®)

Bevacizumab (rhuMAb-VEGF) is a humanized murine monoclonal antibody targeting the vascular endothelial growth factor ligand (VEGF) approved by the FDA in 2004 for the front line or first line treatment in combination with chemotherapy of metatstatic colorectal cancer. VEGF regulates both vascular proliferation and permeability and functions as an antiapoptotic factor for newly formed blood vessels (Chen et al., 2000; Ferrara, 2005; Rosen, 2002). In addition to its approved indication in colorectal cancer, bevacizumab has shown promising efficacy in combination with cytotoxic drugs for the treatment of non-small cell lung cancer (Wakelee and Belani, 2005), renal cell carcinoma (Stadler, 2005), pancreatic cancer (Bruckner et al., 2005), breast cancer (de Gramont and Van Cutsem, 2005) and prostate cancer (Berry and Eisenberger, 2005). Unlike cetuximab, the development of bevacizumab has not included a diagnostic eligibility test. Neither direct measurement of VEGF expression in tumor, circulating VEGF levels in serum or urine or assessment of tumor microvessel density have been incorporated into the clinical trials or linked to the response rates to the antibody. To date a number of theories have been proposed as to the actual mechanism of action of bevacizumab and the relative contributions of direct anti-angiogenesis and other tumor vasculature stabilization and cytotoxic chemotherapy potentiation effects of the antibody (Blagosklonny, 2005; Hurwitz and Kabbinavar, 2005). In summary, currently used without an integrated diagnostic eligibility test, bevacizumab cannot be considered a true targeted therapy, and further development of this agent for use in prostatic, breast, lung, renal, and other cancers may will be inhibited by the inability to individually select patients who will be more likely to benefit from its use, either alone or in combination with other traditional cytotoxic drugs, antibodies and novel drugs.

Edrecolomab (Panorex®)

Edrecolomab is a murine IgG_{2A} monoclonal antibody that targets the human tumor-associated antigen Ep-CAM (17-1A). Edrecolomab has been approved in Europe (Germany) since 1995 to date has not been approved by the FDA. In a study of 189 patients with resected stage III colorectal cancer, treatment with edrecolomab resulted in a 32% increase in overall survival compared with no treatment ($P < 0.01$) (Schwartzberg, 2001). Edrecolomab's antitumor effects are mediated through antibody-dependent cellular cytotoxicity, complement-mediated cytolysis, and the induction of an antiidiotypic network (Haller, 2001). Edrecolomab is also currently being tested in large multicenter adjuvant phase III studies in stage II/III rectal cancer and stage II colon cancer. Edrecolomab was well tolerated when used as monotherapy and added little to chemotherapy-related side effects when used in combination. Sequential treatment of patients with metastatic breast cancer with edrecolomab after adjuvant chemotherapy reduced levels of disseminated tumor cells in the bone marrow and eliminated Ep-CAM-positive micrometastases (Kirchner et al., 2002).

huJ-591 (Anti-$PSMA_{EXT}$)

Prostate-specific membrane antigen (PSMA) is a membrane-bound glycoprotein restricted to normal prostatic epithelial cells, prostate cancer, and the endothelium of the neovasculature of a wide variety of non-prostatic carcinomas and other solid tumors (Figure 42.2) (Israeli et al., 1994; Liu et al., 1997; Ross, 2005). PSMA expression per cell progressively increases in primary prostate cancer, metastatic hormone sensitive prostate cancer, and hormone refractory metastatic disease. PSMA expression is

increased further in association with clinically advanced prostatic cancer, particularly in hormone refractory disease, and appears to be an ideal molecule for use in targeting prostatic cancer cells. Increasing expression levels of PSMA in resected primary prostate cancer is associated with increased rates of subsequent disease recurrence (Ross et al., 2003c). Humanized and fully human antibodies specific for the extracellular domain of PSMA have been developed. A phase I clinical trial of one these antibodies, huJ591 conjugated with ^{90}Y, has yielded promising results (Milowsky et al., 2004). Programs using toxin conjugates with anti-PSMA antibodies have completed preclinical development (Fracasso et al., 2002) and are currently showing promising results in early stage clinical trials for hormone-refractory advanced metastatic prostate cancer (Milowsky et al., 2005). Finally, antibodies to PSMA have been used as diagnostic imaging agents (Figure 42.2), including the commercially available Prostascint® Freeman et al., 2002).

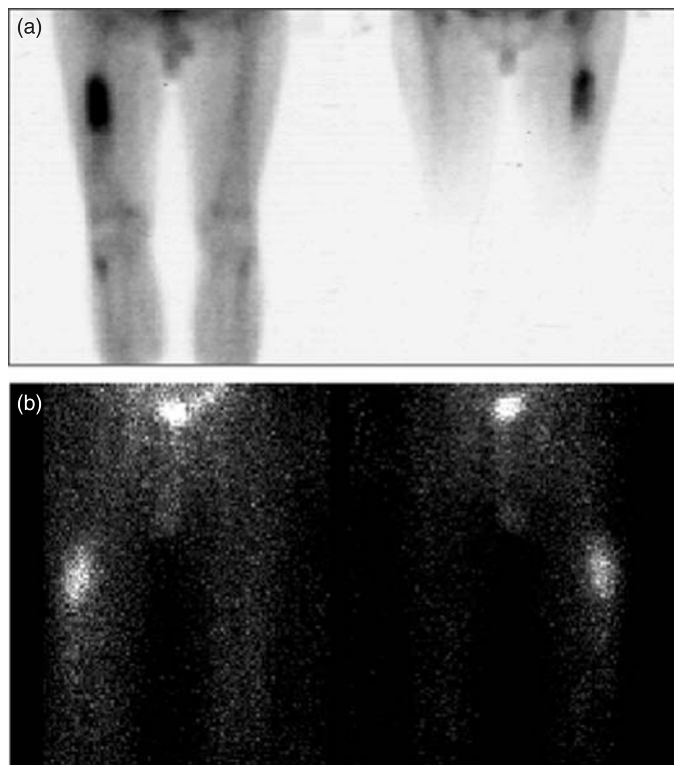


Figure 42.2 PSMA expression in non-prostate cancer. (a) Traditional bone scan demonstrating bilateral activity in the femur indirectly indicating the presence of metastatic renal cell carcinoma. (b) ^{111}I-huJ591$_{EXT}$ diagnostic immunoscintiscan of the same patient showing direct localization of the anti-PSMA antibody conjugate to the sites of metastatic renal cell carcinoma that feature PSMA expression in the tumor neovasculature. Reprinted from Ross, J.S. and Foster, C.S. eds. *The Molecular Oncology of Prostate Cancer*. 2006. Sudbury, MA: Jones and Bartlett, Inc. with permission by the publisher.

SELECTED TARGETED ANTICANCER THERAPIES USING SMALL MOLECULES

Table 42.4 lists selected small molecule drugs designed to target specific genetic events and biologic pathways critical to cancer growth, invasion, and metastasis.

Targeted Small Molecule Drugs for Hematologic Malignancies

ATRA

Arguably the first truly targeted therapy after the development of hormonal therapy for breast cancer was the development of ATRA for the treatment of acute promyelocytic leukemia, a subset of acute nonlymphocytic leukemia featuring a disease-defining retinoic acid receptor activating t(15:17) reciprocal translocation (Fang et al., 2002; Parmar and Tallman, 2003). For these selected patients, direct targeting of the retinoic acid receptor with ATRA has resulted in very high response rates, delay in disease progression, and long-term cures for these patients (Fang et al., 2002; Parmar and Tallman, 2003).

Imatinib (Gleevec®)

The development of imatinib for patients with chronic myelogenous leukemia in 2001 ushered in a new excitement both in the scientific and public communities for targeted anticancer therapy. Imatinib received fast-track approval by the FDA as an ATP-competitive selective inhibitor of *bcr-abl* and has unprecedented efficacy for the treatment of early stage chronic myelogenous leukemia typically achieving durable complete hematologic and complete cytogenetic remissions, with minimal toxicity (Druker, 2003; Goldman and Melo, 2003; O'Brien et al., 2003). Imatinib is a true targeted therapy for leukemia in that a test for the *bcr/abl* translocation must be performed before a patient will be considered as eligible to receive the drug. The prediction of resistance to imatinib in early phase Chronic Myeloid Leukemia (CML) has been the subject of numerous studies (Lange et al., 2005; O'hare et al., 2005). It is the current goal to predict resistance emergence with gene mutation testing and employ novel tyrosine kinase inhibitors to attempt to overcome blast cells that have lost the ability to bind imatinib to the ATP binding pocket of the fusion gene (Lange et al., 2005; O'hare et al., 2005).

Imatinib has also achieved regulatory approval for the treatment of relapsed and metastatic GISTs, which characteristically feature an activating point mutation in the c-*kit* receptor tyrosine kinase gene (von Mehren, 2003). For GISTs, the response to imatinib treatment appears to be predictable based on the location of the c-*kit* mutation (Verweij et al., 2003). The use of imatinib in GIST is also an example of targeted therapy as a measurement of c-*kit* expression usually performed by IHC, required to confirm the diagnosis and render the patient eligible for treatment. Interestingly, most commercially available

TABLE 42.4 Selected small molecule dugs designed to target specific genetic events and biologic pathways critical to cancer growth and progression

Target	Drug	Source	Clinical development status	Comment
PML-RAR-α in PML	ATRA	Promega	Approved	First true targeted therapy since the introduction of ER testing and hormonal therapy for breast cancer
Bcr/abl in CML	Imatinib	Novartis	Approved	Has emerged as standard of care for early stage CML
c-Kit in GIST PDGF-α	Imatinib	Novartis	Approved	Responses in relapsed/metastatic GIST can be predicted by the location of the activating *c-kit* mutation
Flt-3 in AML	SU5416 PKC412 MLN-518	Pfizer Novartis Millennium	Early Stage Clinical Trials	Small molecule drugs that target the *flt-3* internal tandem duplication seen in 30% of AML
EGFR in NSCLC	Gefitinib	Astra Zeneca	Approved/withdrawn	No survival benefit. Returned to clinical trials.
EGFR in NSCLC and pancreatic cancer	Erlotinib	Genentech/OSI	Approved	Survival benefit demonstrated. No diagnostic test current used to select patients.
Anti-angiogenesis in renal cell carcinoma	BAY 43-9006	Bayer	Approved	Raf kinase inhibitor also targets PDGFR and VEGFR.
Anti-angiogenesis in Myelodysplastic Syndrome	Lenolidamide	Celgene	Approved	Also in clinical trials for the treatment of multiple myeloma
Other anti-angiogenesis	Thalidomide Sunitinib	Celgene Pfizer/Sugen	Approved Approved	Multiple myeloma Gastrointestinal tumor
Bcl-2	G3135	Genta	Pending	Anti-sense oligonucleotide targets the anti-apoptotic gene, *bcl-2*
Proteasome in multiple myeloma	Bortezomib	Millennium	Approved	Proteasome inhibition effective in hematologic malignancies, but of uncertain potential for the treatment of solid tumors

Adapted from Ross, J.S. and Hortobagyi, G.H. eds. *The Molecular Oncology of Breast Cancer*. Jones and Bartlett, Inc., Sudbury, MA with permission by the publisher.

antibodies for *c-kit* recognize the total *c-kit* and do not distinguish the activated or phosphorylated version, which is the actual target of imatinib. Currently, the high treatment failure rate is directly linked to the test used to characterize the patients. It is anticipated that either the use of specific antibodies designed to identify the activated *c-kit* gene or directed sequencing of the *c-kit* gene may be required before imatinib is prescribed for patients with recurrent or metastatic GIST. An alternative to *c-kit* mutation testing for the prediction of resistance to imatinib, functional imaging after initial dosing of the drugs has been employed for patients with metastatic GIST (Heinicke et al., 2005). In early 2006, the anti-angiogenesis agent sunitinib was approved by the FDA for the treatment of advanced GIST that has become resistant to imatinib therapy.

Flt-3 Targeted Therapy

In approximately 30% of cases of acute myelogenous leukemia and less frequently in other forms of leukemia, a *flt-3* gene mutation creates an internal tandem duplication that creates an abnormal FLT3 receptor that promotes the growth and survival of the leukemic cells (Advani, 2005; Gilliland and Griffin, 2002; Kelly et al., 2002; Sawyers, 2002). Three small molecule compounds are in clinical trials for the treatment of acute myelogenous leukemia by targeting the *flt-3* internal tandem duplication. These drugs are also examples of potential true targeted therapies in that a test for detecting an internal tandem duplication that causes the *flt-3* gene activation will likely be required and incorporated into the FDA drug approval label should these agents be successful in future clinical trials.

Targeted Small Molecule Drugs for Solid Tumors

Gefitinib (Iressa®)

Gefitinib was originally approved by the FDA in 2003 as a monotherapy for the treatment of patients with locally advanced or metastatic non-small cell lung cancer after failure of both platinum based and docetaxel chemotherapies (Ranson, 2002; Schiller, 2003). Gefitinib is a small molecule drug that targets the EGFR. In contrast with the approval of trastuzumab, this approval of gefitinib did not include an eligibility requirement reference to a specific tumor diagnostic test designed to select patients that were more likely to respond to the drug. Overexpression of EGFR typically identified by IHC is extremely common in both lung and breast cancers (Campiglio et al., 2004; Ranson, 2002; Schiller, 2003), but in contrast with HER-2 overexpression, which is virtually limited to cases with gene amplification, multiple mechanisms of dysregulation of EGFR and associated activation of signaling pathways have been described for both of these tumors (Campiglio et al., 2004; Ranson, 2002; Schiller, 2003). Thus, it has been difficult to develop this drug for expanded indications or combination therapies in the absence of a well-defined efficacy test. However, more recently, two independent groups reported their similar discovery of a specific activating mutation in the tyrosine kinase domain of the EGFR receptor that was associated with a high response of patients with non-small cell lung cancer to gefitinib (Lynch et al., 2004; Paez et al., 2004). Of interest have been the consistent observations that both a bronchioloalveolar histology and a persistent skin rash have been the best clinical signals of gefitinib response in lung cancer (Dudek et al., 2005). In addition, although specific activating mutations in the EGFR gene have been reproduced in a number of studies (Chan et al., 2006), some studies have failed to demonstrate this association and other biomarkers including EGFR gene amplification have also been found to be predictive of tumor response (Carbone, 2004; Kobayashi et al., 2005). Most recently, follow-on studies of gefitinib in lung revealed that the increased response rates that led to the approval of the drug were not accompanied by a clinical survival advantage (Twombly, 2005). This has led to a current withdrawal of the drug while further research and clinical trials are performed. It is possible the gefitinib will reappear on the market for the treatment of lung cancer with an integrated diagnostic test designed to boost the response rates by limiting the treatment to tumors with specific histologic and molecular features.

Erlotinib (Tarceva®)

Erlotinib is another targeted small molecule inhibitor of EGFR that was approved by the FDA in 2005 for the treatment of non-small cell lung cancer and pancreatic cancer (Moore, 2005; Smith, 2005). To date, similar to gefitinib, the clinical trials and FDA approval for erlotinib have not included an assessment of the EGFR status or other diagnostic test for eligibility to receive the drug. In lung cancer the predictors of tumor response including skin rash and brochioloalveolar histology have also applied to erlotinib as have the somewhat conflicting associations of both activating EGFR mutations and EGFR gene amplification as predictors of drug response (Chan et al., 2006; Silvestri and Rivera, 2005). Clinical trials have demonstrated that erlotinib does add a survival benefit to the treatment of both lung and pancreatic cancers and the drug remains on the market currently without an integrated diagnostic eligibility test.

BAY 43-9006 (Sorafenib®)

BAY 43-9006 is a RAF kinase inhibitor that also inhibits the VEGFR and PDGFR growth factor receptors. It is thus considered to be an anti-angiogenesis drug. This oral agent was approved in late 2005 by the FDA for the treatment of metastatic renal cell carcinoma (Staehler et al., 2005). Currently, there are no diagnostic tests associated with the selection of this agent and clinical trials for other types of cancer are on-going.

Other Small Molecule Antiangiogenesis Agents (SU5416, Thalidomide [Thalomid®], Lenalidamide [Revlimid®], Endostatin/Angiostatin, and Marimastat)

A variety of small molecule drugs are currently in clinical trials for the treatment of solid tumors that target the establishment and growth of tumor blood vessels (Khalil et al., 2003; Mendel et al., 2000; Thomas and Kantarjian, 2000; Zogakis and Libutti, 2001). Additional compounds that target matrix metalloproteases, such as the drug marimastat, are also considered to be angiogenesis

inhibitors (Brown, 2000; Dell'Eva et al., 2002; Miller et al., 2002). The anti-angiogenesis drug, lenalidomide (Revlimid®) was approved by the FDA in late 2005 for the treatment of myelodysplastic syndrome (List, 2005). To date, none of these compounds has a linked diagnostic test such as tumor microvessel density or the expression of an angiogenesis promoting gene or protein in their clinical development plans.

G3139 (Genasense®)

Another strategy in anticancer therapy is the targeting of chemotherapy resistance by overcoming the antiapoptosis mechanisms of cancer cells. An example of this approach is the novel antisense oligonucleotide G3139, which targets the antiapoptotic gene *bcl-2* (Tamm et al., 2001; Tolcher, 2002). This agent has been the most widely tested antisense therapy and has been mostly focused in hematologic malignancies (Stein et al., 2005).

Bortezomib (Velcade®)

Recently, drugs targeting the proteasome have been developed that are designed to impact downstream pathways regulating angiogenesis, tumor growth, adhesion, and resistance to apoptosis (Adams, 2002; Elliott and Ross, 2001). One of these agents, bortezomib (PS-341), has recently been approved for the treatment of advanced refractory multiple myeloma (Richardson et al., 2003). Bortezomib has shown both preclinical activity in animal studies and biologic activity in early clinical trials involving patients with a variety of solid tumors, but, to date, no trials using this agent alone or in combination with other drugs has progressed to Phase III. Although pharmacogenomic studies of bortezomib use in multiple myeloma have been conducted, to date, no specific pattern of gene expression or other specific test has emerged that could be a guide to the selection of patients for treatment.

2009 UPDATE

In the past year, targeted therapies featuring combinations of drugs and diagnostics tests have become the standard of care in several approved indications. The resulting high level of uptake, coupled with their premium prices, make targeted therapies the leading therapy class in the oncology market with, according to Datamonitor, global sales growing at a near 33% rate and reaching $17.3 billion in 2007. A number of targeted therapy cancer brands have achieved blockbuster sales and have become important sources of revenue for some of the leading pharmaceutical and biotech companies such as Roche, Genentech, and Novartis. Of note has been the introduction of the anti-HER1/2 tyrosine kinase inhibitor lapatinib in HER2 positive breast cancer; additional indications for use of the antibody therapeutic Trastuzumab in earlier stages of the disease; expansion in the use of the tyrosine kinase inhibitors Gleevec and Dasatinib in chronic myelogenous leukemia; additional indications for use of the anti-EGFR antibody therapies Erbitux and Panitumab; expansion of use of the anti-VEGF ligand antibody Bevacizumab in multiple solid tumors including colon, lung, and breast cancers; and the clinical use of small molecule antiangiogenesis inhibitors such as Sunitinib and Sorafenib (Dy and Adjei, 2008).

During the last year, several significant developments in the clinical management of breast and colorectal cancer featured further refinements in the use of integrated diagnostics. In breast cancer, continued improvement in HER2 test accuracy has been achieved with the widespread adoption of the ASCO-CAP HER2 testing guidelines (Wolff et al., 2007). There has also been continued interest in the preliminary results from HER2 testing of circulating tumor cells, suggesting that HER2-negative primary tumors may evolve into HER2-positive metastatic disease more often than previously believed (Ignatiatis et al., 2008). Finally, testing was reported for the 2D6 haplotype in the cytochrome P450 gene as a means of predicting the failure of a subset of estrogen receptor positive breast cancer patients to activate and subsequently respond to the antiestrogen receptor drug, tamoxifen (Jin et al., 2005).

In colorectal cancer, clinical trial-based data indicated that the presence of a mutation in the K-*ras* oncogene predicted resistance to the anti-EGFR antibody therapeutics Erbitux and Panitumab independent of the original tumor EGFR protein expression, gene copy number, or presence of gene sequence mutations (Karapetis et al., 2008). Widespread testing of colorectal cancer for K-*ras* mutations has now commenced, and it is anticipated that this will also be applied to newly diagnosed non-small cell lung cancers as well.

2009 UPDATE REFERENCES

Dy, G.K. and Adjei, A.A. (2008) Systemic cancer therapy: Evolution over the last 60 years. *Cancer* 113(7 Suppl), 1857–1887.

Ignatiadis, M., Kallergi, G., Ntoulia, M. et al. (2008). Prognostic value of the molecular detection of circulating tumor cells using a multimarker reverse transcription-PCR assay for cytokeratin 19, mammaglobin A, and HER2 in early breast cancer. *Clin Cancer Res* 14, 2593–2600.

Jin, Y., Desta, Z., Stearns, V. et al. (2005). CYP2D6 genotype, antidepressant use, and tamoxifen metabolism during adjuvant breast cancer treatment. *J Natl Cancer Inst* 97, 30–39.

Karapetis, C.S., Khambata-Ford, S., Jonker, D.J. et al. (2008). K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N Engl J Med* 359, 1757–1765.

Wolff, A.C., Hammond, M.E., Schwartz, J.N. et al. (2007). American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *Arch Pathol Lab Med* 131, 18.

REFERENCES

Adams, J. (2002). Development of the proteasome inhibitor PS-341. *Oncologist* 7, 9–16.

Advani, A.S. (2005). FLT3 and acute myelogenous leukemia: Biology, clinical significance and therapeutic applications. *Curr Pharm Des* 11, 3449–3457.

Amos, J. and Patnaik, M. (2002). Commercial molecular diagnostics in the U.S.: The Human Genome Project to the clinical laboratory. *Hum Mutat* 19, 324–333.

Baselga, J. (2001). The EGFR as a target for anticancer therapy – focus on cetuximab. *Eur J Cancer* 37(Suppl 4), S16–S22.

Bast, R.C., Jr and Hortobagyi, G.N. (2004). Individualized care for patients with cancer-a work in progress. *N Engl J Med* 351, 2865–2867.

Bast, R.C., Jr, Ravdin, P., Hayes, D.F. et al. (2000). update of recommendations for the use of tumor markers in breast and colorectal cancer: Clinical practice guidelines of the American Society of Clinical Oncology. *J Clin Oncol* 19, 1865–1878.

Berry, W. and Eisenberger, M. (2005). Achieving treatment goals for hormone-refractory prostate cancer with chemotherapy. *Oncologist* 10(Suppl 3), 30–39.

Bieche, I., Onody, P., Laurendeau, I. et al. (1999). Real-time reverse transcription-PCR assay for future management of ERBB2-based clinical applications. *Clin Chem* 45, 1148–1156.

Blagosklonny, M.V. (2005). How avastin potentiates chemotherapeutic srugs: Action and reaction in antiangiogenic therapy. *Cancer Biol Ther* 4, 1307–1310.

Bottles, K. (2001). A revolution in genetics: Changing medicine, changing lives. *Physician Exec* 27, 58–63.

Bross, P.F., Beitz, J., Chen, G. et al. (2001). Approval summary: Gemtuzumab ozogamicin in relapsed acute myeloid leukemia. *Clin Cancer Res* 7, 1490–1496.

Brown, E., Lewis, P.H. and Nocera, J. (2001). In search of the silver bullet. *Fortune* 143, 166–170.

Brown, P.D. (2000). Ongoing trials with matrix metalloproteinase inhibitors. *Expert Opin Investig Drugs* 9, 2167–2177.

Bruckner, H.W., Hrehorovich, V.R. and Sawhney, H.S. (2005). Bevacizumab as treatment for chemotherapy-resistant pancreatic cancer. *Anticancer Res* 25, 3637–3639.

Burke, J.M., Jurcic, J.G. and Scheinberg, D.A. (2002). Radioimmunotherapy for acute leukemia. *Cancer Control* 9, 106–113.

Buzdar, A.U., Robertson, J.F., Eiermann, W. et al. (2006). An overview of the pharmacology and pharmacokinetics of the newer generation aromatase inhibitors anastrozole, letrozole, and exemestane. *Cancer* 95, 2006–2016.

Campiglio, M., Locatelli, A., Olgiati, C. et al. (2004). Inhibition of proliferation and induction of apoptosis in breast cancer cells by the epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor ZD1839 ("Iressa") is independent of EGFR expression level. *J Cell Physiol* 198, 259–268.

Carbone, D.P. (2004). Biomarkers of response to gefitinib in non-small-cell lung cancer. *Nat Clin Pract Oncol* 1, 66–67.

Carswell, C.I., Plosker, G.L. and Wagstaff, A.J. (2001). Daclizumab: A review of its use in the management of organ transplantation. *BioDrugs* 15, 745–773.

Chan, S.K., Gullick, W.J. and Hill, M.E. (2006). Mutations of the epidermal growth factor receptor in non-small cell lung cancer-search and destroy. *Eur J Cancer* 42, 17–23.

Chen, H.X., Gore-Langton, R.E. and Cheson, B.D. (2000). Clinical trials referral resource: Current clinical trials of the anti-VEGF monoclonal antibody bevacizumab. *Oncology* 15, 1023–1026.

Cheson, B.D. (2002). Bexxar (Corixa/GlaxoSmithKline). *Curr Opin Investig Drugs* 3, 165–170.

Chester, K.A. and Hawkins, R.E. (1995). Clinical issues in antibody design. *Trend Biotechnol* 13, 294–300.

Ciocca, D.R. and Elledge, R. (2000). Molecular markers for predicting response to tamoxifen in breast cancer patients. *Endocrine* 13, 1–10.

Coiffier, B. (2002). Rituximab in the treatment of diffuse large B-cell lymphomas. *Semin Oncol* 29(1 Suppl 2), 30–35.

de Gramont, A. and Van Cutsem, E. (2005). Investigating the potential of bevacizumab in other indications: Metastatic renal cell, non-small cell lung, pancreatic and breast cancer. *Oncology* 69(Suppl 3), 46–56.

Dell'Eva, R., Pfeffer, U., Indraccolo, S. et al. (2002). Inhibition of tumor angiogenesis by angiostatin: From recombinant protein to gene therapy. *Endothelium* 9, 3–10.

Dillman, R.O. (2001). Monoclonal antibody therapy for lymphoma: An update. *Cancer Pract* 9, 71–80.

Dillman, R.O. (2002). Radiolabeled anti-CD20 monoclonal antibodies for the treatment of B-cell lymphoma. *J Clin Oncol* 20, 3545–3557.

Dressman, M.A., Baras, A., Malinowski, R. *et al.* (2003). Gene expression profiling detects gene amplification and differentiates tumor types in breast cancer. *Cancer Res* 63, 2194–2199.

Druker, B.J. (2003). Imatinib alone and in combination for chronic myeloid leukemia. *Semin Hematol* 40, 50–58.

Dudek, A.Z., Kmak, K.L., Koopmeiners, J. et al. (2006). Skin rash and bronchoalveolar histology correlates with clinical benefit in patients treated with gefitinib as a therapy for previously treated advanced or metastatic non-small cell lung cancer. *Lung Cancer*, 51, 89–96.

Dumont, F.J. (2002). CAMPATH (alemtuzumab) for the treatment of chronic lymphocytic leukemia and beyond. *Expert Rev Anticancer Ther* 2, 23–35.

Elliott, P.J. and Ross, J.S. (2001). The proteasome: A new target for novel drug therapies. *Am J Clin Pathol* 116, 637–646.

Evans, W.E. and McLeod, H.L. (2003). Pharmacogenomics—drugdisposition, drug targets, and side effects. *N Engl J Med* 348, 538–549.

Fang, J., Chen, S.J., Tong, J.H. et al. (2002). Treatment of acute promyelocytic leukemia with ATRA and As2O3: A model of molecular target-based cancer therapy. *Cancer Biol Ther* 1, 614–620.

Ferrara, N. (2005). VEGF as a therapeutic target in cancer. *Oncology* 69(Suppl 3), 11–16.

Fornier, M.N., Seidman, A.D., Schwartz, M.K. et al. (2005). Serum HER2 extracellular domain in metastatic breast cancer patients treated with weekly trastuzumab and paclitaxel: Association with HER2 status bimmunohistochemistry and fluorescence in situ hybridization and with response rate. *Ann Oncol* 16, 234–239.

Fracasso, G., Bellisola, G., Cingarlini, S. et al. (2002). Anti-tumor effects of toxins targeted to the prostate specific membrane antigen. *Prostate* 53, 9–23.

Freeman, L.M., Krynyckyi, B.R., Li, Y. et al. (2002). National Prostascint study group. The role of $^{(111)}$In Capromab Pendetide (ProstaScintR) immunoscintigraphy in the management of prostate cancer. *Q J Nucl Med* 46, 131–137.

Gilliland, D.G. and Griffin, J.D. (2002). The roles of FLT3 in hematopoiesis and leukemia. *Blood* 100, 1532–1542.

Gleissner, B. and Thiel, E. (2001). Detection of immunoglobulin heavy chain gene rearrangements in hematologic malignancies. *Expert Rev Mol Diagn* 1, 191–200.

Goldenberg, D.M. (2002). Targeted therapy of cancer with radiolabeled antibodies. *J Nucl Med* 43, 693–713.

Goldman, J.M. and Melo, J.V. (2003). Chronic myeloid leukemia—advances in biology and new approaches to treatment. *N Engl J Med* 349, 1451–1464.

Gordon, L.I., Witzig, T.E., Wiseman, G.A. et al. (2002). Yttrium 90 ibritumomab tiuxetan radioimmunotherapy for relapsed or refractory low-grade non-Hodgkin's lymphoma. *Semin Oncol* 29(1 suppl 2), 87–92.

Grillo-Lopez, A.J. (2002). AntiCD20 mAbs: Modifying therapeutic strategies and outcomes in the treatment of lymphoma patients. *Expert Rev Anticancer Ther* 2, 323–329.

Grillo-Lopez, A.J., Hedrick, E., Rashford, M. et al. (2002). Rituximab: Ongoing and future clinical development. *Semin Oncol* 29(1 suppl 2), 105–112.

Grimwade, D. and Lo Coco, F. (2002). Acute promyelocytic leukemia: A model for the role of molecular diagnosis and residual disease monitoring in directing treatment approach in acute myeloid leukemia. *Leukemia* 16, 1959–1973.

Haller, D.G. (2001). Update of clinical trials with edrecolomab: A monoclonal antibody therapy for colorectal cancer. *Semin Oncol* 28(1 suppl 1), 25–30.

Hammond, M.E., Fitzgibbons, P.L., Compton, C.C. et al. (2000). College of American Pathologists Conference XXXV: Solid tumor prognostic factors-which, how and so what? Summary document and recommendations for implementation. Cancer Committee and Conference Participants. *Arch Pathol Lab Med* 124, 958–965.

Harari, P.M. and Huang, S. (2006). Radiation combined With EGFR signal inhibitors: Head and neck cancer focus. *Semin Radiat Oncol* 16, 38–44.

Hayes, D.F. and Thor, A.D. (2002). c-erbB-2 in breast cancer: Development of a clinically useful marker. *Semin Oncol* 29, 231–245.

Heinicke, T., Wardelmann, E., Sauerbruch, T. et al. (2005). Very early detection of response to imatinib mesylate therapy of gastrointestinal stromal tumours using 18fluoro deoxyglucose-positron emission tomography. *Anticancer Res* 25, 4591–4594.

Hemminki, A. (2002). From molecular changes to customised therapy. *Eur J Cancer* 38, 333–338.

Herbst, R.S. and Langer, C.J. (2002). Epidermal growth factor receptors as a target for cancer treatment: The emerging role of IMC-C225 in the treatment of lung and head and neck cancers. *Semin Oncol* 29(1 suppl 4), 27–36.

Hortobagyi, G.N. (2001). Overview of treatment results with trastuzumab (Herceptin) in metastatic breast cancer. *Semin Oncol* 28, 43–47.

Hurwitz, H. and Kabbinavar, F. (2005). Bevacizumab combined with standard fluoropyrimidine-based chemotherapy regimens to treat colorectal cancer. *Oncology* 69(Suppl 3), 17–24.

Ibrahim, N.K. and Hortobagyi, G.N. (1999). The evolving role of specific estrogen receptor modulators (SERMs). *Surg Oncol* 8, 103–123.

Isaacs, J.D. (2001). From bench to bedside: Discovering rules for antibody design, and improving serotherapy with monoclonal antibodies. *Rheumatology* 40, 724–738.

Israeli, R.S., Powell, C.T., Corr, J.G. et al. (1994). Expression of the prostate-specific membrane antigen. *Cancer Res* 54, 1807–1811.

Jones, P.T., Dear, P.H., Foote, J. et al. (1986). Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature* 321, 522–525.

Jordan, V.C. (2003). Antiestrogens and selective estrogen receptor modulators as multifunctional medicines 1 Receptor interactions. *J Med Chem* 46, 883–908.

Jordan, V.C. (2003). Antiestrogens and selective estrogen receptor modulators as multifunctional medicines. 2. Clinical considerations and new agents. *J Med Chem* 46, 1081–1111.

Jordan, V.C. (2003). Tamoxifen: A most unlikely pioneering medicine. *Nat Rev Drug Discov* 2, 205–213.

Kelly, L.M., Yu, J.C., Boulton, C.L. et al. (2002). CT53518, a novel selective FLT3 antagonist for the treatment of acute myelogenous leukemia (AML). *Cancer Cell* 1, 421–432.

Khalil, M.Y., Grandis, J.R. and Shin, D.M. (2003). Targeting epidermal growth factor receptor: Novel therapeutics in the management of cancer. *Expert Rev Anticancer Ther* 3, 367–380.

Kirchner, E.M., Gerhards, R. and Voigtmann, R. (2002). Sequential immunochemotherapy and edrecolomab in the adjuvant therapy of breast cancer: Reduction of 17–1A-positive disseminated tumour cells. *Ann Oncol* 13, 1044–1048.

Kiyoi, H. and Naoe, T. (2005). FLT3 in human hematologic malignancies. *Leuk Lymph* 43, 1541–1547.

Kobayashi, S., Boggon, T.J., Dayaram, T. et al. (2005). EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N Engl J Med* 352, 786–792.

Krasner, C. and Joyce, R.M. (2001). Zevalin: 90yttrium labeled anti-CD20 (ibritumomab tiuxetan), a new treatment for non-Hodgkin's lymphoma. *Curr Pharm Biotechnol* 2, 341–349.

Kreitman, R.J., Chaudhary, V.K., Kozak, R.W. et al. (1992). Recombinant toxins containing the variable domains of the anti-Tac monoclonal antibody to the interleukin-2 receptor kill malignant cells from patients with chronic lymphocytic leukemia. *Blood* 80, 2344–2352.

Kuus-Reichel, K., Grauer, L.S., Karavodin, L.M. et al. (1994). Will immunogenicity limit the use, efficacy, and future development of therapeutic monoclonal antibodies?. *Clin Diagn Lab Immunol* 1, 365–372.

Lange, T., Park, B., Willis, S.G. et al. (2005). BCR-ABL kinase domain mutations in chronic myeloid leukemia: Not quite enough to cause resistance to imatinib therapy?. *Cell Cycle* 4, 1761–1766.

Larson, R.A., Boogaerts, M., Estey, E. et al. (2002). Antibody-targeted chemotherapy of older patients with acute myeloid leukemia in first relapse using Mylotarg (gemtuzumab ozogamicin). *Leukemia* 16, 1627–1636.

Lemonick, M.D. and Park, A. (2001). New hope for cancer. *Time* 157, 62–69.

Leonard, D.S., Hill, A.D., Kelly, L. et al. (2002). Anti-human epidermal growth factor receptor 2 monoclonal antibody therapy for breast cancer. *Br J Surg* 89, 262–271.

Ligibel, J.A. and Winer, E.P. (2002). Trastuzumab/chemotherapy combinations in metastatic breast cancer. *Semin Oncol* 29, 38–43.

Linenberger, M.L., Maloney, D.G. and Bernstein, I.D. (2002). Antibody-directed therapies for hematological malignancies. *Trends Mol Med* 8, 69–76.

List, A.F. (2005). Emerging data on IMiDs in the treatment of myelodysplastic syndromes (MDS). *Semin Oncol* 32, S31–S35.

Liu, H., Moy, P., Kim, S. et al. (1997). Monoclonal antibodies to the extracellular domain of prostate-specific membrane antigen also react with tumor vascular endothelium. *Cancer Res* 57, 3629–3634.

Locker, G.Y. (1998). Hormonal therapy of breast cancer. *Cancer Treat Rev* 24, 221–240.

Lynch, T.J., Bell, D.W., Sordella, R. et al. (2004). Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to Gefitinib. *N Engl J Med* 350, 2129–2139.

Masood, S. and Bui, M.M. (2002). Prognostic and predictive value of HER2/neu oncogene in breast cancer. *Microsc Res Tech* 59, 102–108.

Mauro, M.J. and Druker, B.J. (2001). STI571: Targeting BCR-ABL as therapy for CML. *Oncologist* 6, 233–238.

McKeage, K. and Perry, C.M. (2002). Trastuzumab: A review of its use in the treatment of metastatic breast cancer overexpressing HER2. *Drugs* 62, 209–243.

Mendel, D.B., Laird, A.D., Smolich, B.D. et al. (2000). Development of SU5416, a selective small molecule inhibitor of VEGF receptor tyrosine kinase activity, as an anti-angiogenesis agent. *Anticancer Drug Des* 15, 29–41.

Mendelsohn, J. and Baselga, J. (2000). The EGF receptor family as targets for cancer therapy. *Oncogene* 19, 6550–6565.

Merluzzi, S., Figini, M., Colombatti, A. et al. (2000). Humanized antibodies as potential drugs for therapeutic use. *Adv Clin Path* 4, 77–85.

Milenic, D.E. (2002). Monoclonal antibody-based therapy strategies: Providing options for the cancer patient. *Curr Pharm Des* 8, 1749–1764.

Miller, K.D., Gradishar, W., Schuchter, L. et al. (2002). A randomized phase II pilot trial of adjuvant marimastat in patients with early-stage breast cancer. *Ann Oncol* 13, 1220–1224.

Miller, W.R., Anderson, T.J. and Dixon, J.M. (2002). Anti-tumor effects of letrozole. *Cancer Invest* 20, 15–21.

Milowsky, M.I., Nanus, D.M., Kostakoglu, L. et al. (2004). Phase I Trial of Yttrium-90-Labeled Anti-Prostate-Specific Membrane Antigen Monoclonal Antibody J591 for Androgen-Independent Prostate Cancer. *J Clin Oncol* 22, 2522–2531.

Milowsky, M.I., Galsky, J., Lewin, C.P. et al. (2005). Phase 1/2 dose escalation trial of the prostate-specific membrane antigen (PSMA)-targeted immunoconjugate MLN2704 in patients with progressive metastatic castration-resistant prostate cancer. *Proceed ASCO* 24, 4615.

Moore, M.J. (2005). Brief communication: A new combination in the treatment of advanced pancreatic cancer. *Semin Oncol* 32, 5–6.

Nabhan, C. and Tallman, M.S. (2002). Early phase I/II trials with gemtuzumab ozogamicin (Mylotarg®) in acute myeloid leukemia. *Clin Lymph* 2(Suppl 1), S19–S23.

Nielsen, U.B. and Marks, J.D. (2000). Internalizing antibodies and targeted cancer therapy: Direct selection from phage display libraries. *PSTT* 3, 282–291.

O'Brien, S.G., Guilhot, F., Larson, R.A. et al. (2003). Imatinib compared with interferon and low-dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukemia. *N Engl J Med* 348, 994–1004.

O'Dwyer, M.E. and Druker, B.J. (2001). Chronic myelogenous leukaemia—new therapeutic principles. *J Intern Med* 250, 3–9.

O'hare, T., Corbin, A.S. and Druker, B.J. (2006). Targeted CML therapy: Controlling drug resistance, seeking cure. *Curr Opin Genet Dev* 16, 92–99.

Osborne, C.K. (1998). Steroid hormone receptors in breast cancer management. *Breast Cancer Res Treat* 51, 227–238.

Paez, J.G., Janne, P.A., Lee, J.C. et al. (2004). EGFR mutations in lung cancer: Correlation with clinical response to gefitinib therapy. *Science* 304, 1497–5000.

Paik, S., Tan-chui, E., Bryan, J. et al. (2002). Successful quality assurance program for HER-2 testing in the NSAPB trial for Herceptin. *Breast Cancer Res Treat* 76(Suppl 1), S31.

Paik, S., Shak, S., Tang, G. et al. (2004). A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 351, 2817–2826.

Paik, S., Shak, S., Tang, G. et al. (2005). Expression of the 21 genes in the recurrence score assay and tamoxifen clinical benefit in the NSABP study B-14 of node negative, estrogen receptor positive breast cancer. *Proc ASCO* 24, 510.

Pangalis, G.A., Dimopoulou, M.N., Angelopoulou, M.K. et al. (2001). Campath-1H (anti-CD52) monoclonal antibody therapy in lymphoproliferative disorders. *Med Oncol* 18, 99–107.

Parmar, S. and Tallman, M.S. (2003). Acute promyelocytic leukaemia: A review. *Expert Opin Pharmacother* 4, 1379–1392.

Pawlowski, V., Revillion, F., Hornez, L. et al. (2000). A real-time one-step reverse transcriptase-polymerase chain reaction method to quantify c-erbB-2 expression in human breast cancer. *Cancer Detect Prev* 24, 212–223.

Pimm, M.V. (1994). Possible consequences of human antibody responses on the biodistribution of fragments of human, humanized or chimeric monoclonal antibodies: A note of caution. *Life Sci* 55, PL45–PL49.

Pusztai, L., Ayers, M., Stec, J. et al. (2003). Gene expression profiles obtained from fine-needle aspirations of breast cancer reliably identify routine prognostic markers and reveal large-scale molecular differences between estrogen-negative and estrogen-positive tumors. *Clin Cancer Res* 9, 2406–2415.

Ranson, M. (2002). ZD1839 (Iressa): For more than just non-small cell lung cancer. *Oncologist* 7(Suppl 4), 16–24.

Reff, M.E. and Heard, C. (2001). A review of modifications to recombinant antibodies: Attempt to increase efficacy in oncology applications. *Crit Rev Oncol Hematol* 40, 25–35.

Reff, M.E., Hariharan, K. and Braslawsky, G. (2002). Future of monoclonal antibodies in the treatment of hematologic malignancies. *Cancer Control* 9, 152–166.

Reichert, J.M. (2002). Therapeutic monoclonal antibodies: Trends in development and approval in the US. *Curr Opin Mol Ther* 4, 110–118.

Reilly, R.M., Sandhu, J., Alvarez-Diez, T.M. et al. (1995). Problems of delivery of monoclonal antibodies pharmaceutical and pharmacokinetic solutions. *Clin Pharmacokinet* 28, 126–142.

Reynolds, T. (2002). Biotech firm faces challenges from FDA, falling stock prices. *J Natl Cancer Inst* 94, 326–328.

Richardson, P.G., Barlogie, B., Berenson, J. et al. (2003). A phase 2 study of bortezomib in relapsed, refractory myeloma. *N Engl J Med* 348, 2609–2617.

Rosen, L.S. (2002). Clinical experience with angiogenesis signaling inhibitors: Focus on vascular endothelial growth factor (VEGF) blockers. *Cancer Control* 9(2 suppl), 36–44.

Ross, J.S. (2005). Genomic microarrays in cancer molecular diagnostics: Just biomarker discovery tools or future bedside clinical assays?. *Expert Rev Mol Diagn* 5, 837–838.

Ross, J.S., Fletcher, J.A., Linette, G.P. et al. (2003). The Her-2/neu gene and protein in breast cancer 2003: Biomarker and target of therapy. *Oncologist* 8, 307–325.

Ross, J.S., Gray, K., Gray, G.S. et al. (2003). Anticancer antibodies. *Am J Clin Pathol* 119, 472–485.

Ross, J.S., Sheehan, C.E., Fisher, H.A. et al. (2003). Correlation of primary tumor prostate-specific membrane antigen expression with disease recurrence in prostate cancer. *Clin Cancer Res* 9, 6357–6362.

Ross, J.S., Gray, K., Webb, I.A. et al. (2005). Antibody based therapeutics: Focus on prostate cancer. *Cancer Metastasis Rev* 24, 517–533.

Rubnitz, J.E. and Pui, C.H. (1999). Molecular diagnostics in the treatment of leukemia. *Curr Opin Hematol*, 229–235.

Saltz, L. (2005). Epidermal growth factor receptor-negative colorectal cancer: Is there truly such an entity?. *Clin Colorectal Cancer* 5(Suppl 2), S98–S100.

Sawyers, C.L. (2002). Finding the next Gleevec: FLT3 targeted kinase inhibitor therapy for acute myeloid leukemia. *Cancer Cell* 1, 413–415.

Schiller, J.H. (2003). New directions for ZD1839 in the treatment of solid tumors. *Semin Oncol* 30(1 Suppl 1), 49–55.

Schnitt, S.J. and Jacobs, T.W. (2001). Current status of HER2 testing: Caught between a rock and a hard place. *Am J Clin Pathol* 116, 806–810.

Schwartzberg, L.S. (2001). Clinical experience with edrecolomab: A monoclonal antibody therapy for colorectal carcinoma. *Crit Rev Oncol Hematol* 40, 17–24.

Shawver, L.K., Slamon, D. and Ullrich, A. (2002). Smart drugs: Tyrosine kinase inhibitors in cancer therapy. *Cancer Cell* 1, 117–123.

Sievers, E.L. and Linenberger, M. (2001). Mylotarg: Antibody-targeted chemotherapy comes of age. *Curr Opin Oncol* 13, 522–527.

Silvestri, G.A. and Rivera, M.P. (2005). Targeted therapy for the treatment of advanced non-small cell lung cancer: A review of the epidermal growth factor receptor antagonists. *Chest* 128, 3975–3984.

Slamon, D.J., Leyland-Jones, B., Shak, S. et al. (2001). Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* 344, 783–792.

Smith, J. (2005). Erlotinib: Small-molecule targeted therapy in the treatment of non-small-cell lung cancer. *Clin Ther* 27, 1513–1534.

Stadler, W.M. (2005). Targeted agents for the treatment of advanced renal cell carcinoma. *Cancer* 104, 2323–2333.

Stadtmauer, E.A. (2002). Trials with gemtuzumab ozogamicin (Mylotarg®) combined with chemotherapy regimens in acute myeloid leukemia. *Clin Lymphoma* 2(Suppl 1), S24–S28.

Staehler, M., Rohrmann, K., Haseke, N. et al. (2005). Targeted agents for the treatment of advanced renal cell carcinoma. *Curr Drug Targets* 6, 835–846.

Stein, C.A., Benimetskaya, L. and Mani, S. (2005). Antisense strategies for oncogene inactivation. *Semin Oncol* 32, 563–572.

Stevenson, G.T., Anderson, V.A. and Leong, W.S. (2002). Engineered antibody for treating lymphoma. *Recent Results Cancer Res* 159, 104–112.

Tamm, I., Dorken, B. and Hartmann, G. (2001). Antisense therapy in oncology: New hope for an old idea?. *Lancet* 358, 489–497.

Tanner, M., Gancberg, D., Di Leo, A. et al. (2000). Chromogenic in situ hybridization: A practical alternative for fluorescence in situ hybridization to detect HER-2/neu oncogene amplification in archival breast cancer samples. *Am J Pathol* 157, 1467–1472.

Thomas, D.A. and Kantarjian, H.M. (2000). Current role of thalidomide in cancer treatment. *Curr Opin Oncol* 12, 564–573.

Tolcher, A.W. (2002). Regulators of apoptosis as anticancer targets. *Hematol Oncol Clin North Am* 16, 1255–1267.

Tubbs, R.R., Pettay, J.D., Roche, P.C. et al. (2001). Discrepancies in clinical laboratory testing of eligibility for trastuzumab therapy: apparent immunohistochemical false-positives do not get the message. *J Clin Oncol* 19, 2714–2721.

Twombly, R. (2005). Failing survival advantage in crucial trial, future of Iressa is in jeopardy. *J Natl Cancer Inst* 97, 249–250.

Verweij, J., van Oosterom, A., Blay, J.Y. et al. (2003). Imatinib mesylate (STI-571 Glivec, Gleevec) is an active agent for gastrointestinal stromal tumours, but does not yield responses in other soft-tissue sarcomas that are unselected for a molecular target Results from an EORTC Soft Tissue and Bone Sarcoma Group phase II study. *Eur J Cancer* 39, 2006–2011.

von Mehren, M. (2003). Gastrointestinal stromal tumors: A paradigm for molecularly targeted therapy. *Cancer Invest* 21, 553–563.

Wagner, H.N., Jr, Wiseman, G.A., Marcus, C.S. et al. (2002). Administration guidelines for radioimmunotherapy of non-Hodgkin's lymphoma with $^{(90)}$Y-labeled anti CD20 monoclonal antibody. *J Nucl Med* 43, 267–272.

Wakelee, H. and Belani, C.P. (2005). Optimizing first-line treatment options for patients with advanced NSCLC. *Oncologist* 10(Suppl 3), 1–10.

Wang, S., Saboorian, M.H., Frenkel, E. et al. (2000). Laboratory assessment of the status of Her-2/neu protein and oncogene in breast cancer specimens: Comparison of immunohistochemistry assay with fluorescence in situ hybridization assays. *J Clin Pathol* 53, 374–381.

Wang, S., Saboorian, M.H., Frenkel, E.P. et al. (2001). Assessment of HER-2/neu status in breast cancer. Automated Cellular Imaging System (ACIS)-assisted quantitation of immunohistochemical assay achieves high accuracy in comparison with fluorescence in situ hybridization assay as the standard. *Am J Clin Pathol* 116, 495–503.

Watkins, N.A. and Ouwehand, W.H. (2000). Introduction to antibody engineering and phage display. *Vox Sang* 78, 72–79.

Weinshilboum, R. (2003). Inheritance and drug response. *N Engl J Med* 348, 529–537.

Wilbur, D.C., Willis, J., Mooney, R.A. et al. (1992). Estrogen and progesterone detection in archival formalin-fixed paraffin embedded tissue from breast carcinoma: A comparison of immunocytochemistry with dextran coated charcoal assay. *Mod Pathol* 5, 79–84.

Winter, G. and Harris, W.J. (1993). Humanized antibodies. *Immunol Today* 14, 243–246.

Wiseman, G.A., Gordon, L.I., Multani, P.S. et al. (2002). Ibritumomab tiuxetan radioimmunotherapy for patients with relapsed or refractory non-Hodgkin lymphoma and mild thrombocytopenia: A phase II multicenter trial. *Blood* 99, 4336–4342.

Wong, S.F. (2005). Cetuximab: An epidermal growth factor receptor monoclonal antibody for the treatment of colorectal cancer. *Clin Ther* 27, 684–694.

Zelenetz, A.D. (2003). A clinical and scientific overview of tositumomab and iodine I 131 tositumomab. *Semin Oncol* 30(2 suppl 4), 22–30.

Zhao, J., Wu, R., Au, A. et al. (2002). Determination of HER2 gene amplification by chromogenic in situ hybridization (CISH) in archival breast carcinoma. *Mod Pathol* 15, 657–665.

Zogakis, T.G. and Libutti, S.K. (2001). General aspects of anti-angiogenesis and cancer therapy. *Expert Opin Biol Ther* 1, 253–275.

This page intentionally left blank

Clinical (Inflammatory Disease)

Section 6

43. Genomics in the Evaluation and Management of Rheumatoid Arthritis

44. Genomic Evaluation of Multiple Sclerosis

45. Genomic Assessment of Inflammatory Bowel Disease

46. Asthma Genomics

47. Genomic Evaluation and Management of Chronic Obstructive Pulmonary Disease

48. Genetics and Genomics of Interstitial Lung Disease

49. Peptic Ulcer Disease

This page intentionally left blank

CHAPTER

Genomics in the Evaluation and Management of Rheumatoid Arthritis

Robert M. Plenge and Michael E. Weinblatt

INTRODUCTION

Rheumatoid arthritis (RA) is a systemic, chronic inflammatory disorder whose root cause is unclear. The clinical hallmark of RA is an inflammatory arthritis with a predilection for specific diarthrodial (freely movable) joints. It is the most common form of inflammatory arthritis, with an estimated prevalence of up to ~1% in the adult population. Females are at greater risk than males for developing the disease, with a female:male ratio of ~2.5:1. While the disease can occur at any age, the peak age on onset is in the 40s, with an increasing incident with age.

As with many complex diseases – those influenced by multiple genes and environmental exposures – there is substantial clinical heterogeneity. The clinical features of new-onset RA are highlighted in Table 43.1. With longstanding disease, articular erosions and joint deformities occur. Autoantibodies (RF and CCP) have important diagnostic and prognostic features, and have proven very useful in clinical management of RA. Most efforts aimed at understanding the molecular basis of RA have focused of genetic studies of disease susceptibility. To date, there are only two genes that have been convincingly demonstrated to influence risk of RA (*PTPN22* and *HLA-DRB1*), although only a small fraction of the genome has been adequately interrogated using available molecular genetic techniques. Other genomic technologies such as large-scale expression and proteomic profiling in RA are less mature, but offer promise in understanding disease etiology.

"Genomic medicine" – translating genomic information into prediction of disease susceptibility, characterization of gene–environment interactions, identification of new therapeutic targets, and development of novel gene-based diagnostics – has had very little impact thus for on the clinical management of patients with RA. The association of susceptibility to *HLA-DRB1*

TABLE 43.1 Common clinical features of new-onset rheumatoid arthritis

Symptoms

- Joint swelling
- Pain/stiffness (commonly in morning and lasting >1 h)
- Fatigue
- Malaise

Articular characteristics

- Palpation tenderness
- Synovial thickening
- Effusion
- Erythema

Distribution

- Symmetrical
- Distal (e.g., hands and feet) more commonly than proximal (e.g., spine)
- PIP, MCP/MTP, wrist/ankle more commonly than elbow/knee, shoulder/hip

PIP = proximal interphalangeal joint, MCP = metacarpophalangeal joint, MTP = metatarsophalangeal joint.

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

alleles has been known since the 1970s, yet has not translated into novel diagnostics or therapeutic treatments. The most remarkable advancement in the clinical management of RA, which emerged from intense basic science research, is the development of inhibitors of an important cytokine, tumor necrosis factor-alpha (TNFα). Advances in genomic medicine hope to parallel the success of TNFα-inhibitors, thus leading to improved patient care for this debilitating disease.

PREDISPOSITION

Despite decades of research, the root cause of RA is unclear (Firestein, 2003). Genes and environment together contribute to development of RA, although only two genes (*PTPN22* and *HLA-DRB1*), and one environmental factor (smoking), have been associated with RA susceptibility across multiple, independent studies. Antibodies directed against cyclic citrullinated peptides (anti-CCP antibodies) have emerged as a specific marker for RA. That anti-CCP antibodies predate the diagnosis of RA by years suggests that these autoantibodies are pathogenic rather than simply a marker of chronic inflammation. Together, these risk factors suggest a hypothetical model that an environment trigger (e.g., smoking) invokes a generalized inflammatory response in genetically susceptible hosts, which leads to the formation of autoantibodies and eventually RA.

Genetic Basis of RA

The genetic contribution to RA susceptibility in humans has been demonstrated through twin studies (MacGregor et al., 2000), family studies (Bali et al., 1999), and genome-wide linkage scans (Amos et al., 2006; Cornelis et al., 1998; Etzel et al., 2006; Jawaheer et al., 2001, 2003; MacKay et al., 2002; Shiozawa et al., 1998). Heritability refers to the amount of phenotypic variation due to additive genetic factors, rather than common environmental factors, stochastic variation, gene–environment interactions, and gene–gene interactions. One such study demonstrated that approximately 60% of disease variability is inherited (MacGregor et al., 2000). Another measure of genetic contribution to disease activity is to compare prevalence of disease in family members compared to the general population. Whereas the population risk of RA is ~1%, the monozygotic (MZ) twin of a patient with RA has a risk of ~15% (Aho et al., 1986; Jarvinen and Aho, 1994; Silman et al., 1993). Moreover, the relative risk (RR) to the sibling of a proband with RA is ~5 for RA (Deighton et al., 1989; Hasstedt et al., 1994; Wolfe et al., 1988), although the number varies depending on the population studied (Jawaheer et al., 2001).

The MHC-Region and HLA-DRB1 Susceptibility Alleles

The major histocompatibility complex (MHC) region spans ~3.6 megabases (Mb) on the short arm of human chromosome 6, and contains hundreds of genes, including many involved in immune function (Horton et al., 2004; Stewart et al., 2004). It has been estimated that the MHC-region of the human genome accounts for approximately one-third of the overall genetic component of RA risk (Deighton et al., 1989; Rigby et al., 1991). Genome-wide linkage scans using both microsatellite (Cornelis et al., 1998; Jawaheer et al., 2001; MacKay et al., 2002; Shiozawa et al., 1998) and single nucleotide polymorphism (SNP) markers (Amos et al., 2006) have identified consistently this region as important in RA pathogenesis. These genome-wide scans have demonstrated that the MHC region has the largest genetic contribution in RA, and the relative contribution of MHC genes (λMHC) was found to be ~1.75 (Cornelis et al., 1998; Jawaheer et al., 2001).

Much, but probably not all, of the risk attributable to the MHC-region is associated with alleles within the *HLA-DRB1* gene. An association between RA and the Class II HLA proteins was first noted in the 1970s, when the mixed lymphocyte culture (MCL) type Dw4 (related to the serological subtype DR4) was observed to be more common among patients with RA compared to controls (Stastny, 1978; Stastny and Fink, 1977). Subsequently, investigation of the molecular diversity of Class II proteins (subunits of HLA-DR, -DQ and -DP) localized the serological Dw4 subtype to the *HLA-DRB1* gene (Gregersen et al., 1986a, b). When the susceptible DR subtypes were considered as a group, Gregersen et al., noted a shared amino acid (a.a.) sequence at positions 70–74 of the HLA-DRB1 protein (Gregersen et al., 1987). These residues are important in peptide binding, and thus it was hypothesized that RA-associated alleles bind specific peptides, which in turn facilitates the development of auto-reactive T cells. These alleles are now known collectively as "shared epitope" alleles due to the related sequence composition in the third hypervariable region (Table 43.2): the susceptibility alleles result in missense a.a. changes, where the shared susceptibility a.a. motif is ^{70}Q/R-K/R-R-A-A^{74}.

The *HLA-DRB1* gene encodes for a protein that is part of MHC Class II molecules. These molecules, heterodimers of alpha and beta proteins, are found on professional antigen-presenting cells (APCs), and display peptides derived from extracellular proteins to CD4+ T cells (T "helper" cells). The genes that encode the Class II molecules are found in three subregions (DR, DQ, and DP) spanning ~1 Mb within the MHC region. Within the DR subregion, one alpha- and three beta-chain genes have been described; the a-chain gene and two beta-chain genes, DRB1 and DRB2, are clearly expressed. The DQ subregion contains two sets of alpha and beta-chain genes, DX and DQ-alpha and beta. With the exception of DR-alpha, all of the expressed genes display considerable allelic diversity. Classification of HLA-DRB1 molecules includes serological nomenclature (e.g., DR1, DR4, and DR10), MCL nomenclature (e.g., Dw4) and DNA-sequence based nomenclature (see below).

Since the initial observation, a large number of population studies have confirmed the association between RA and allelic variants at *HLA-DRB1* (Ollier and Thomson, 1992). At the level of DNA, the most common (>5% population frequency) *HLA-DRB1* shared epitope susceptibility alleles include *0101,

TABLE 43.2 RA HLA-DRB1 "shared epitope" alleles

DRB1 alleles	Low-resolution	a.a (location)				
		70	71	72	73	74
		Q	R	R	A	A
DRB1*0101	DR1	Q	R	R	A	A
DRB1*0102	DR1	–	–	–	–	–
DRB1*0103	DR1	D	E	–	–	–
DRB1*03	DR3	–	K	–	G	R
DRB1*0401	DR4	–	K	–	–	–
DRB1*0402	DR4	D	E	–	–	–
DRB1*0403	DR4	–	–	–	–	E
DRB1*0404	DR4	–	–	–	–	–
DRB1*0405	DR4	–	–	–	–	–
DRB1*0407	DR4	–	–	–	–	E
DRB1*0408	DR4	–	–	–	–	–
DRB1*0411	DR4	–	–	–	–	E
DRB1*07	DR7	D	–	–	G	Q
DRB1*08	DR8	D	–	–	–	L
DRB1*0901	DR9	R	–	–	–	E
DRB1*1001	DR10	R	–	–	–	–
DRB1*1101	DR11	D	–	–	–	–
DRB1*1102	DR11	D	E	–	–	–
DRB1*1103	DR11	D	E	–	–	–
DRB1*1104	DR11	D	–	–	–	–
DRB1*12	DR12	D	–	–	–	–
DRB1*1301	DR13	D	E	–	–	–
DRB1*1302	DR13	D	E	–	–	–
DRB1*1303	DR13	D	K	–	–	–
DRB1*1323	DR13	D	E	–	–	–
DRB1*1401	DR14	R	–	–	–	–
DRB1*1402	DR14	–	–	–	–	–
DRB1*1404	DR14	R	–	–	–	E
DRB1*15	DR2	–	A	–	–	–
DRB1*16	DR16	D	–	–	–	–

Genes associated with RA HLA-DRB1"shared epitope" alleles are classified by amino acid (a.a.) sequence at positions 70–74. The consensus a.a. sequence (QRRAA) is shown at the top; the identical a.a. is indicated by a dash (-) and variable a.a. indicated by appropriate nomenclature. Alleles in bold are associated with RA susceptibility.

*0401, and *0404 in individuals of European ancestry, and *0405 and *0901 in individuals of Asian ancestry; less common shared epitope alleles include *0102, *0104, *0408, *0413, *0416, *1001, and *1402. Of note, the *0901 allele observed among Asian populations does not strictly conform to the SE a.a. sequence motif (^{70}R-R-R-A-E^{74}, see Table 43.3), and the classic SE alleles may not contribute to risk in African-American and Hispanic-American RA populations (McDaniel et al., 1995; Teller et al., 1996). Thus, additional exploration of the molecular basis of *HLA-DRB1* susceptibility alleles is needed in the future.

While *HLA-DRB1* susceptibility alleles are often considered as a group, the strength of the genetic association to RA susceptibility differs across the *DRB1* alleles. There are at least two classes of *HLA-DRB1* risk alleles (high and moderate). In general, DRB1*0401 and *0405 alleles exhibit a high level or risk, with a RR of approximately 3. The DRB1*0101, *0404, *0901, and *1001 alleles exhibit a more moderate RR in the range of 1.5.

It is becoming increasingly clear that *HLA-DRB1* shared epitope alleles only influence the development of seropositive RA, and more specifically anti-CCP+ RA (see below for discussion on autoantibodies) (Huizinga et al., 2005; Irigoyen et al., 2005). Collectively, the shared epitope alleles have an odds ratio (OR) of over 5 if CCP+ RA patients are compared to matched healthy controls. Because these alleles are quite common in the general population (collectively, allele frequency ~40% in individuals of European ancestry), the attributable risk for SE alleles is quite high.

Several investigators have proposed a refined classification of shared epitope alleles, as this hypothesis alone cannot explain all of the genetic risk attributable to the *HLA-DRB1* locus (de Vries et al., 2002; Gao et al., 1991; Michou et al., 2006; Zanelli et al., 1998). No consensus has emerged, however. Some of these studies suggest that a protective allele may be in linkage disequilibrium with the *HLA-DRB1* alleles. It has been hypothesized that the presence of an asparagine amino acid at position 70 of the HLA-DRB1 protein (D70) may be associated with protection from the development of RA (once the effect of the shared epitope alleles has been taking into consideration) (Mattey et al., 2001a; Ruiz-Morales et al., 2004).

Numerous studies have shown that *HLA-DRB1* susceptibility alleles influence disease severity in longstanding disease, particularly the development of bony erosions (e.g., Chen et al., 2002; Gorman et al. 2004; Moxley and Cohen, 2002). More recently, however, it has been suggested that this association is primarily due to the presence of CCP autoantibodies (Huizinga et al., 2005). It remains to be determined whether *HLA-DRB1* alleles contribute additional risk of developing erosive disease independent of CCP autoantibodies (van der Helm-van Mil et al., 2006). This more recent observation may be an important explanation for why some studies have demonstrated that SE alleles predict erosive changes, but only in RF- patients (El-Gabalawy et al., 1999; Mattey et al., 2001b). One hypothesis to explain this observation is that RF- patients in these older studies are actually CCP+ [and it is known that SE alleles have a stronger association with CCP+ than RF+ RA (Irigoyen et al., 2005)]. In

TABLE 43.3 Non-MHC associations in RA

Gene	OR	Comments
PTPN22	1.75	Clear association with missense SNP (rs2476601) and CCP+ RA
CTLA4	1.20	Possible association with CT60 SNP (rs3087243) and RA
PADI4	1.20	Possible association with haplotype tagged by SNP (rs2240340) and RA
SLC22A4	1.25	Possible association with SNP (rs2073838) and RA

the future, it will be important to assess the relationship between *HLA-DRB1* alleles and clinical outcome, controlling for the effect of CCP as well as other important clinical variables.

Despite decades of research, it is not fully known how the *HLA-DRB1* alleles cause risk of RA, and direct functional proof has been difficult (Goronzy and Weyand, 1993; Nepom, 2001). Hypotheses include that the SE alleles influence (1) thresholds for T-cell activation [based on avidity between the T-cell receptor, MHC, and peptide, especially in the context of post-translational modification events important in RA pathogenesis (Hill et al., 2003)]; (2) thymic selection of high-affinity self-reactive T cells (based on the T-cell synovial repertoire) (Yang et al., 1999); and (3) molecular mimicry of microbial antigens (Albani et al., 1992). It is worthwhile noting that the third hypervariable region of the protein (location of SE allelic variants) contains a peptide-binding groove that serves to present peptides to CD4+ T cells (Seyfried et al., 1988), and that citrullination of certain peptides triggers a strong immune response to citrullinated peptides in *HLA-DRB1* *0401 transgenic mice (Hill et al., 2003).

Other MHC-Region Genes

Several studies suggest that additional genes within the MHC likely contribute to disease susceptibility, once the effect of *HLA-DRB1* has been taken into consideration (Jawaheer et al., 2002; Kochi et al., 2004; Mulcahy et al., 1996; Singal et al., 1999; Zanelli et al., 2001). For example, an extended haplotype that includes *HLA-DRB1* DR3 alleles may be associated with RA (Jawaheer et al., 2002). The associated haplotype spans ~500 kb, and contains Class III MHC genes, including the TNF-alpha region implicated in other studies (Mulcahy et al., 1996; Ota et al., 2001; Waldron-Lynch et al., 2001). One study suggests that this association is restricted to CCP- patients (Irigoyen et al., 2005).

No single study has tested comprehensively DNA variants within the MHC in a patient population large enough to detect subtle effects beyond *HLA-DRB1* alleles. Only recently have genetic linkage disequilibrium maps become available to thoroughly test the hypothesis that non-*HLA-DRB1* alleles influence the risk of developing RA (Miretti et al., 2005; Walsh et al., 2003). Application of high-density SNP genotyping in large patient collections should provide additional insight into the role of the MHC region in susceptibility and severity of RA.

Non-MHC Genes

The identity of genes contributing to RA that lie outside the MHC region has been more elusive. A sizeable portion of genetic variation is attributable to such non-HLA genes – up to two-thirds in some studies. The gene with the most convincing evidence of replication, *PTPN22*, influences threshold of T-cell activation (Begovich et al., 2004; Vang et al., 2005). The statistical evidence of association for other genes (e.g., *CTLA4*, *PADI4*, and *SLC22A4*) is not yet conclusive, and therefore it is not possible to draw broader conclusions on functional classification of genes associated with RA beyond *HLA-DRB1* and *PTPN22* (which implicate class II presentation and T-cell activation).

The most convincing non-HLA gene associated with RA is *PTPN22* (Begovich et al., 2004), a finding that has been replicated across multiple independent studies (e.g., Dieude et al., 2005; Harrison et al., 2006; Hinks et al., 2005; Lee et al., 2005; Orozco et al., 2005; Plenge et al., 2005; Steer et al., 2005; Viken et al., 2005; Zhernakova et al., 2005). The susceptibility allele is a missense variant that changes an arginine to tryptophan amino acid (R620W), resulting in alteration of T-cell activation (Begovich et al., 2004; Vang et al., 2005). The magnitude of the genetic effect, as measured by the OR, is substantially less than for HLA-DRB1*0401 but similar to other SE alleles (*PTPN22* OR ~1.75). Interestingly, this allele is absent in East Asians, and thus not associated with susceptibility in Japanese populations (Ikari et al., 2006).

Extensive genotype-phenotype correlations have not been conducted for *PTPN22* (as they have for *HLA-DRB1*). Like *HLA-DRB1* alleles, *PTPN22* only associates with seropositive RA. There is some evidence that *PTPN22* influences age of onset (Plenge et al., 2005), and may have a more significant effect in males compared to females (Pierer et al. 2006; Plenge et al., 2005), but no evidence that it influences disease activity or radiographic erosions (Harrison et al., 2006; Wesoly et al., 2005).

Additional genes have been implicated in genetic association studies of RA, but have not yet reached the level of statistical significance required to solidify their position as true RA-susceptibility genes (see Table 43.3). Some encouraging candidate genes include *CTLA4*, *PADI4*, and *SLC22A4*.

1. The association between variants with *CTLA4* and susceptibility to autoimmunity is most convincing in type 1 diabetes and autoimmune thyroiditis (Ueda et al., 2003). CTLA4 is a negative regulator of T-cell activation. In these diseases, an allele in the 3′ untranslated region (UTR) of the gene causes a modest increase in disease risk (OR ~1.2–1.5). Several studies have extended these findings to RA (Lei et al., 2005; Plenge et al., 2005), where the magnitude of the genetic effect is again modest (~1.20).
2. *PADI4* encodes for an enzyme that post-translationally modifies arginine to citrulline, and may therefore be important in generating anti-CCP autoantibodies. An initial report in Japanese patients implicated a common variant (population allele frequency ~35%) in disease risk (Suzuki et al., 2003); subsequent reports have been less convincing

statistically, but nonetheless support an association with RA susceptibility (Iwamoto et al., 2006).

3. Finally, a common allele within the *SLC22A4* gene may be associated with RA susceptibility in East Asian populations (Tokuhiro et al., 2003), but this result has not been replicated in RA patients of European ancestry (Plenge et al., 2005). The protein product of *SLC22A4* is an organic cation transporter expressed in hematological and immunological tissues. Of note, the putative causal allele, which disrupts a RUNX1-binding site, is at reduced population frequency in RA patients of European ancestry compared to East Asian ancestry (0.08 versus 0.30 minor allele frequency), thus limiting power to detect an association in RA patients with European ancestry.

Candidate gene studies in RA, as with many other complex diseases, highlight the current challenge of genetic association studies (Plenge and Rioux, 2006). First, most studies to date have tested a small fraction of genetic variation in the human genome. It is estimated that over 10 million common variants (population allele frequency >1%) exist in the human genome, yet most studies test a vanishingly small fraction of these variants. Second, the expected genetic effect for most disease alleles is quite modest (OR <1.5). Therefore, thousands of patients are required to detect the genetic effect – and most studies have been conducted on far fewer, limiting power to detect a true positive association.

In the immediate future, genome-wide association studies, which test the majority of common genetic variants in the human genome, will be conducted in RA and other autoimmune diseases. If appropriately designed and interpreted (Hirschhorn and Daly, 2005), it is expected that these studies will greatly expand the list of RA-susceptibility genes. Several genome-wide association studies are underway with genotyping platforms that capture >60% of common genetic variants (as estimated by Phase II HapMap).

Non-Genetic Risk Factors

Sex Bias and Hormonal Factors

Perhaps the strongest non-genetic risk factor for the development of RA is female sex: females are more than twice as likely to develop RA compared to males, and this disparity is even greater at a younger age (Linos et al., 1980; Symmons et al., 1994). In females, the risk of developing RA increases with age and peaks around the time of menopause (Doran et al., 2002; Goemaere et al., 1990; Karlson et al., 2004). These observations have lead to considerable effort in examining the role of hormonal and pregnancy factors in disease occurrence. These studies include:

1. Conditions associated with excess estrogen and/or progesterone may be protective against developing RA. The risk of developing RA is increased in the 12-month postpartum period (when serum hormone levels fall) (Silman et al., 1992), and symptoms are less severe during the post-ovulatory phase of the menstrual cycle and during pregnancy (when levels are elevated) (Latman, 1983; Ostensen et al., 1983).
2. Duration of breast-feeding may be associated with risk of developing RA. Women with >24 total months of breast-feeding have a twofold reduced risk of developing RA (Karlson et al., 2004).
3. Cross-sectional studies of serum androgen levels consistently demonstrate lower serum testosterone levels in RA patients (Cutolo and Accardo, 1991; Masi et al., 1995).
4. The hypothalamic-pituitary axis (HPA) is altered in RA patients (Chikanza et al., 1992), and conditions leading to panhypopituitarism and hypoadrenalism are associated with the onset of RA (Wilder, 1996).
5. Castration of female mice exaggerates arthritis (Holmdahl et al., 1986), and estrogen replacement therapy suppresses arthritis in the collagen-induced mouse model of arthritis (Holmdahl et al., 1987) and in adjuvant arthritis in the rat (Kappas et al., 1963), consistent with the role of estrogen in the human.
6. Finally, a subset of RA patients undergoes striking remission during pregnancy. Although fetal-maternal genetic relationships have been invoked to explain this phenomenon (Nelson et al., 1993), other data fails to fully support this hypothesis (Brennan et al., 2000; Gregersen, 2000), and thus it is likely that hormonal changes during pregnancy at least partially explain disease remission in pregnancy.

While it has been proposed that genes within sex steroid hormone pathways may contribute to RA inherited risk, no gene has been conclusively identified to date.

Environmental Factors

Smoking is the environmental risk factor with strongest association with developing RA. Over the past 20 years, many studies have convincingly shown that smoking increases risk of developing RA in both males and females by a factor of ~2.0 (Gorman, 2006; Stolt et al., 2003). Most of these studies demonstrate that the increase risk is in developing autoantibody positive RA, and that the risk is greatest for heavy, current smokers; the risk remains, however, for >10 years following smoking cessation. Because the link between smoking and autoantibody positive RA is reminiscent of the genetic association between *HLA-DRB1* "shared epitope" alleles and autoantibody positive RA, Padyukov et al. investigated the risk of developing disease in carriers of these two established risk factors: a 16-fold risk of developing RF+ RA in smokers who carry two copies of the SE alleles was observed (Padyukov et al., 2004).

Other putative environmental risk factors that have not been as widely replicated as smoking include blood transfusions, obesity, occupational silica and mineral oil exposure, and socioeconomic class (Silman and Pearson, 2002; Symmons, 2003). It is interesting that several of these putative environmental factors are inhaled through the lungs. Silica dust and mineral oil exposure, similar to exposure to cigarette smoke, was a risk factor only for seropositive RA. Mineral oil can also act as an adjuvant capable of inducing experimental arthritis in rodent models.

Infectious Agents

There is indirect, but certainly not conclusive evidence that exposure to an infectious agent(s) may trigger the development of RA (Silman and Pearson, 2002). Despite decades of investigation, however, no single infectious agent has emerged as influencing risk of RA.

Support for an infectious etiology contributing to RA risk include: (a) certain forms of arthritides in humans are triggered by bacteria, including enterogenic and urogenic infections (e.g., reactive arthritis, rheumatic fever, Lyme disease, and Whipple's disease); (b) bacterial components are able to induce chronic arthritis closely resembling RA in animal models; (c) antibodies, T-cell clones, or cellular immune responses specific to certain bacteria as well as bacterial components have been observed in synovial fluid or peripheral blood in patients with RA; and (d) and paleontological (Rothschild et al., 1988) and epidemiological data (Doran et al., 2002; Kaipiainen-Seppanen et al., 1996; Shichikawa et al., 1999) demonstrating the first appearance of RA in modern man, yet with a decrease in incidence over the last 40 years (with the improvement in public health measures and the widespread use of antimicrobial therapy).

If an infectious agent is causal, one could imagine different mechanisms by which it might lead to RA. Exposure might lead to a generalized over-active immune system, thus clearing the infection but resulting in autoimmunity later in life. A more specific response to a single infectious agent might induce an immunological response against the infection, yet lead to cross-reactivity with antigens in human synovial tissue ("molecular mimcry").

Autoantibodies

RF and CCP Autoantibodies

Autoantibodies have proven useful in the diagnosis and prognosis of RA patients. Autoantibodies are detected in approximately two-thirds of patients with RA and predict severe disease. The two major types of autoantibodies used clinically to create RA subsets are RF, which is an immunoglobulin specific to the Fc region of IgG, and anti-cyclic citrullinated peptide (CCP) antibodies, which are antibodies directed against peptides that have arginine posttranslationally modified to citrulline (Schellekens et al., 2000). These autoantibodies are strongly correlated but may represent distinct clinical subsets of RA.

RF autoantibodies are part of the diagnostic criteria for RA (Arnett et al., 1988). The RF assay, however, remains suboptimal as a diagnostic test, as it lacks sensitivity (50–90%) and specificity (50–90%) (Shmerling and Delbanco, 1991). Furthermore, it is present in many other disease states (including those that mimic RA) and patients that smoke, and its incidence increases with age. In contrast, anti-CCP antibodies have moderate sensitivity (60–80%) but increased specificity for RA (>90%), and predict functional status and radiographic erosions in patients with early-onset RA (van Jaarsveld et al., 1999; van Zeben et al., 1992; Kroot et al., 2000).

There has been much debate about whether these autoantibodies are causal or whether they represent a non-specific response to systemic inflammation. Several lines of evidence suggest that these autoantibodies – and CCP in particular – are pathogenic: (1) anti-CCP antibodies appear before disease onset (Rantapaa-Dahlqvist et al., 2003; Schellekens et al., 2000); (2) the presence of CCP is very specific to RA; (3) antibodies against citrullinated proteins enhance tissue injury in a murine model of arthritis (collagen-induced arthritis) (Kuhn et al., 2006); and (4) genetic variation in an enzyme, *PADI4*, involved in the citrullination pathway appears associated with RA susceptibility (Suzuki et al., 2003) and see "*Non-MHC genes*" above).

Finally, presence of RF and CCP antibodies may facilitate the identification of genes in RA. As with other complex human diseases, substantial clinical heterogeneity exists. Because presence or absence of RF and CCP antibodies provides a rationale clinical subset, these have been used to subset patients in the search for RA genes. Indeed, the two genes with a clear association to RA (*HLA-DRB1* and *PTPN22*) are both association with CCP+ patients, but not CCP- patients.

Hypothetical Model of RA Predisposition

Based upon the above information, it is tempting to speculate about a hypothetical model of RA pathogenesis. One model, put forth by Klareskog et al (Klareskog et al., 2006), is that smoking or other air pollutants lead to inflammation and citrullination of proteins in the lungs in genetically susceptible individuals. This event leads to the formation of anti-CCP antibodies, which have a direct pathogenic role in the joint. Under this hypothesis, citrullinated proteins are presented in class II MHC molecules to induce an immune response, and *HLA-DRB1* variants augment the autoimmune response. The *PTPN22* missense allele may also augment the autoimmune response, or may allow autoreactive T-cells to escape selection in the thymus.

While aspects of this model may be correct, it almost certainly does not explain RA pathogenesis in all patients. Continued integration of clinical and basic science research is necessary to revise, or refine, models of RA pathogenesis. The advent of genome-wide association studies should facilitate identification of new RA-related genes, facilitating our understanding of disease pathogenesis.

Other Genomic Studies

Other genomic resources are available to study predisposition to RA, including mRNA expression profiling and examination of serum protein levels, although neither has been widely implemented in large patient collections. The field of proteomics is still limited by the availability of cost-effective, high-throughput assays. The search for auto-antigens that predispose to RA has also been challenging. The goals of these efforts include identifying novel biomarkers that allow early and specific disease diagnosis or predict severe disease, or biomarkers that may serve as therapeutic targets. The success of RF and anti-CCP autoantibodies provide proof-of-concept that a novel biomarker might have direct clinical utility.

mRNA Expression Profiling

Genome-wide analysis of mRNA expression patterns has been performed in peripheral blood and synovial fluid derived from RA patients. This approach has proven insightful in the peripheral blood of patients with systemic lupus erythematosis (SLE) (Baechler et al., 2003), but has been less fruitful to date in RA. Studies in peripheral blood have compared RA patients versus normal controls (Batliwalla et al., 2005; van der Pouw Kraan et al., 2007), RF+ versus RF- patients (Bovin et al., 2004), and early versus chronic RA (Olsen et al., 2004). Similar techniques have been applied to synvovial tissue, where inflammatory RA synovium is compared to non-inflammatory OA synovium (Devauchelle et al., 2004; van der Pouw Kraan et al., 2003a, b). No clear consensus has emerged from these studies, but there is some evidence that there may be an alteration in the number or activation state of monocytes (Batliwalla et al., 2005) or a type I interferon signature (similar to SLE, [van der Pouw Kraan et al., 2007]). Many studies are too small (e.g., <15 patients) to draw any significant conclusions given the large number of hypotheses (i.e., expressed genes) tested (Bovin et al., 2004; Devauchelle et al., 2004; Haas et al., 2006; Szodoray et al., 2006).

Proteomics

Genome-wide genetic association studies and mRNA expression profiling allow for exploration of novel hypotheses, effectively exploring hundreds of thousands of genes for influence on disease risk. These platforms are becoming increasingly cost-effective and robust. In contrast, the field of proteomics has been hampered by equivalent technical capacity. Mass spectrometry offers such a platform, but is not yet widely in use (Domon and Aebersold, 2006). Consequently, most "proteomics" research in RA and other complex diseases has been limited to ELISA-based methodologies that are not capable of high-throughput multiplexing – and thus limited to a small number (hundreds not thousands) of proteins of known relevance.

It is clear that cytokines play a crucial role in RA pathogenesis (Lee and Weinblatt, 2001). Cytokines are small soluble proteins that mediate intercellular communication within the immune system. Important examples include the proinflammatory cytokines interleukin 1 (IL1) and tumor necrosis factor alpha (TNF-alpha), as well as their soluble receptors (p55 and p75 for TNF; IL1R1, IL1R2, and IL1Ra for IL1). These cytokines are secreted primarily by macrophages within the synovium, and lead to inflammation and stimulation of synovial tissue effector functions (e.g., cellular proliferation, expression of metalloproteinases and adhesion-molecules, and secretion of prostaglandins and other cytokines). Perhaps the most incriminating evidence for the role of cytokines in RA pathogenesis is in neutralizing therapies directed at the proteins themselves in patients with active RA: antibodies directed against TNF-alpha (or its soluble receptor) have been remarkably useful in the treatment of RA; anti-IL1 therapies are somewhat effective, but not to the same degree as anti-TNF-alpha therapies.

While cytokines are important in RA, what is not clear is whether alterations in cytokine secretion *per se* leads to the development of RA, or whether dysfunction is simply the result of an over-active immune system. No longitudinal cohort study has yet demonstrated alterations in these molecules prior to the development of RA – evidence that would be necessary to establish causality. An alternate strategy to establish causality would be to demonstrate that genetic variation within a cytokine gene leads to the development of RA. Numerous studies have suggested that alleles within the *TNF* gene influence RA risk, but it has yet to be conclusively shown that this effect is independent of linkage disequilibrium with *HLA-DRB1* or neighboring genes within the 3.6 Mb MHC region. In theory, all alleles that influence secretion of a given cytokine could be mapped and then tested to determine whether these alleles also influence the development of RA. To date, no such comprehensive list of alleles exists.

It is worth highlighting studies that have implemented newer proteomic techniques such as mass spectrometry (Saulot et al., 2002; de Seny et al., 2005; Liao et al., 2004; Uchida et al., 2002), multiplex cytokine arrays (Hueber et al., 2006) and antigen microarrays (Hueber et al., 2005) to analyze synovial fluid and serum in RA patients. A study by Liao et al. used a two-step mass spectrometry approach to generate protein profiles of synovial fluid from patients with either erosive or non-erosive RA (Liao et al., 2004). Among 33 proteins elevated in the synovial fluid of patients with erosive RA were C-reactive protein (CRP) and 6 members of the S100 protein family of calcium-binding proteins. The authors demonstrated that levels of CRP, S100A8 (calgranulin A), S100A9 (calgranulin B), and S100A12 (calgranulin C) proteins were also elevated in the serum of patients with erosive RA. A different study by de Seny et al., which included 34 patients with longstanding disease, identified 5 proteins elevated in the serum of patients with CCP+ RA (compared to patients with other inflammatory diseases), one of which was hypothesized to be myeloid-related protein 8 (MRP8) (de Seny et al., 2005). The MRP8 protein may be enriched in the synovial fluid of RA compared to osteoarthritis patients (Uchida et al., 2002). Finally, Hueber et al. used microarrays in a set of early RA patients to show that autoantibody reactivity against citrullinated epitopes is more common in patients with high-serum levels of TNFα, IL-1β, IL-6, IL-13, IL-15, and GM-CSF (Hueber et al., 2006).

SCREENING

No reliable methods to screen a population prior to disease onset exist for RA. It may be that the presence of CCP autoantibodies together with genetic susceptibility variants (*PTPN22* and *HLA-DRB1*) will increase risk substantially to the point where screening the healthy population because cost-effective (Berglin et al., 2004; Johansson et al., 2005). Under this scenario, it remains to be determined whether effective intervention to prevent RA will emerge – currently, RA therapies are directed at symptoms rather than a cure. Moreover, this simple approach does nothing to screen the 1/3 of RA patients who develop

autoantibody negative disease. Overall, no clear paradigm has emerged as to how screening the population would influence clinical decision-making, and additional attention should be paid to this area in the future.

DIAGNOSIS, PROGNOSIS, AND MONITORING

The diagnosis of RA is based on established clinical criteria (Arnett et al., 1988). The clinical course for any individual patient is highly variable. Extent of synovial inflammation and presence of autoantibodies at the time of initial diagnosis portend a poor prognosis, but cannot alone predict prognosis in any individual patient. Similarly, the presence of *HLA-DRB1* SE alleles predicts more severe disease, but does not routinely enter into decision-making in clinical patient care. Because treatment strategies have improved dramatically over the last decade with the advent of TNF-alpha inhibitors, prognosis is more influenced by response to treatment rather than aggressiveness of the underlying disease in an untreated patient.

A major goal of therapy is early institution of *d*isease-*m*odifying *a*nti-*r*heumatic *d*rugs, or DMARDs. The most popular disease-modifying therapy is low dose weekly methotrexate (MTX). This drug is used as a mono-therapy or in combination with other synthetic molecules (antimalarials, sulfasalazine, leflunomide) or with biologic response modifiers. Biologics that neutralize the pro-inflammatory cytokine TNF-alpha have been a major advance in the therapy of RA. Monoclonal antibodies (infliximab and adalimumab) or the p75 TNF-alpha soluble receptor fusion protein (etanercept) are effective as a mono-therapy or in combination with MTX in reducing the signs and symptoms of the disease, as well as improving quality of life and slowing the rate of radiographic progression. These therapies are very effective when combined with MTX. Newer biologics recently approved for use in RA include abatacept (blocks the CTLA4 co-stimulation pathway) and rituximab (a B-cell depleting strategy). Early intervention with MTX and combination therapy (including anti-TNF-alpha therapy) have had a major impact on this disease.

2009 UPDATE

The last two years has observed a dramatic increase in the number of validated risk alleles for complex traits, in large part due to the implementation of genome-wide association studies and related methodologies (Altshuler et al., 2008). In RA genetics, the greatest recent advance has been the discovery and validation of RA risk alleles outside of the MHC (Barton et al., 2008a, b; Chang et al., 2008; Consortium, 2007; Kurreeman et al., 2007; Plenge et al., 2007a, b; Raychaudhuri et al., 2008; Remmers et al., 2007; Suzuki et al., 2008; Thomson et al., 2007; Zhernakova et al., 2007). In addition, large-scale SNP genotyping across the MHC provides continued support for additional alleles, including alleles that are found on the conserved A1-B8-DR3 (8.1) haplotype and alleles near the *HLA-DPB1* gene (Lee et al., 2008).

There are six new validated non-MHC RA risk loci (overall $p < 5 \times 10^{-8}$ with independent replication), including 2q32/*STAT4* (Barton et al., 2008a; Kobayashi et al., 2008; Lee et al., 2007; Martinez et al., 2008; Orozco et al., 2008; Palomino-Morales et al., 2008; Remmers et al., 2007; Zervou et al., 2008), 9q33/*TRAF1-C5* (Barton et al., 2008a; Chang et al., 2008; Kurreeman et al., 2008; Plenge et al., 2007b; Zervou et al., 2008), 6q23/*TNFAIP3* (Plenge et al., 2007a; Thomson et al., 2007), 20q13/*CD40* (Raychaudhuri et al., 2008), 10p15/*PRKCQ* (Barton et al., 2008b; Raychaudhuri et al., 2008), and 12q13/*PIP4K2C* (Barton et al., 2008b; Raychaudhuri et al., 2008). There are at least six "highly suggestive" RA risk loci ($10^{-6} < p < 5 \times 10^{-8}$), which require additional replication before they can be considered validated RA risk loci: 2q33/*CTLA4* (Raychaudhuri et al., 2008), 4q27/*IL2-IL21* (Raychaudhuri et al., 2008; Zhernakova et al., 2007), 9p13/*CCL21* (Raychaudhuri et al., 2008), 1p36/*TNFRSF14* (Barton et al., 2008b; Raychaudhuri et al., 2008), 22q13/*IL2RB* (Barton et al., 2008b), and 1q22/*CD244* (Suzuki et al., 2008). For each of these loci, the causal allele and causal gene has not yet been definitively identified, indicating the need for fine-mapping and functional studies. The non-MHC risk alleles contribute between 3–5% of the genetic burden of RA (Raychaudhuri et al., 2008), compared to approximately one-third for the MHC (Cornelis et al., 1998; Deighton et al., 1989; Jawaheer et al., 2001; Rigby et al., 1991). Among individuals of European ancestry, the *PTPN22* locus (Begovich et al., 2004) continues to be the non-MHC region with the strongest effect, as estimated by the OR. All other non-MHC risk alleles are common in the general population and have modest effect on disease risk (OR $\sim$ 1.10).

The success in mapping RA risk loci has been paralleled by the success in other autoimmune and inflammatory diseases, including type 1 diabetes (Smyth et al., 2008; Todd et al., 2007), celiac disease (Smyth et al., 2008; van Heel et al., 2007), SLE (Graham et al., 2008; Harley et al., 2008; Hom et al., 2008), inflammatory bowel disease (Barrett et al., 2008), and multiple sclerosis (Hafler et al., 2007). An emerging theme is that many of the validated autoimmune risk loci predispose to more than one autoimmune disease (Behrens et al.,

2008; Consortium, 2008; Fung et al., 2008; Plenge, 2008; Raychaudhuri et al., 2008; Remmers et al., 2007; Smyth et al., 2008; Zhernakova et al., 2007). In RA, up to half of the known non-MHC risk loci also appear to confer risk to at least one additional autoimmune disease. In most examples, it is the same allele that is associated with risk in both diseases (e.g., a common allele at the *STAT4* gene locus and risk of both RA and SLE [Remmers et al., 2007]). There are gene loci, however, that appear to have multiple autoimmune alleles (e.g., the *6q23/TNFAIP3* locus and risk of both RA and SLE [Graham et al., 2008; Musone et al., 2008; Plenge et al., 2007a; Thomson et al., 2007]). Finally, there are loci, such as *PTPN22*, in which an allele *predisposes* to one autoimmune disease (e.g., RA [Begovich et al., 2004]), but the same allele is *protective* in another autoimmune disease (e.g., Crohn's disease [Barrett et al., 2008]).

These human genetic discoveries have identified pathways that are important in RA pathogenesis. One is the nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) signaling pathway (Hammaker et al., 2003; Sen and Baltimore, 1986). NF-κB, a protein complex that acts as a transcription factor, is found in almost all cell types and is involved in a variety of cellular responses to stimuli. Both TNF and CD40 signal through NF-κB to exert their effects on the immune system. Several of the RA risk loci contain genes that are involved in NF-κB signaling, including *CD40*, *TRAF1*, *TNFAIP3*, *PRKCQ*, and *TNFRSF14*.

2009 UPDATE REFERENCES

Altshuler, D., Daly, M.J. and Lander, E.S. (2008). Genetic mapping in human disease. *Science* 322, 881–888.

Barrett, J.C., Hansoul, S., Nicolae, D.L., Cho, J.H., Duerr, R.H., Rioux, J.D., Brant, S.R., Silverberg, M.S., Taylor, K.D., Barmada, M.M. et al. (2008). Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* 40(8), 955–962.

Barton, A., Thomson, W., Ke, X., Eyre, S., Hinks, A., Bowes, J., Gibbons, L., Plant, D., Wilson, A.G., Marinou, I. et al. (2008a). Re-evaluation of putative rheumatoid arthritis susceptibility genes in the post-genome wide association study era and hypothesis of a key pathway underlying susceptibility. *Hum Mol Genet* 17(15), 2274–2279.

Barton, A., Thomson, W., Ke, X., Eyre, S., Hinks, A., Bowes, J., Plant, D., Gibbons, L.J., Wilson, A.G., Bax, D.E. et al. (2008b). Rheumatoid arthritis susceptibility loci at chromosomes 10p15, 12q13 and 22q13. *Nat Genet* 40(10), 251–257.

Begovich, A.B., Carlton, V.E., Honigberg, L.A., Schrodi, S.J., Chokkalingam, A.P., Alexander, H.C., Ardlie, K.G., Huang, Q., Smith, A.M., Spoerke, J.M. et al. (2004). A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am J Hum Genet* 75, 330–337.

Behrens, E.M., Finkel, T.H., Bradfield, J.P., Kim, C.E., Linton, L., Casalunovo, T., Frackelton, E.C., Santa, E., Otieno, F.G., Glessner, J.T. et al. (2008). Association of the TRAF1-C5 locus on chromosome 9 with juvenile idiopathic arthritis. *Arthritis Rheum* 58, 2206–2207.

Chang, M., Rowland, C.M., Garcia, V.E., Schrodi, S.J., Catanese, J.J., van der Helm-van Mil, A.H., Ardlie, K.G., Amos, C.I., Criswell, L.A., Kastner, D.L. et al. (2008). A large-scale rheumatoid arthritis genetic study identifies association at chromosome 9q33.2. *PLoS Genet* 4, e1000107.

Consortium. (2008). The expanding genetic overlap between multiple sclerosis and type I diabetes. *Genes Immun* 10(1), 11–14.

Consortium, W.T.C.C. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661–678.

Cornelis, F., Faure, S., Martinez, M., Prud'homme, J.F., Fritz, P., Dib, C., Alves, H., Barrera, P., de Vries, N., Balsa, A. et al. (1998). New susceptibility locus for rheumatoid arthritis suggested by a genome-wide linkage study. *Proc Natl Acad Sci U S A* 95, 10746–10750.

Deighton, C.M., Walker, D.J., Griffiths, I.D. and Roberts, D.F. (1989). The contribution of HLA to rheumatoid arthritis. *Clin Genet* 36, 178–182.

Fung, E., Smyth, D.J., Howson, J.M., Cooper, J.D., Walker, N.M., Stevens, H., Wicker, L.S. and Todd, J.A. (2008). Analysis of 17 autoimmune disease-associated variants in type 1 diabetes identifies 6q23/TNFAIP3 as a susceptibility locus. *Genes Immun* 10(2), 188–191.

Graham, R.R., Cotsapas, C., Davies, L., Hackett, R., Lessard, C.J., Leon, J.M., Burtt, N.P., Guiducci, C., Parkin, M., Gates, C. et al. (2008). Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. *Nat Genet* 40(9), 1059–1061.

Hafler, D.A., Compston, A., Sawcer, S., Lander, E.S., Daly, M.J., De Jager, P.L., de Bakker, P.I., Gabriel, S.B., Mirel, D.B., Ivinson, A.J. et al. (2007). Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med* 357, 851–862.

Hammaker, D., Sweeney, S. and Firestein, G.S. (2003). Signal transduction networks in rheumatoid arthritis. *Ann Rheum Dis* 62(Suppl 2), ii86–ii89.

Harley, J.B., Alarcon-Riquelme, M.E., Criswell, L.A., Jacob, C.O., Kimberly, R.P., Moser, K.L., Tsao, B.P., Vyse, T.J., Langefeld, C.D., Nath, S.K. et al. (2008). Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. *Nat Genet* 40, 204–210.

Hom, G., Graham, R.R., Modrek, B., Taylor, K.E., Ortmann, W., Garnier, S., Lee, A.T., Chung, S.A., Ferreira, R.C., Pant, P.V. et al. (2008). Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. *N Engl J Med* 358, 900–909.

Jawaheer, D., Seldin, M.F., Amos, C.I., Chen, W.V., Shigeta, R., Monteiro, J., Kern, M., Criswell, L.A., Albani, S., Nelson, J.L. et al. (2001). A genomewide screen in multiplex rheumatoid arthritis families suggests genetic overlap with other autoimmune diseases. *Am J Hum Genet* 68, 927–936.

Kobayashi, S., Ikari, K., Kaneko, H., Kochi, Y., Yamamoto, K., Shimane, K., Nakamura, Y., Toyama, Y., Mochizuki, T., Tsukahara, S. et al. (2008). Association of STAT4 with susceptibility to rheumatoid arthritis and systemic lupus erythematosus in the Japanese population. *Arthritis Rheum* 58, 1940–1946.

Kurreeman, F.A., Padyukov, L., Marques, R.B., Schrodi, S.J., Seddighzadeh, M., Stoeken-Rijsbergen, G., van der Helm-van Mil, A.H., Allaart, C.F., Verduyn, W., Houwing-Duistermaat, J. et al. (2007). A candidate gene approach identifies the TRAF1/C5 region as a risk factor for rheumatoid arthritis. *PLoS medicine* 4, e278.

Kurreeman, F.A., Rocha, D., Houwing-Duistermaat, J., Vrijmoet, S., Teixeira, V.H., Migliorini, P., Balsa, A., Westhovens, R., Barrera, P., Alves, H. et al. (2008). Replication of the tumor necrosis factor receptor-associated factor 1/complement component 5 region as a susceptibility locus for rheumatoid arthritis in a European family-based study. *Arthritis Rheum* 58, 2670–2674.

Lee, H.S., Lee, A.T., Criswell, L.A., Seldin, M.F., Amos, C.I., Carulli, J.P., Navarrete, C., Remmers, E.F., Kastner, D.L., Plenge, R.M. et al. (2008). Several regions in the major histocompatibility complex confer risk for anti-CCP-antibody positive rheumatoid arthritis, independent of the DRB1 locus. *Mol Med* 14, 293–300.

Lee, H.S., Remmers, E.F., Le, J.M., Kastner, D.L., Bae, S.C. and P.K. Gregersen (2007). Association of STAT4 with Rheumatoid Arthritis in the Korean Population. *Mol Med* 13, 455–460.

Martinez, A., Varade, J., Marquez, A., Cenit, M.C., Espino, L., Perdigones, N., Santiago, J.L., Fernandez-Arquero, M., de la Calle, H., Arroyo, R. et al. (2008). Association of the STAT4 gene with increased susceptibility for some immune-mediated diseases. *Arthritis Rheum* 58, 2598–2602.

Musone, S.L., Taylor, K.E., Lu, T.T., Nititham, J., Ferreira, R.C., Ortmann, W. Shifrin, N., Petri, M.A., Ilyas Kamboh, M., Manzi, S. et al. (2008). Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus. *Nat Genet* 40(9), 1062–1064.

Orozco, G., Alizadeh, B.Z., Delgado-Vega, A.M., Gonzalez-Gay, M.A., Balsa, A., Pascual-Salcedo, D., Fernandez-Gutierrez, B., Gonzalez-Escribano, M.F., Petersson, I.F., van Riel, P.L. et al. (2008). Association of STAT4 with rheumatoid arthritis: A replication study in three European populations. *Arthritis Rheum* 58, 1974–1980.

Palomino-Morales, R.J., Rojas-Villarraga, A., Gonzalez, C.I., Ramirez, G., Anaya, J.M. and Martin, J. (2008). STAT4 but not TRAF1/C5 variants influence the risk of developing rheumatoid arthritis and systemic lupus erythematosus in Colombians. *Genes Immun* 9, 379–382.

Plenge, R.M. (2008). Shared genetic risk factors for type 1 diabetes and celiac disease. *N Engl J Med* 359(26), 2837–2838.

Plenge, R.M., Cotsapas, C., Davies, L., Price, A.L., de Bakker, P.I., Maller, J., Pe'er, I., Burtt, N.P., Blumenstiel, B., Defelice, M. et al. (2007a). Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat Genet* 39, 1477–1482.

Plenge, R.M., Seielstad, M., Padyukov, L., Lee, A.T., Remmers, E.F., Ding, B., Liew, A., Khalili, H., Chandrasekaran, A., Davies, L.R. et al. (2007b). TRAF1-C5 as a Risk Locus for Rheumatoid Arthritis – A Genomewide Study. *N Engl J Med* 357, 1199–1209.

Raychaudhuri, S., Remmers, E.F., Lee, A.T., Hackett, R., Guiducci, C., Burtt, N.P., Gianniny, L., Korman, B.D., Padyukov, L., Kurreeman, F.A. et al. (2008). Common variants at CD40 and other loci confer risk of rheumatoid arthritis. *Nat Genet* 40(10), 1216–1223.

Remmers, E.F., Plenge, R.M., Lee, A.T., Graham, R.R., Hom, G., Behrens, T.W., de Bakker, P.I., Le, J.M., Lee, H.S., Batliwalla, F. et al. (2007). STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N Engl J Med* 357, 977–986.

Rigby, A.S., Silman, A.J., Voelm, L., Gregory, J.C., Ollier, W.E., Khan, M.A., Nepom, G.T. and Thomson, G. (1991). Investigating the HLA component in rheumatoid arthritis: An additive (dominant) mode of inheritance is rejected, a recessive mode is preferred. *Genet Epidemiol* 8, 153–175.

Sen, R. and Baltimore, D. (1986). Multiple nuclear factors interact with the immunoglobulin enhancer sequences. *Cell* 46, 705–716.

Smyth, D.J., Plagnol, V., Walker, N.M., Cooper, J.D., Downes, K., Yang, J.H., Howson, J.M., Stevens, H., McManus, R., Wijmenga, C. et al. (2008). Shared and distinct genetic variants in type 1 diabetes and celiac disease. *N Engl J Med* 359, 2767–2777.

Suzuki, A., Yamada, R., Kochi, Y., Sawada, T., Okada, Y., Matsuda, K., Kamatani, Y., Mori, M., Shimane, K., Hirabayashi, Y. et al. (2008). Functional SNPs in CD244 increase the risk of rheumatoid arthritis in a Japanese population. *Nat Genet* 40(10), 1224–1229.

Thomson, W., Barton, A., Ke, X., Eyre, S., Hinks, A., Bowes, J., Donn, R., Symmons, D., Hider, S., Bruce, I.N. et al. (2007). Rheumatoid arthritis association at 6q23. *Nat Genet* 39, 1431–1433.

Todd, J.A., Walker, N.M., Cooper, J.D., Smyth, D.J., Downes, K., Plagnol, V., Bailey, R., Nejentsev, S., Field, S.F., Payne, F. et al. (2007). Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet* 39, 857–864.

van Heel, D.A., Franke, L., Hunt, K.A., Gwilliam, R., Zhernakova, A., Inouye, M., Wapenaar, M.C., Barnardo, M.C., Bethel, G., Holmes, G.K. et al. (2007). A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet* 39, 827–829.

Zervou, M.I., Sidiropoulos, P., Petraki, E., Vazgiourakis, V., Krasoudaki, E., Raptopoulou, A., Kritikos, H., Choustoulaki, E., Boumpas, D.T. and Goulielmos, G.N. (2008). Association of a TRAF1 and a STAT4 gene polymorphism with increased risk for rheumatoid arthritis in a genetically homogeneous population. *Hum Immunol* 69(9), 567–571.

Zhernakova, A., Alizadeh, B.Z., Bevova, M., van Leeuwen, M.A., Coenen, M.J., Franke, B., Franke, L., Posthumus, M.D., van Heel, D.A., van der Steege, G. et al. (2007). Novel association in chromosome 4q27 region with rheumatoid arthritis and confirmation of type 1 diabetes point to a general risk locus for autoimmune diseases. *Am J Hum Genet* 81, 1284–1288.

REFERENCES

Aho, K., Koskenvuo, M., Tuominen, J. and Kaprio, J. (1986). Occurrence of rheumatoid arthritis in a nationwide series of twins. *J Rheumatol* 13, 899–902.

Albani, S., Carson, D.A. and Roudier, J. (1992). Genetic and environmental factors in the immune pathogenesis of rheumatoid arthritis. *Rheum Dis Clin North Am* 18, 729–740.

Amos, C.I., Chen, W.V., Lee, A., Li, W., Kern, M., Lundsten, R., Batliwalla, F., Wener, M., Remmers, E., Kastner, D.A. et al. (2006). High-density SNP analysis of 642 Caucasian families with rheumatoid arthritis identifies two new linkage regions on 11p12 and 2q33. *Genes Immun* 7, 277–286.

Arnett, F.C., Edworthy, S.M., Bloch, D.A., McShane, D.J., Fries, J.F., Cooper, N.S., Healey, L.A., Kaplan, S.R., Liang, M.H., Luthra, H.S. et al. (1988). The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 31, 315–324.

Baechler, E.C., Batliwalla, F.M., Karypis, G., Gaffney, P.M., Ortmann, W.A., Espe, K.J., Shark, K.B., Grande, W.J., Hughes, K.M., Kapur, V. et al. (2003). Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc Natl Acad Sci USA* 100, 2610–2615.

Bali, D., Gourley, S., Kostyu, D.D., Goel, N., Bruce, I., Bell, A., Walker, D.J., Tran, K., Zhu, D.K., Costello, T.J. et al. (1999). Genetic analysis of multiplex rheumatoid arthritis families. *Genes Immun* 1, 28–36.

Batliwalla, F.M., Baechler, E.C., Xiao, X., Li, W., Balasubramanian, S., Khalili, H., Damle, A., Ortmann, W.A., Perrone, A., Kantor, A.B. et al. (2005). Peripheral blood gene expression profiling in rheumatoid arthritis. *Genes Immun* 6, 388–397.

Begovich, A.B., Carlton, V.E., Honigberg, L.A., Schrodi, S.J., Chokkalingam, A.P., Alexander, H.C., Ardlie, K.G. et al. (2004). A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am J Hum Genet* 75, 330–337.

Berglin, E., Padyukov, L., Sundin, U., Hallmans, G., Stenlund, H., Van Venrooij, W.J., Klareskog, L. and Dahlqvist, S.R. (2004). A combination of autoantibodies to cyclic citrullinated peptide (CCP) and HLA-DRB1 locus antigens is strongly associated with future onset of rheumatoid arthritis. *Arthritis Res Ther* 6, R303–R308.

Bovin, L.F., Rieneck, K., Workman, C., Nielsen, H., Sorensen, S.F., Skjodt, H., Florescu, A., Brunak, S. and Bendtzen, K. (2004). Blood cell gene expression profiling in rheumatoid arthritis. Discriminative genes and effect of rheumatoid factor. *Immunol Lett* 93, 217–226.

Brennan, P., Barrett, J., Fiddler, M., Thomson, W., Payton, T. and Silman, A. (2000). Maternal-fetal HLA incompatibility and the course of inflammatory arthritis during pregnancy. *J Rheumatol* 27, 2843–2848.

Chen, J.J., Mu, H., Jiang, Y., King, M.C., Thomson, G. and Criswell, L.A. (2002). Clinical usefulness of genetic information for predicting radiographic damage in rheumatoid arthritis. *J Rheumatol* 29, 2068–2073.

Chikanza, I.C., Petrou, P., Kingsley, G., Chrousos, G. and Panayi, G.S. (1992). Defective hypothalamic response to immune and inflammatory stimuli in patients with rheumatoid arthritis. *Arthritis Rheum* 35, 1281–1288.

Cornelis, F., Faure, S., Martinez, M., Prud'homme, J.F., Fritz, P., Dib, C., Alves, H. et al. (1998). New susceptibility locus for rheumatoid arthritis suggested by a genome-wide linkage study. *Proc Natl Acad Sci USA* 95, 10746–10750.

Cutolo, M. and Accardo, S. (1991). Sex hormones, HLA and rheumatoid arthritis. *Clin Exp Rheumatol* 9, 641–646.

de Seny, D., Fillet, M., Meuwis, M.A., Geurts, P., Lutteri, L., Ribbens, C., Bours, V., Wehenkel, L., Piette, J., Malaise, M. et al. (2005). Discovery of new rheumatoid arthritis biomarkers using the surface-enhanced laser desorption/ionization time-of-flight mass spectrometry ProteinChip approach. *Arthritis Rheum* 52, 3801–3812.

de Vries, N., Tijssen, H., van Riel, P.L. and van de Putte, L.B. (2002). Reshaping the shared epitope hypothesis: HLA-associated risk for rheumatoid arthritis is encoded by amino acid substitutions at positions 67–74 of the HLA-DRB1 molecule. *Arthritis Rheum* 46, 921–928.

Deighton, C.M., Walker, D.J., Griffiths, I.D. and Roberts, D.F. (1989). The contribution of HLA to rheumatoid arthritis. *Clin Genet* 36, 178–182.

Devauchelle, V., Marion, S., Cagnard, N., Mistou, S., Falgarone, G., Breban, M., Letourneur, F., Pitaval, A., Alibert, O., Lucchesi, C. et al. (2004). DNA microarray allows molecular profiling of rheumatoid arthritis and identification of pathophysiological targets. *Genes Immun* 5, 597–608.

Dieude, P., Garnier, S., Michou, L., Petit-Teixeira, E., Glikmans, E., Pierlot, C., Lasbleiz, S., Bardin, T., Prum, B. and Cornelis, F. (2005). Rheumatoid arthritis seropositive for the rheumatoid factor is linked to the protein tyrosine phosphatase nonreceptor 22-620W allele. *Arthritis Res Ther* 7, R1200–R1207.

Domon, B. and Aebersold, R. (2006). Mass spectrometry and protein analysis. *Science* 312, 212–217.

Doran, M.F., Pond, G.R., Crowson, C.S., O'Fallon, W.M. and Gabriel, S.E. (2002). Trends in incidence and mortality in rheumatoid arthritis in Rochester, Minnesota, over a forty-year period. *Arthritis Rheum* 46, 625–631.

El-Gabalawy, H.S., Goldbach-Mansky, R., Smith, D., II, Arayssi, T., Bale, S., Gulko, P., Yarboro, C., Wilder, R.L., Klippel, J.H. and Schumacher, H.R., Jr (1999). Association of HLA alleles and clinical features in patients with synovitis of recent onset. *Arthritis Rheum* 42, 1696–1705.

Etzel, C.J., Chen, W.V., Shepard, N., Jawaheer, D., Cornelis, F., Seldin, M.F., Gregersen, P.K. and Amos, C.I. (2006). Genome-wide meta-analysis for rheumatoid arthritis. *Hum Genet* 119, 634–641.

Firestein, G.S. (2003). Evolving concepts of rheumatoid arthritis. *Nature* 423, 356–361.

Gao, X., Gazit, E., Livneh, A. and Stastny, P. (1991). Rheumatoid arthritis in Israeli Jews: Shared sequences in the third hypervariable region of DRB1 alleles are associated with susceptibility. *J Rheumatol* 18, 801–803.

Goemaere, S., Ackerman, C., Goethals, K., De Keyser, F., Van der Straeten, C., Verbruggen, G., Mielants, H. and Veys, E.M. (1990). Onset of symptoms of rheumatoid arthritis in relation to age, sex and menopausal transition. *J Rheumatol* 17, 1620–1622.

Gorman, J.D. (2006). Smoking and rheumatoid arthritis: Another reason to just say no. *Arthritis Rheum* 54, 10–13.

Gorman, J.D., Lum, R.F., Chen, J.J., Suarez-Almazor, M.E., Thomson, G. and Criswell, L.A. (2004). Impact of shared epitope genotype and ethnicity on erosive disease: a meta-analysis of 3,240 rheumatoid arthritis patients. *Arthritis Rheum* 50, 400–412.

Goronzy, J.J. and Weyand, C.M. (1993). Interplay of T lymphocytes and HLA-DR molecules in rheumatoid arthritis. *Curr Opin Rheumatol* 5, 169–177.

Gregersen, P.K. (2000). Pregnant with controversy. *J Rheumatol* 27, 2738–2739.

Gregersen, P.K., Moriuchi, T., Karr, R.W., Obata, F., Moriuchi, J., Maccari, J., Goldberg, D., Winchester, R.J. and Silver, J. (1986a). Polymorphism of HLA-DR beta chains in DR4, -7, and -9 haplotypes: Implications for the mechanisms of allelic variation. *Proc Natl Acad Sci USA* 83, 9149–9153.

Gregersen, P.K., Shen, M., Song, Q.L., Merryman, P., Degar, S., Seki, T., Maccari, J., Goldberg, D., Murphy, H., Schwenzer, J. et al. (1986b). Molecular diversity of HLA-DR4 haplotypes. *Proc Natl Acad Sci USA* 83, 2642–2646.

Gregersen, P.K., Silver, J. and Winchester, R.J. (1987). The shared epitope hypothesis an approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis Rheum* 30, 1205–1213.

Haas, C.S., Creighton, C.J., Pi, X., Maine, I., Koch, A.E., Haines, G.K., Ling, S., Chinnaiyan, A.M. and Holoshitz, J. (2006). Identification of genes modulated in rheumatoid arthritis using complementary DNA microarray analysis of lymphoblastoid B cell lines from disease-discordant monozygotic twins. *Arthritis Rheum* 54, 2047–2060.

Harrison, P., Pointon, J.J., Farrar, C., Brown, M.A. and Wordsworth, B.P. (2006). Effects of PTPN22 C1858T polymorphism on susceptibility and clinical characteristics of British Caucasian rheumatoid arthritis patients. *Rheumatology (Oxford)* 45, 1009–1011.

Hasstedt, S.J., Clegg, D.O., Ingles, L. and Ward, R.H. (1994). HLA-linked rheumatoid arthritis. *Am J Hum Genet* 55, 738–746.

Hill, J.A., Southwood, S., Sette, A., Jevnikar, A.M., Bell, D.A. and Cairns, E. (2003). Cutting edge: The conversion of arginine to citrulline allows for a high-affinity peptide interaction with the rheumatoid arthritis-associated HLA-DRB1*0401 MHC class II molecule. *J Immunol* 171, 538–541.

Hinks, A., Barton, A., John, S., Bruce, I., Hawkins, C., Griffiths, C.E., Donn, R., Thomson, W., Silman, A. and Worthington, J. (2005). Association between the PTPN22 gene and rheumatoid arthritis and juvenile idiopathic arthritis in a UK population: Further support that PTPN22 is an autoimmunity gene. *Arthritis Rheum* 52, 1694–1699.

Hirschhorn, J.N. and Daly, M.J. (2005). Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 6, 95–108.

Holmdahl, R., Jansson, L. and Andersson, M. (1986). Female sex hormones suppress development of collagen-induced arthritis in mice. *Arthritis Rheum* 29, 1501–1509.

Holmdahl, R., Jansson, L., Meyerson, B. and Klareskog, L. (1987). Oestrogen induced suppression of collagen arthritis: I. Long term oestradiol treatment of DBA/1 mice reduces severity and incidence of arthritis and decreases the anti type II collagen immune response. *Clin Exp Immunol* 70, 372–378.

Horton, R., Wilming, L., Rand, V., Lovering, R.C., Bruford, E.A., Khodiyar, V.K., Lush, M.J., Povey, S., Talbot, C.C., Wright, M.W. et al. (2004). Gene map of the extended human MHC. *Nat Rev Genet* 5, 889–899.

Hueber, W., Kidd, B.A., Tomooka, B.H., Lee, B.J., Bruce, B., Fries, J.F., Sonderstrup, G., Monach, P., Drijfhout, J.W., van Venrooij, W.J. et al. (2005). Antigen microarray profiling of autoantibodies in rheumatoid arthritis. *Arthritis Rheum* 52, 2645–2655.

Hueber, W., Tomooka, B.H., Zhao, X., Kidd, B.A., Drijfhout, J.W., Fries, J.F., Van Venrooij, W.J., Metzger, A.L., Genovese, M.C. and Robinson, W.H. (2006). Proteomic analysis of secreted proteins in early rheumatoid arthritis: Anti-citrulline reactivity is associated with upregulation of proinflammatory cytokines. *Ann Rheum Dis*.

Huizinga, T.W., Amos, C.I., van der Helm-van Mil, A.H., Chen, W., van Gaalen, F.A., Jawaheer, D., Schreuder, G.M., Wener, M., Breedveld, F.C., Ahmad, N. et al. (2005). Refining the complex rheumatoid arthritis phenotype based on specificity of the HLA-DRB1 shared epitope for antibodies to citrullinated proteins. *Arthritis Rheum* 52, 3433–3438.

Ikari, K., Momohara, S., Inoue, E., Tomatsu, T., Hara, M., Yamanaka, H. and Kamatani, N. (2006). Haplotype analysis revealed no association between the PTPN22 gene and RA in a Japanese population. *Rheumatology (Oxford)* 45, 1345–1348.

Irigoyen, P., Lee, A.T., Wener, M.H., Li, W., Kern, M., Batliwalla, F., Lum, R.F., Massarotti, E., Weisman, M., Bombardier, C. et al. (2005). Regulation of anti-cyclic citrullinated peptide antibodies in rheumatoid arthritis: contrasting effects of HLA-DR3 and the shared epitope alleles. *Arthritis Rheum* 52, 3813–3818.

Iwamoto, T., Ikari, K., Nakamura, T., Kuwahara, M., Toyama, Y., Tomatsu, T., Momohara, S. and Kamatani, N. (2006). Association between PADI4 and rheumatoid arthritis: A meta-analysis. *Rheumatology (Oxford)* 45, 804–807.

Jarvinen, P. and Aho, K. (1994). Twin studies in rheumatic diseases. *Semin Arthritis Rheum* 24, 19–28.

Jawaheer, D., Seldin, M.F., Amos, C.I., Chen, W.V., Shigeta, R., Monteiro, J., Kern, M., Criswell, L.A., Albani, S., Nelson, J.L. et al. (2001). A genomewide screen in multiplex rheumatoid arthritis families suggests genetic overlap with other autoimmune diseases. *Am J Hum Genet* 68, 927–936.

Jawaheer, D., Li, W., Graham, R.R., Chen, W., Damle, A., Xiao, X., Monteiro, J., Khalili, H., Lee, A., Lundsten, R. et al. (2002). Dissecting the genetic complexity of the association between human leukocyte antigens and rheumatoid arthritis. *Am J Hum Genet* 71, 585–594.

Jawaheer, D., Seldin, M.F., Amos, C.I., Chen, W.V., Shigeta, R., Etzel, C., Damle, A. et al. (2003). Screening the genome for rheumatoid arthritis susceptibility genes: A replication study and combined analysis of 512 multicase families. *Arthritis Rheum* 48, 906–916.

Johansson, M., Arlestig, L., Hallmans, G. and Rantapaa-Dahlqvist, S. (2005). PTPN22 polymorphism and anti-cyclic citrullinated peptide antibodies in combination strongly predicts future onset of rheumatoid arthritis and has a specificity of 100% for the disease. *Arthritis Res Ther* 8, R19.

Kaipiainen-Seppanen, O., Aho, K., Isomaki, H. and Laakso, M. (1996). Incidence of rheumatoid arthritis in Finland during 1980-1990. *Ann Rheum Dis* 55, 608–611.

Kappas, A., Jones, H.E. and Roitt, I.M. (1963). Effects of steroid sex hormones on immunological phenomena. *Nature* 198, 902.

Karlson, E.W., Mandl, L.A., Hankinson, S.E. and Grodstein, F. (2004). Do breast-feeding and other reproductive factors influence future risk of rheumatoid arthritis? Results from the Nurses' Health Study. *Arthritis Rheum* 50, 3458–3467.

Klareskog, L., Stolt, P., Lundberg, K., Kallberg, H., Bengtsson, C., Grunewald, J., Ronnelid, J., Harris, H.E., Ulfgren, A.K., Rantapaa-Dahlqvist, S. et al. (2006). A new model for an etiology of rheumatoid arthritis: Smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. *Arthritis Rheum* 54, 38–46.

Kochi, Y., Yamada, R., Kobayashi, K., Takahashi, A., Suzuki, A., Sekine, A., Mabuchi, A., Akiyama, F., Tsunoda, T., Nakamura, Y. et al. (2004). Analysis of single-nucleotide polymorphisms in Japanese rheumatoid arthritis patients shows additional susceptibility markers besides the classic shared epitope susceptibility sequences. *Arthritis Rheum* 50, 63–71.

Kroot, E.J., de Jong, B.A., van Leeuwen, M.A., Swinkels, H., van den Hoogen, F.H., van't Hof, M., van de Putte, L.B., van Rijswijk, M.H., van Venrooij, W.J. and van Riel, P.L. (2000). The prognostic value of

anti-cyclic citrullinated peptide antibody in patients with recent-onset rheumatoid arthritis. *Arthritis Rheum* 43, 1831–1835.

Kuhn, K.A., Kulik, L., Tomooka, B., Braschler, K.J., Arend, W.P., Robinson, W.H. and Holers, V.M. (2006). Antibodies against citrullinated proteins enhance tissue injury in experimental autoimmune arthritis. *J Clin Invest* 116, 961–973.

Latman, N.S. (1983). Relation of menstrual cycle phase to symptoms of rheumatoid arthritis. *Am J Med* 74, 957–960.

Lee, A.T., Li, W., Liew, A., Bombardier, C., Weisman, M., Massarotti, E.M., Kent, J., Wolfe, F., Begovich, A. and Gregersen, P.K. (2005). The PTPN22 R620W polymorphism associates with RF positive rheumatoid arthritis in a dose dependent manner but not with HLA shared epitope status. *Genes and Immunity* 6, 129–133.

Lee, D.M. and Weinblatt, M.E. (2001). Rheumatoid arthritis. *Lancet* 358, 903–911.

Lei, C., Dongqing, Z., Yeqing, S., Oaks, M.K., Lishan, C., Jianzhong, J., Jie, Q., Fang, D., Ningli, L., Xinghai, H. et al. (2005). Association of the CTLA-4 gene with rheumatoid arthritis in Chinese Han population. *Eur J Hum Genet* 13, 823–828.

Liao, H., Wu, J., Kuhn, E., Chin, W., Chang, B., Jones, M.D., O'Neil, S., Clauser, K.R., Karl, J., Hasler, F. et al. (2004). Use of mass spectrometry to identify protein biomarkers of disease severity in the synovial fluid and serum of patients with rheumatoid arthritis. *Arthritis Rheum* 50, 3792–3803.

Linos, A., Worthington, J.W., O'Fallon, W.M. and Kurland, L.T. (1980). The epidemiology of rheumatoid arthritis in Rochester, Minnesota: A study of incidence, prevalence, and mortality. *Am J Epidemiol* 111, 87–98.

MacGregor, A.J., Snieder, H., Rigby, A.S., Koskenvuo, M., Kaprio, J., Aho, K. and Silman, A.J. (2000). Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum* 43, 30–37.

MacKay, K., Eyre, S., Myerscough, A., Milicic, A., Barton, A., Laval, S., Barrett, J., Lee, D., White, S., John, S. et al. (2002). Whole-genome linkage analysis of rheumatoid arthritis susceptibility loci in 252 affected sibling pairs in the United Kingdom. *Arthritis Rheum* 46, 632–639.

Masi, A.T., Feigenbaum, S.L. and Chatterton, R.T. (1995). Hormonal and pregnancy relationships to rheumatoid arthritis: Convergent effects with immunologic and microvascular systems. *Semin Arthritis Rheum* 25, 1–27.

Mattey, D.L., Dawes, P.T., Gonzalez-Gay, M.A., Garcia-Porrua, C., Thomson, W., Hajeer, A.H. and Ollier, W.E. (2001a). HLA-DRB1 alleles encoding an aspartic acid at position 70 protect against development of rheumatoid arthritis. *J Rheumatol* 28, 232–239.

Mattey, D.L., Hassell, A.B., Dawes, P.T., Cheung, N.T., Poulton, K.V., Thomson, W., Hajeer, A.H. and Ollier, W.E. (2001b). Independent association of rheumatoid factor and the HLA-DRB1 shared epitope with radiographic outcome in rheumatoid arthritis. *Arthritis Rheum* 44, 1529–1533.

McDaniel, D.O., Alarcon, G.S., Pratt, P.W. and Reveille, J.D. (1995). Most African-American patients with rheumatoid arthritis do not have the rheumatoid antigenic determinant (epitope). *Ann Intern Med* 123, 181–187.

Michou, L., Croiseau, P., Petit-Teixeira, E., du Montcel, S.T., Lemaire, I., Pierlot, C., Osorio, J., Frigui, W., Lasbleiz, S., Quillet, P. et al. (2006). Validation of the reshaped shared epitope HLA-DRB1 classification in rheumatoid arthritis. *Arthritis Res Ther* 8, R79.

Miretti, M.M., Walsh, E.C., Ke, X., Delgado, M., Griffiths, M., Hunt, S., Morrison, J., Whittaker, P., Lander, E.S., Cardon, L.R. et al. (2005). A high-resolution linkage-disequilibrium map of the human major histocompatibility complex and first generation of tag single-nucleotide polymorphisms. *Am J Hum Genet* 76, 634–646.

Moxley, G. and Cohen, H.J. (2002). Genetic studies, clinical heterogeneity, and disease outcome studies in rheumatoid arthritis. *Rheum Dis Clin North Am* 28, 39–58.

Mulcahy, B., Waldron-Lynch, F., McDermott, M.F., Adams, C., Amos, C.I., Zhu, D.K., Ward, R.H., Clegg, D.O., Shanahan, F. and Molloy, M.G. (1996). Genetic variability in the tumor necrosis factor-lymphotoxin region influences susceptibility to rheumatoid arthritis. *Am J Hum Genet* 59, 676–683.

Nelson, J.L., Hughes, K.A., Smith, A.G., Nisperos, B.B., Branchaud, A.M. and Hansen, J.A. (1993). Maternal-fetal disparity in HLA class II alloantigens and the pregnancy-induced amelioration of rheumatoid arthritis. *N Engl J Med* 12, 466–471.

Nepom, G.T. (2001). The role of the DR4 shared epitope in selection and commitment of autoreactive T cells in rheumatoid arthritis. *Rheum Dis Clin North Am* 27, 305–315.

Ollier, W. and Thomson, W. (1992). Population genetics of rheumatoid arthritis. *Rheum Dis Clin North Am* 18, 741–759.

Olsen, N., Sokka, T., Seehorn, C.L., Kraft, B., Maas, K., Moore, J. and Aune, T.M. (2004). A gene expression signature for recent onset rheumatoid arthritis in peripheral blood mononuclear cells. *Ann Rheum Dis* 63, 1387–1392.

Orozco, G., Sanchez, E., Gonzalez-Gay, M.A., Lopez-Nevot, M.A., Torres, B., Caliz, R., Ortego-Centeno, N., Jimenez-Alonso, J., Pascual-Salcedo, D., Balsa, A. et al. (2005). Association of a functional single-nucleotide polymorphism of PTPN22, encoding lymphoid protein phosphatase, with rheumatoid arthritis and systemic lupus erythematosus. *Arthritis Rheum* 52, 219–224.

Ostensen, M., Aune, B. and Husby, G. (1983). Effect of pregnancy and hormonal changes on the activity of rheumatoid arthritis. *Scand J Rheumatol* 12, 69–72.

Ota, M., Katsuyama, Y., Kimura, A., Tsuchiya, K., Kondo, M., Naruse, T., Mizuki, N., Itoh, K., Sasazuki, T. and Inoko, H. (2001). A second susceptibility gene for developing rheumatoid arthritis in the human MHC is localized within a 70-kb interval telomeric of the TNF genes in the HLA class III region. *Genomics* 71, 263–270.

Padyukov, L., Silva, C., Stolt, P., Alfredsson, L. and Klareskog, L. (2004). A gene-environment interaction between smoking and shared epitope genes in HLA-DR provides a high risk of seropositive rheumatoid arthritis. *Arthritis Rheum* 50, 3085–3092.

Pierer, M., Kaltenhauser, S., Arnold, S., Wahle, M., Baerwald, C., Hantzschel, H. and Wagner, U. (2006). Association of PTPN22 1858 single-nucleotide polymorphism with rheumatoid arthritis in a German cohort: Higher frequency of the risk allele in male compared to female patients. *Arthritis Res Ther* 8, R75.

Plenge, R. and Rioux, J.D. (2006). Identifying susceptibility genes for immunological disorders: Patterns, power, and proof. *Immunol Rev* 210, 40–51.

Plenge, R.M., Padyukov, L., Remmers, E.F., Purcell, S., Lee, A.T., Karlson, E.W., Wolfe, F., Kastner, D.L., Alfredsson, L., Altshuler, D. et al. (2005). Replication of putative candidate-gene associations with rheumatoid arthritis in >4,000 samples from North America and Sweden: Association of susceptibility with PTPN22, CTLA4, and PADI4. *Am J Hum Genet* 77, 1044–1060.

Rantapaa-Dahlqvist, S., de Jong, B.A., Berglin, E., Hallmans, G., Wadell, G., Stenlund, H., Sundin, U. and van Venrooij, W.J. (2003). Antibodies against cyclic citrullinated peptide and IgA rheumatoid factor predict the development of rheumatoid arthritis. *Arthritis Rheum* 48, 2741–2749.

Rigby, A.S., Silman, A.J., Voelm, L., Gregory, J.C., Ollier, W.E., Khan, M.A., Nepom, G.T. and Thomson, G. (1991). Investigating the HLA component in rheumatoid arthritis: An additive (dominant) mode of inheritance is rejected, a recessive mode is preferred. *Genet Epidemiol* 8, 153–175.

Rothschild, B.M., Turner, K.R. and DeLuca, M.A. (1988). Symmetrical erosive peripheral polyarthritis in the Late Archaic Period of Alabama. *Science* 241, 1498–1501.

Ruiz-Morales, J.A., Vargas-Alarcon, G., Flores-Villanueva, P.O., Villarreal-Garza, C., Hernandez-Pacheco, G., Yamamoto-Furusho, J.K., Rodriguez-Perez, J.M., Perez-Hernandez, N., Rull, M., Cardiel, M.H. et al. (2004). HLA-DRB1 alleles encoding the "shared epitope" are associated with susceptibility to developing rheumatoid arthritis whereas HLA-DRB1 alleles encoding an aspartic acid at position 70 of the beta-chain are protective in Mexican Mestizos. *Hum Immunol* 65, 262–269.

Saulot, V., Vittecoq, O., Charlionet, R., Fardellone, P., Lange, C., Marvin, L., Machour, N., Le Loet, X., Gilbert, D. and Tron, F. (2002). Presence of autoantibodies to the glycolytic enzyme alpha-enolase in sera from patients with early rheumatoid arthritis. *Arthritis Rheum* 46, 1196–1201.

Schellekens, G.A., Visser, H., de Jong, B.A., van den Hoogen, F.H., Hazes, J.M., Breedveld, F.C. and van Venrooij, W.J. (2000). The diagnostic properties of rheumatoid arthritis antibodies recognizing a cyclic citrullinated peptide. *Arthritis Rheum* 43, 155–163.

Seyfried, C.E., Mickelson, E., Hansen, J.A. and Nepom, G.T. (1988). A specific nucleotide sequence defines a functional T-cell recognition epitope shared by diverse HLA-DR specificities. *Hum Immunol* 21, 289–299.

Shichikawa, K., Inoue, K., Hirota, S., Maeda, A., Ota, H., Kimura, M., Ushiyama, T. and Tsujimoto, M. (1999). Changes in the incidence and prevalence of rheumatoid arthritis in Kamitonda, Wakayama, Japan, 1965-1996. *Ann Rheum Dis* 58, 751–756.

Shiozawa, S., Hayashi, S., Tsukamoto, Y., Goko, H., Kawasaki, H., Wada, T., Shimizu, K., Yasuda, N., Kamatani, N., Takasugi, K. et al. (1998). Identification of the gene loci that predispose to rheumatoid arthritis. *Int Immunol* 10, 1891–1895.

Shmerling, R.H. and Delbanco, T.L. (1991). The rheumatoid factor: An analysis of clinical utility. *Am J Med* 91, 528–534.

Silman, A., Kay, A. and Brennan, P. (1992). Timing of pregnancy in relation to the onset of rheumatoid arthritis. *Arthritis Rheum* 35, 152–155.

Silman, A.J., MacGregor, A.J., Thomson, W., Holligan, S., Carthy, D., Farhan, A. and Ollier, W.E. (1993). Twin concordance rates for rheumatoid arthritis: Results from a nationwide study. *Br J Rheumatol* 32, 903–907.

Silman, A.J. and Pearson, J.E. (2002). Epidemiology and genetics of rheumatoid arthritis. *Arthritis Res* 4(Suppl 3), S265–S272.

Singal, D.P., Li, J. and Lei, K. (1999). Genetics of rheumatoid arthritis (RA): Two separate regions in the major histocompatibility complex contribute to susceptibility to RA. *Immunol Lett* 69, 301–306.

Stastny, P. (1978). Association of the B-cell alloantigen DRw4 with rheumatoid arthritis. *N Engl J Med* 298, 869–871.

Stastny, P. and Fink, C.W. (1977). HLA-Dw4 in adult and juvenile rheumatoid arthritis. *Transplant Proc* 9, 1863–1866.

Steer, S., Lad, B., Grumley, J.A., Kingsley, G.H. and Fisher, S.A. (2005). Association of R602W in a protein tyrosine phosphatase gene with a high risk of rheumatoid arthritis in a British population: Evidence for an early onset/disease severity effect. *Arthritis Rheum* 52, 358–360.

Stewart, C.A., Horton, R., Allcock, R.J., Ashurst, J.L., Atrazhev, A.M., Coggill, P., Dunham, I. et al. (2004). Complete MHC haplotype sequencing for common disease gene mapping. *Genome Res* 14, 1176–1187.

Stolt, P., Bengtsson, C., Nordmark, B., Lindblad, S., Lundberg, I., Klareskog, L. and Alfredsson, L. (2003). Quantification of the influence of cigarette smoking on rheumatoid arthritis: Results from a population based case-control study, using incident cases. *Ann Rheum Dis* 62, 835–841.

Suzuki, A., Yamada, R., Chang, X., Tokuhiro, S., Sawada, T., Suzuki, M., Nagasaki, M., Nakayama-Hamada, M., Kawaida, R., Ono, M. et al. (2003). Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nat Genet* 34, 395–402.

Symmons, D.P. (2003). Environmental factors and the outcome of rheumatoid arthritis. *Best Pract Res Clin Rheumatol* 17, 717–727.

Symmons, D.P., Barrett, E.M., Bankhead, C.R., Scott, D.G. and Silman, A.J. (1994). The incidence of rheumatoid arthritis in the United Kingdom: Results from the Norfolk Arthritis Register. *Br J Rheumatol* 33, 735–739.

Szodoray, P., Alex, P., Frank, M.B., Turner, M., Turner, S., Knowlton, N., Cadwell, C., Dozmorov, I., Tang, Y., Wilson, P.C. et al. (2006). A genome-scale assessment of peripheral blood B-cell molecular homeostasis in patients with rheumatoid arthritis. *Rheumatology (Oxford)* 45, 1466–1476.

Teller, K., Budhai, L., Zhang, M., Haramati, N., Keiser, H.D. and Davidson, A. (1996). HLA-DRB1 and DQB typing of Hispanic American patients with rheumatoid arthritis: The "shared epitope" hypothesis may not apply. *J Rheumatol* 23, 1363–1368.

Tokuhiro, S., Yamada, R., Chang, X., Suzuki, A., Kochi, Y., Sawada, T., Suzuki, M., Nagasaki, M., Ohtsuki, M., Ono, M. et al. (2003). An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis. *Nat Genet* 35, 341–348.

Uchida, T., Fukawa, A., Uchida, M., Fujita, K. and Saito, K. (2002). Application of a novel protein biochip technology for detection and identification of rheumatoid arthritis biomarkers in synovial fluid. *J Proteome Res* 1, 495–499.

Ueda, H., Howson, J.M., Esposito, L., Heward, J., Snook, H., Chamberlain, G., Rainbow, D.B. et al. (2003). Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* 423, 506–511.

van der Helm-van Mil, A.H., Verpoort, K.N., Breedveld, F.C., Huizinga, T.W., Toes, R.E. and de Vries, R.R. (2006). The HLA-DRB1 shared epitope alleles are primarily a risk factor for anti-cyclic citrullinated peptide antibodies and are not an independent risk factor for development of rheumatoid arthritis. *Arthritis Rheum* 54, 1117–1121.

van der Pouw Kraan, T.C., van Gaalen, F.A., Huizinga, T.W., Pieterman, E., Breedveld, F.C. and Verweij, C.L. (2003a). Discovery of distinctive gene expression profiles in rheumatoid synovium using cDNA microarray technology: Evidence for the existence of multiple pathways of tissue destruction and repair. *Genes Immun* 4, 187–196.

van der Pouw Kraan, T.C., van Gaalen, F.A., Kasperkovitz, P.V., Verbeet, N.L., Smeets, T.J., Kraan, M.C., Fero, M., Tak, P.P., Huizinga, T.W., Pieterman, E. et al. (2003b). Rheumatoid arthritis is a heterogeneous disease: Evidence for differences in the activation of the STAT-1 pathway between rheumatoid tissues. *Arthritis Rheum* 48, 2132–2145.

van der Pouw Kraan, T.C., Wijbrandts, C.A., van Baarsen, L.G., Voskuyl, A.E., Rustenburg, F., Baggen, J.M., Ibrahim, S.M., Fero, M., Dijkmans, B.A., Tak, P.P. et al. (2007). Rheumatoid Arthritis subtypes identified by genomic profiling of peripheral

blood cells: Assignment of a type I interferon signature in a subpopulation of patients. *Ann Rheum Dis*.

van Jaarsveld, C.H., ter Borg, E.J., Jacobs, J.W., Schellekens, G.A., Gmelig-Meyling, F.H., van Booma-Frankfort, C., de Jong, B.A., van Venrooij, W.J. and Bijlsma, J.W. (1999). The prognostic value of the antiperinuclear factor, anti-citrullinated peptide antibodies and rheumatoid factor in early rheumatoid arthritis. *Clin Exp Rheumatol* 17, 689–697.

van Zeben, D., Hazes, J.M., Zwinderman, A.H., Cats, A., van der Voort, E.A. and Breedveld, F.C. (1992). Clinical significance of rheumatoid factors in early rheumatoid arthritis: Results of a follow up study. *Ann Rheum Dis* 51, 1029–1035.

Vang, T., Congia, M., Macis, M.D., Musumeci, L., Orru, V., Zavattari, P., Nika, K., Tautz, L., Tasken, K., Cucca, F. et al. (2005). Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant. *Nat Genet* 37, 1317–1319.

Viken, M.K., Amundsen, S.S., Kvien, T.K., Boberg, K.M., Gilboe, I.M., Lilleby, V., Sollid, L.M., Forre, O.T., Thorsby, E., Smerdel, A. et al. (2005). Association analysis of the 1858C > T polymorphism in the PTPN22 gene in juvenile idiopathic arthritis and other autoimmune diseases. *Genes Immun* 6, 271–273.

Waldron-Lynch, F., Adams, C., Amos, C., Zhu, D.K., McDermott, M.F., Shanahan, F., Molloy, M.G. and O'Gara, F. (2001). Tumour necrosis factor 5′ promoter single nucleotide polymorphisms influence susceptibility to rheumatoid arthritis (RA) in immunogenetically defined multiplex RA families. *Genes Immun* 2, 82–87.

Walsh, E.C., Mather, K.A., Schaffner, S.F., Farwell, L., Daly, M.J., Patterson, N., Cullen, M., Carrington, M., Bugawan, T.L., Erlich, H. et al. (2003). An integrated haplotype map of the human major histocompatibility complex. *Am J Hum Genet* 73, 580–590.

Wesoly, J., van der Helm-van Mil, A.H., Toes, R.E., Chokkalingam, A.P., Carlton, V.E., Begovich, A.B. and Huizinga, T.W. (2005). Association of the PTPN22 C1858T single-nucleotide polymorphism with rheumatoid arthritis phenotypes in an inception cohort. *Arthritis Rheum* 52, 2948–2950.

Wilder, R.L. (1996). Adrenal and gonadal steroid hormone deficiency in the pathogenesis of rheumatoid arthritis. *J Rheumatol Suppl* 44, 10–12.

Wolfe, F., Kleinheksel, S.M. and Khan, M.A. (1988). Prevalence of familial occurrence in patients with rheumatoid arthritis. *Br J Rheumatol* 27(Suppl 2), 150–152.

Yang, H., Rittner, H., Weyand, C.M. and Goronzy, J.J. (1999). Aberrations in the primary T-cell receptor repertoire as a predisposition for synovial inflammation in rheumatoid arthritis. *J Investig Med* 47, 236–245.

Zanelli, E., Huizinga, T.W., Guerne, P.A., Vischer, T.L., Tiercy, J.M., Verduyn, W., Schreuder, G.M., Breedveld, F.C. and de Vries, R.R. (1998). An extended HLA-DQ-DR haplotype rather than DRB1 alone contributes to RA predisposition. *Immunogenetics* 48, 394–401.

Zanelli, E., Jones, G., Pascual, M., Eerligh, P., van der Slik, A.R., Zwinderman, A.H., Verduyn, W., Schreuder, G.M., Roovers, E., Breedveld, F.C. et al. (2001). The telomeric part of the HLA region predisposes to rheumatoid arthritis independently of the class II loci. *Hum Immunol* 62, 75–84.

Zhernakova, A., Eerligh, P., Wijmenga, C., Barrera, P., Roep, B.O. and Koeleman, B.P. (2005). Differential association of the PTPN22 coding variant with autoimmune diseases in a Dutch population. *Genes Immun* 6, 459–461.

CHAPTER

Genomic Evaluation of Multiple Sclerosis

Francisco J. Quintana and Howard L. Weiner

INTRODUCTION

Multiple sclerosis (MS) is an autoimmune disorder in which the central nervous system (CNS) is targeted by the dysregulated activity of the immune system, resulting in focal lesions and progressive neurological dysfunction. MS is heterogeneous in its clinical symptoms, rate of progression and response to therapy, probably reflecting the existence of several pathogenic mechanisms that make different contributions to the disease. MS is a T cell-mediated autoimmune disease thought to result from a combination of genetic and environmental factors (Weiner, 2004). In this chapter we will analyze the contribution of genomics, transcriptomics, immunomics and proteomics in delineating these factors, as well as their utility for monitoring disease progression and response to therapy, identifying new targets for therapeutic intervention and detecting individuals at risk of developing the disease later on in life.

GENOMICS IN MS

The first observation suggesting a genetic contribution to MS susceptibility was the identification of familial aggregation: first, second and third degree relatives of MS patients have an increased risk of developing the disease (Mackay, 1950); a sibling of an MS patient, for example, has a 20-times greater lifetime risk of developing MS than an individual from the general population (Sawcer et al., 2005). Further studies have showed that familial aggregation in MS results from sharing predisposing genetic elements and not from the exposure to environmental factors (Dyment et al., 2004).

MS is considered a complex genetic disease in which many polymorphic genes have small or at most moderate effects on the overall MS risk, disease severity, rate of progression and age of onset among several clinical outcomes. To date, the major histocompatibility complex (MHC) locus on chromosome 6p21 MHC remains the strongest and most convincingly chromosomal region linked to MS: The DRB1*15 and the CRB1*17 alleles increase the risk for MS while the CRB1*14 has a disease-protective effect (Games, 2003; Jersild et al., 1972; Oksenberg and Barcellos, 2005). Several non-MHC candidate loci have also been linked to MS (Oksenberg and Barcellos, 2005), but it has proven difficult to validate their association in independent studies. The difficulty in the identification of non-MHC genes associated to MS might result from the genetic heterogeneity existing among MS patients, meaning that different combinations of genes might lead to the same end phenotype: MS. In this scenario, methods like linkage analysis might not be sensitive enough for the detection of genes bearing only modest effects on MS susceptibility, and thus association studies in large cohorts of patients and controls might be needed.

Briefly, two approaches are used for the identification of genes linked to MS pathogenesis and progression: linkage and association mapping.

Linkage mapping is based on the study of the co-inheritance of genetic markers and phenotypes in families over several generations. Linkage mapping is successful in finding genes for rare

Copyright © 2010, Elsevier Inc.
All rights reserved.

Mendelian monogenic diseases inherited in a dominant fashion. However, in diseases like MS where several loci have a small contribution to the phenotype under study, linkage studies only identify those loci that have the strongest influence. In addition, one of the premises implicit in linkage studies is that all the families studied have their susceptibility determined by the same genes, an assumption at odds with mounting evidence suggesting that the susceptibility to MS is genetically heterogeneous.

A recent study used 4506 genetic markers to analyze 2692 individuals in 730 families of Northern European descent looking for the co-inheritance of genetic markers and MS (Sawcer et al., 2005). Multipoint non-parametric linkage analysis could only find one significant linkage, which unsurprisingly pointed to the MHC locus on chromosome 6p21. This study therefore confirms the identification of the MHC locus as a genetic determinant of the susceptibility to MS, but it also highlights two limitations of linkage mapping. First, linkage mapping tends to identify large chromosomal fragments because of the small number of recombination events analyzed in familial pedigrees; the MHC locus, for example, contains more than 200 genes. Second, the authors also found suggestive linkage on chromosomes 17q23 and 5q33 and 19p13, but the data were inconclusive about these loci even when a high density of genetic markers was used. Other mapping strategies are needed to identify loci with modest effects on MS.

Association mapping looks for genetic markers with higher frequencies in MS patients than controls, suggesting an association between a disease phenotype and allelic variation (Hafler and De Jager, 2005).

Although genetic susceptibility to MS has been linked to the MHC locus, the particular gene or genes underlying susceptibility to MS within this locus are a matter of discussion, particularly because of the punctuated pattern of linkage disequilibrium observed for this chromosomal location (Jeffreys et al., 2001). As a result, other genes within the MHC locus besides HLA-DRB1 might be associated with MS; *tnf* and other loci in HLA class III and HLA class I are logical candidates. Lincoln and coworkers genotyped 4203 individuals from Finland and Canada with a high-density SNP panel to identify the gene or genes in the MHC locus responsible for the increased susceptibility to MS (Lincoln et al., 2005). Their results identified the DRB1 allele in the HLA class II region as the single major susceptibility locus in the MHC region, although there is still a small chance that a closely adjacent locus contains the true susceptibility variant. Thus, association mapping confined the MS susceptibility to the approximately 100 kb of sequence flanking HLA-DRB1. On the one hand, this work illustrates the increased power of association over linkage mapping studies. On the other hand, it highlights the dependency of association mapping on linkage disequilibrium: the uneven distribution of recombination events in this region leads to the identification of a locus for MS susceptibility big enough to contain several genes.

Admixture mapping is an association mapping strategy based on: (a) a difference on the prevalence of a disease between two ethnic groups and (b) the existence of a third admixed population (Smith and O'Brien, 2005). Northern Europeans and Africans have different susceptibilities to many autoimmune, circulatory and metabolic disorders, among them MS (Smith and O'Brien, 2005). This differential susceptibility is also reflected in North America where MS is more prevalent in European Americans than African Americans. On average, there have been only six generations since African and European populations came into contact in North America, resulting in little recombination between chromosomes of African and European ancestry in the history of African American populations. Thus, the chromosomal segments of one or the other ancestry are long in the admixed population, and few genetic markers can be used to classify the genome of an African American into sections with African or European origins. It is therefore possible to study African Americans to identify genomic regions where individuals with MS tend to have an unusually high proportion of ancestry from either Europeans or Africans, indicative of the presence of an MS variant that differs in frequency between the two ancestral populations. This mapping strategy was initially suggested several years ago (Chakraborty and Weiss, 1988), but it could only be fully implemented with the recent advent of genetic markers that identify human populations of different ancestry (Sachidanandam et al., 2001).

Admixture mapping led to the identification of a new locus associated with MS risk around the chromosome 1 centromere as a result of the analysis of 1166 genetic markers in 1648 samples (605 MS patients and 1043 controls) (Reich et al., 2005). Notably, this linkage could not be replicated using an independent set of 143 African Caribbeans with MS from Martinique and the United Kingdom, suggesting that either the locus in chromosome 1 does not have a role in African Caribbean populations or that the African Caribbean cohort was not large enough. Future studies should attempt to clone the susceptibility gene on chromosome 1.

Genome scans aimed at identifying non-MHC genes linked to MS usually fail when initial promising candidates cannot be validated on independent populations. New experimental techniques and analytical methods, however, are trying to change this trend. A recently published study combined the results of whole genome screens for linkage or association in 18 populations and superimposed them in a combined genomic map (Abdeen et al., 2006). The regions identified by this meta-analysis were then verified in a different set of samples (Abdeen et al., 2006), leading to the identification of several non-MHC candidate genes that modify the risk for MS. The list of candidates includes genes involved in CNS development and regeneration (NTN1, NCAM1, ADAM22 and ADAMTS10) in addition to genes directly linked to inflammation (TGFA, TGFBR, IL18 and IL10RA) (Abdeen et al., 2006). Although these new gene candidates are still awaiting independent validation, the meta-analysis used for their identification might constitute a new method for shortlisting new candidates to be targeted in independent replication studies.

TRANSCRIPTOMICS IN MS

Large-scale studies of mRNA expression in MS have been directed at characterizing the two main components of the disease: the lesion and the immune response.

Characterization of the Lesion

The hallmark of MS is the presence of demyelinated plaques in the white matter area of the CNS (Lassmann, 1998). These lesions are heterogeneous, reflecting the contribution of diverse mechanisms to disease progression (Lassmann et al., 2001). The study of the transcriptional activity in the lesions is therefore of importance for the characterization of the processes that drive MS, and for the identification of new targets of immune intervention.

Lindberg and coworkers studied the transcriptional activity of lesions and normal appearing white matter (NAWM) in samples taken from 6 secondary progressive MS (SPMS) patients and 12 matched controls (Lindberg et al., 2004). Within the lesion, only 21% of the upregulated genes were associated to the immune response; 77% of those immune-related transcripts corresponded to the cellular response. However, the Four most significant immune upregulated genes were linked to humoral immunity (immunoglobulins). Notably, tertiary lymph nodes that support the maturation of antibody-secreting plasma cells within the CNS have been identified in SPMS brains (Aloisi and Pujol-Borrell, 2006; Corcione et al., 2005). In addition, those genes that showed a decreased expression mainly belonged to one of two categories: genes with well-known anti-inflammatory activities (pointing to a deficit in immunoregulatory mechanisms), and genes involved in neural homeostasis. Interestingly, the NAWM showed a significant upregulation in immune-related genes, mainly involved in signaling and effector functions like blood–brain barrier (BBB) disruption and lymphocyte activation. These data provide a molecular insight into the pathological mechanisms driving MS and confirm the presence of physiological abnormalities on NAWM.

The study of the transcriptional profile in the MS lesion can lead to the identification of new targets for immuno intervention. Lock and coworkers studied the expression profile in the acute MS lesions (with signs of inflammation) and in silent lesions (without inflammation but showing clear signs of demyelination and scarring) (Lock et al., 2002). Both types of lesions showed an upregulated expression of genes associated with MHC class II antigen presentation, immunoglobulin synthesis, complement and pro-inflammatory cytokines. Neuron-associated genes and those associated with myelin production were underexpressed.

Different expression profiles were found in active and silent lesions. α-integrin was found to be elevated in chronic silent MS lesions; notably antibodies to α4-integrin reverse and reduce the rate of relapse in relapsing-remitting experimental autoimmune encephalomyelitis (EAE) (Yednock et al., 1992), and a humanized version of this antibody showed promising effects in the treatment of human MS (Polman et al., 2006).

The expression of IgE and IgG Fc receptors was upregulated in chronic silent MS lesions (Lock et al., 2002). Accordingly, mice harboring impaired IgE and IgG Fc receptors develop a milder EAE than their wild type counterparts (Lock et al., 2002). These effects were stronger from day 20 onwards following the induction of the disease, in accordance with microarray data showing that Fc receptor transcripts are elevated in chronic but not acute MS lesions. Granulocyte colony-stimulating factor (GCSF) was found to be upregulated in acute active lesions; its administration to mice before challenge with an encephalopathogenic peptide also leads to an amelioration of the disease, suggesting that endogenous GCSF might participate of the natural regulation of acute attacks (Lock et al., 2002).

In a separate study, the large-scale sequencing of non-normalized cDNA libraries derived from plaques dissected from brains of patients with MS showed an increased frequency of transcripts coding for osteopontin (OPN), a Th1 cytokine involved in the immune response to infectious diseases (Chabas et al., 2001). OPN transcripts were detected exclusively in the MS mRNA population but not in control brain mRNA. The upregulated expression of OPN was confirmed by immuno-histochemistry in human MS plaques, which showed OPN expression in microvascular endothelial cells and macrophages, and in the white matter areas adjacent to plaques. Similar patterns of expression were seen in mice and rats representing a model of relapsing remitting and monophasic EAE, respectively. To test the relevance of OPN in MS, OPN-deficient mice were generated. These mice showed a reduced severity in EAE. Moreover, neutralization of OPN with neutralizing antibodies also led to an amelioration of EAE (Blom et al., 2003), validating the use of cDNA microarrays in the search for new therapeutic targets for MS. The upregulation of OPN levels in MS plaques (Tajouri et al., 2005) and in the circulation (Comabella et al., 2005; Vogt et al., 2003, 2004) of MS patients was replicated in independent studies, prompting the search for polymorphisms in the *opn* gene associated with MS. Although some controversy still remains (Caillier et al., 2003) polymorphisms in the *opn* locus have been associated with increased levels of circulating OPN and disease course (Chiocchetti et al., 2005). OPN is therefore an example of how results obtained in transcriptomics studies might lead to the identification of genetic polymorphisms linked to MS.

Characterization of the Immune Response

The analysis of the transcriptional profile could also be applied to study the immune response in MS, to monitor the progression of the disease and the response to therapy. Two points, however, should be kept in mind when considering the use of cDNA arrays for the analysis of the immune response in MS patients: First, these studies assume that changes in the periphery somehow reflect the ongoing situation within the CNS. Second, these studies are limited by the normal "noise" that exists in basal gene expression, originated from diverse factors such as the relative proportion of different blood cell subsets, gender, age and the time of day at which the sample was taken (Whitney et al., 2003). This variation is likely to impose severe limitations for the use of cDNA microarrays in a clinical setting.

Follow up of Disease Activity

A recent study by Achiron and co-workers followed the transcriptional activity of peripheral blood mononuclear cells (PBMC) prepared from relapsing-remitting MS (RRMS) patients during the relapses and remission of MS (Achiron et al., 2004). The authors identified a transcriptional signature associated to the relapse, that included genes involved in the recruitment of immune cells, epitope spreading and escape from immune-regulation. Although encouraging, these results should be validated using an independent set of samples and in longitudinal studies to asses their predictive value.

Response to Therapy

Gene expression profiling can also be used to classify patients according to different clinical criteria, such as responder or non-responders to therapy. β−Interferon (β−IFN) is widely used for the treatment of MS (Kappos and Hartung, 2005), however, its precise mechanism of action is not known, and neither are biomarkers that would allow the identification of patients that will benefit from it. Weinstock-Guttman and colleagues used cDNA microarrays to study the effects of IFNβ therapy on the transcriptional activity of monocyte-depleted PBMC (Weinstock-Guttman et al., 2003). This pharmaco-dynamic study found that upon 1 h of IFNβ administration, significant changes are detected in the expression of genes involved in the antiviral response, β−IFN signaling and markers of lymphocyte activation. These studies provided a molecular description of the effects of β−IFN on RRMS patients and were later on extended to identify transcriptional signatures associated with a favorable response to treatment with IFNβ (Sturzebecher et al., 2003). Moreover, they demonstrated that MS patients showing a positive response to treatment with β−IFN, as assessed by longitudinal gadolinium-enhanced MRI scans and clinical disease activity, are characterized by specific patterns of gene expression (Sturzebecher et al., 2003). Based on these observations, Oksenberg and coworkers identified groups of genes whose expression has a prognostic value for the identification of MS patients likely to respond to treatment with β−IFN (Baranzini et al., 2005). The work of Oksenberg and coworkers is remarkable for two reasons: First, it demonstrates that gene expression profiling can be used in the management of MS to select a therapeutic regime suited to the patient's metabolism. Second, it used a methodology (RT-PCR) accessible to clinical laboratories, facilitating the translation of their results into daily medical practice.

The combination of the data generated in transcriptomics and genomics studies can be an invaluable source of information and new hypotheses. Aune et al. compared the genes differentially expressed by lymphocytes in rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), insulin-dependent diabetes mellitus (IDDM) and MS, concluding that they are clustered within chromosomal domains in the genome (Aune et al., 2004). Strikingly, they found that the chromosomal domains containing the genes differentially expressed in autoimmune disorders could be mapped to disease susceptibility loci associated to those diseases by genetic linkage studies (Aune et al., 2004). These results suggest that the expression of disease-associated genes is co-regulated as a result of shared genetic regulatory elements or local patterns of chromatin condensation. Recently Baranzini and coworkers studied the genetic concordance between gene expression and genetic linkage in MS (Fernald et al., 2005). They first compiled the data on gene expression available for MS and EAE, and superimposed it with the all the known susceptibility loci identified in MS and EAE. In their study, Baranzini and coworkers identified the MS susceptibility genes located in the MHC locus as overlapping with clusters of differentially expressed genes in MS and murine EAE. However, they could also identify an interesting region on chromosome X that might contribute to the sexual dimorphism observed in MS. The integration of the data generated by different platforms, like transcriptomics, genomics and proteomics, is therefore likely to deepen our understanding of the mechanisms driving MS.

IMMUNOMICS IN MS

The autoimmune nature of MS suggests that the study of the immune response should be useful for the early diagnosis, prognosis and monitoring of MS patients. With this aim, new tools have been developed, allowing a high-throughput analysis of the T cell- and antibody-mediated immune response.

T Cell Response

MHC arrays have been recently developed based on the immobilization of peptide/MHC tetramers used to activate peptide-specific T cells and follow cytokine secretion or adhesion (Chen et al., 2005; Soen et al., 2003). Their use for the study of human immunology is limited by the high frequency of polymorphisms existing on the MHC locus. Nevertheless, these arrays have been recently used to characterize the cellular response in tumor-vaccinated humans (Chen et al., 2005).

Reverse phase arrays consist of arrays of spotted lysates prepared from primary cells, which are interrogated with antibodies specific for phosphorylated or de-phosphorylated proteins involved in signaling transduction pathways of interest (Chan et al., 2004). By combining different antibodies and inhibitors of specific signal transduction pathways, detailed signaling maps can be constructed describing the molecular events that lead to the activation of a specific cell population. This approach have been recently be used to study $CD4^+CD25^+$ regulatory T cells (Treg) (Chan et al., 2004). Since defects in Treg function have been described in MS patients (Viglietta et al., 2004), these arrays could be useful in the identification of signaling defects in the immune system of MS patients, and how these pathways are modified in immuno-modulatory regimes.

Multidimensional FACS is based on the simultaneous computation of the signals produced by more than 16 colors, and their analysis using advance mathematical algorithms (Perfetto et al., 2004). This technique could be useful to interrogate the immune response to specific antigens using a combination of tetramers, intracellular cytokine staining and surface markers (Chattopadhyay et al., 2005).

Cell migration arrays (Kuschel et al., 2006). To reach the CNS and cause tissue destruction, autoimmune T-cells have to express surface molecules that allow them to cross the BBB and initiate inflammation (Luster et al., 2005). The importance of this process for MS progression is shown by the beneficial effects that natalizumab, a humanized antibody to α4 integrin that interferes with T cell extravasation to the CNS and has showed promising preliminary results in the treatment of MS (Polman et al., 2006). Thus, the characterization of the adhesion properties exhibited by the different cell populations in MS patients might be of help to predict and monitor response to treatment and predict and prevent the development of relapses.

B Cell Response

Antibodies are of interest in MS because they might have a pathological role in MS (Genain et al., 1995), and more importantly, because antibody responses are thought to reflect the activity of the T cell compartment (Robinson et al., 2002b). These antibodies can be produced by B cells in the periphery and make their way to the inflamed CNS via the disrupted BBB, but they are also produced by intrathecal germinal centers that seem to form tertiary lymph nodes (Aloisi and Pujol-Borrell, 2006; Corcione et al., 2005). It is easier to assay antibody reactivity than to follow antigen-specific T-cell responses, thus new efforts have been invested in the development of new technologies for monitoring the humoral response in MS patients and autoimmunity (Quintana et al., 2004; Robinson et al., 2002a).

Antigen arrays can be used to detect changes in the repertoire of antibodies reflecting the antigen spreading that accompanies EAE progression (Robinson et al., 2003). The information obtained about the degree of antigen spreading observed in each mouse was used to design tailored immunomodulatory vaccines to control EAE (Robinson et al., 2003). In humans, have been used to identify new lipid targets of antibodies present in the CSF of MS patients; some of the new lipid targets found were also validated on EAE (Kanter et al., 2006). Future experiments should study the antibody response in the serum of MS patients, searching for patterns of antibody reactivity that predict the progression of MS or the response to therapy, as has been shown for other autoimmune disorders, such as RA (Hueber et al., 2005), autoimmune diabetes (Quintana et al., 2004) and SLE (Li et al., 2005). Antigen arrays have also been shown to identify the individual mice that will develop autoimmune diabetes later in life by studying an experimental model that shows incomplete penetrance. Thus, antigen arrays might be useful for the identification of patients at risk of developing MS, before the overt onset of the symptoms (Quintana et al., 2004). In addition, antigen arrays might be used to interrogate the antibody repertoire in the search of new targets of the MS autoimmune attack and therefore, new targets for immuno-modulation. Indeed, we have preliminary data suggesting that antigen arrays might be used to identify antibody patterns linked to the different forms of MS and the pathology at the site of the lesion (Quintana and Weiner, unpublished results).

PROTEOMICS IN MS

Proteomic studies in MS can identify new targets of the autoimmune process complementing the information provided by antigen array; they can also identify new processes contributing to disease pathology, and biomarkers for the early diagnosis and monitoring of MS patients.

Identification of New Targets of Autoimmunity

Using 2D gels on brain extracts, Almeras and coworkers identified 14 new targets recognized by antibodies present in CSF (Almeras et al., 2004). The targets identified by this study included heat shock proteins, structural proteins and enzymes involved in glucose metabolism (Almeras et al., 2004); although these results have been partially validated in the experimental model of MS EAE (Zephir et al., 2006), they are still waiting validation in an independent patient cohort.

Identification of New Pathogenic Processes

Proteomic studies have also identified new mechanisms that contribute to MS pathology. The existence of a link between Epstein-Barr Virus infection (EBV) and MS has been recently strengthened by the work of Cepok and co-workers, who used protein expression arrays to characterize the reactivity of antibodies in the CSF of MS patients, most of those antibodies recognized EBV epitopes (Cepok et al., 2005). These results suggest that EBV reactivation might elicit and abnormal immune response in susceptible individuals that contributes to MS (Sundstrom et al., 2004).

Brinkmeier and coworkers analyzed the CSF of MS and Guillain-Barre Syndrome (GBS) patients to characterize molecules that might affect ion-channel function (Brinkmeier et al., 2000). They identified and endogenous pentapetide (QYNAD) that works as a reversible Na^+ channel blocker (Brinkmeier et al., 2000). This pentapetide is present in the CSF of healthy individuals, but its levels are upregulated 3–14-fold in MS and GBS patients. This blocking peptide might be involved in the fast exacerbations and relapses commonly seen in demyelinating autoimmune diseases. Moreover, it might become a valuable marker of disease activity, and the target of future therapeutic interventions. These observations, however, could not be replicated by independent researchers (Cummins et al., 2003), highlighting the need for independent validation of proteomic findings.

Identification of Biomarkers

Proteomics can identify biomarkers useful in the monitoring of the different processes that contribute to MS pathology. In this direction, several studies have suggested that cytokines, chemokines, complement and adhesion molecules can be used as indicators of the inflammatory process in MS; while the levels of actin, tubulin, neurofilaments, tau, GFAP and S-100 proteins can be takes as indicators of axonal loss and gliosis (reviewed in Bielekova and Martin, 2004; Miller, 2004; Teunissen et al., 2005). However, for these to be of widespread clinical use, these markers should be easy to collect and provide reproducible results. This

motivation has boosted efforts aimed at finding markers of inflammation and neurodegeneration detectable in blood (Bielekova and Martin, 2004), urine (Bielekova and Martin, 2004) and tears (Devos et al., 2001); and simultaneously it has fostered the development of new technologies aimed at detecting minute amounts of proteins in body fluids (Zhang et al., 2006).

2009 UPDATE

Polymorphisms in the α chain of the IL-7 receptor (IL-7Ra) have been recently associated with MS (Gregory et al., 2007; Hafler et al., 2007; Lundmark et al., 2007). Although these polymorphisms make only a small contribution to the overall genetic susceptibility to MS, they are a significant step towards the identification of genetic determinants for MS outside the MHC locus. The IL-7Ra allele associated with MS favors a relative decrease in the membrane-bound IL-7Ra (Gregory et al., 2007). IL-7 is produced by stromal cells in lymphoid tissues and its availability is controlled through its uptake by the membrane-bound IL-7R on T cells (Mazzucchelli and Durum, 2007). Thus, considering the positive effects that IL-7 has on lymphocyte survival and proliferation (Mazzucchelli and Durum, 2007), the decrease in membrane IL-7R might result in increased levels of IL-7 available to fuel the inflammatory T cell response in MS.

The α chain IL-2 receptor (IL-2Ra) gene has also been recently linked to MS (Hafler et al., 2007). IL-2Ra allelic variation has been previously associated to other autoimmune diseases such as type I diabetes, but at a different genomic position (Lowe et al., 2007). IL-2 is required for the development of regulatory T cells (T_{reg}) (Malek, 2008; Suzuki et al., 1995) and, indeed, deficits in T_{reg} activity characterize relapsing–remitting MS (Viglietta et al., 2004); thus the IL2Ra polymorphisms might be related to the immune dysregulation observed in MS. Notably, IL2Ra-specific antibodies have shown promising beneficial effects for the treatment of MS on phase 2 clinical trials (Bielekova et al., 2004; Rose et al., 2004). Although the link between IL-2Rapolymorphisms and MS is still awaiting further validation, the association of IL-7Raand IL-2Ra variants to MS supports the use of genome-wide studies to delineate pathways contributing to disease pathogenesis.

The transcriptional profiling of $CD4^+$ T cells in MS has been recently used to predict the conversion of clinically isolated syndrome (CIS) patients to definite MS (Corvol et al., 2008). Baranzini and coworkers identified a set of genes whose expression is linked to the conversion to definite MS in less than a year. One of those genes is TOB1, a critical regulator of cell proliferation as a potential player in MS pathogenesis. Decreased TOB1 expression at the RNA and protein levels was found in MS patients and also in EAE. Moreover, a genetic association was observed between TOB1 variation and MS progression. All in all, these results link alterations in the regulation of $CD4^+$T cell quiescence to the progression of MS (Corvol et al., 2008).

A large proteomic study of MS lesions by Steinman and coworkers has recently identified the dysregulation of the clotting cascade as a new player in MS pathogenesis (Han et al., 2008). *In vivo* administration of hirudin or recombinant activated protein C reduced disease severity in EAE and suppressed T_H1 and T_H17 cytokines in astrocytes and immune cells (Han et al., 2008). Administration of mutant forms of recombinant activated protein C showed that both its anticoagulant and its signaling functions were essential for optimal amelioration of EAE (Han et al., 2008). Thus, the coagulation cascade might constitute a new therapeutic target for MS.

Using antigen arrays, we characterized patterns of antibody reactivity in MS serum against a panel of CNS protein, and lipid autoantigens, and heat shock proteins (Quintana et al., 2008b). We found unique autoantibody patterns that distinguished RRMS, SPMS and primary progressive (PPMS) MS from both healthy controls and other neurologic or autoimmune driven diseases, including Alzheimer's disease, adrenoleukodystropy and lupus erythematosus. RRMS was characterized by autoantibodies to heat shock proteins that were not observed in PPMS or SPMS. In addition, RRMS, SPMS and PPMS were characterized by unique patterns of reactivity to CNS antigens (Quintana et al., 2008b). Furthermore, we examined sera from patients with different immunopathologic patterns of MS as determined by brain biopsy, and we identified unique antibody patterns to lipids and CNS-derived peptides that were linked to each type of pathology (Quintana et al., 2008b). A further characterization of the role of these lipids in EAE identified them as direct effectors in MS pathogenesis by the activation of the innate immune system in the CNS (Quintana et al., 2008b). The demonstration of unique serum immune signatures linked to different stages and pathologic processes in MS provides an avenue to monitor MS and to characterize immunopathogenic mechanisms and therapeutic targets in the disease.

Screenings aimed at identifying genes or drugs controlling the immune response cannot be easily undertaken in mice because they are based on crossing, maintaining and screening large numbers of animals, an expensive time-, space- and labor-intensive task; new experimental models are needed. The zebrafish (*Daniorerio*) harbors an adaptive immune system that resembles the mammalian immune system (Langenau and Zon, 2005) and offers several experimental advantages for the study of pathways controlling vertebrate processes of interest. As part of our work on the zebrafish to identify pathways controlling immunity, we have identified the ligand-activated transcription

factor aryl hydrocarbon receptor (AHR) as a regulator of the generation of T_{reg} and T_H17 cells (Quintana et al., 2008a). AHR activation by its ligand 2, 3, 7, 8-tetrachlorodibenzo-p-dioxin (TCDD) results in the generation of functional T_{reg} that inhibit the development of EAE by a TGFβ1-dependent mechanism. Surprisingly, AHR activation by an alternative ligand, 6-formylindolo [3, 2-b]carbazole, interferes with T_{reg} differentiation, boosts T_H17 differentiation and worsens EAE. Thus, AHR regulates both T_{reg} and T_H17 differentiation in a ligand-specific fashion, constituting a unique target for therapeutic immunomodulation. Moreover, our findings suggest that the experimental advantages offered by the zebrafish can be exploited to characterize metabolic pathways controlling immunity in vertebrates and to identify new targets for therapeutic intervention..

2009 UPDATE REFERENCES

Bielekova, B., Richert, N., Howard, T., Blevins, G., Markovic-Plese, S., McCartin, J., Frank, J.A., Wurfel, J., Ohayon, J., Waldmann, T.A., et al. (2004). Humanized anti-CD25 (daclizumab) inhibits disease activity in multiple sclerosis patients failing to respond to interferon beta. *Proc Natl Acad Sci USA* 101, 8705–8708.

Corvol, J.C., Pelletier, D., Henry, R.G., Caillier, S.J., Wang, J., Pappas, D., Casazza, S., Okuda, D.T., Hauser, S.L., Oksenberg, J.R., et al. (2008). Abrogation of T cell quiescence characterizes patients at high risk for multiple sclerosis after the initial neurological event. *Proc Natl Acad Sci USA* 105, 11839-11844.

Gregory, S.G., Schmidt, S., Seth, P., Oksenberg, J.R., Hart, J., Prokop, A., Caillier, S.J., Ban, M., Goris, A., Barcellos, L.F., et al. (2007). Interleukin 7 receptor alpha chain (IL7R) shows allelic and functional association with multiple sclerosis. *Nat Genet* 39, 1083–1091.

Hafler, D.A., Compston, A., Sawcer, S., Lander, E.S., Daly, M.J., De Jager, P.L., de Bakker, P.I., Gabriel, S.B., Mirel, D.B., Ivinson, A.J., et al. (2007). Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med* 357, 851–862.

Han, M.H., Hwang, S.I., Roy, D.B., Lundgren, D.H., Price, J.V., Ousman, S.S., Fernald, G.H., Gerlitz, B., Robinson, W.H., Baranzini, S.E., et al. (2008). Proteomic analysis of active multiple sclerosis lesions reveals therapeutic targets. *Nature* 451, 1076–1081.

Langenau, D.M. and Zon, L.I. (2005). The zebrafish: A new model of T-cell and thymic development. *Nat Rev Immunol* 5, 307-317.

Lowe, C.E., Cooper, J.D., Brusko, T., Walker, N.M., Smyth, D.J., Bailey, R., Bourget, K., Plagnol, V., Field, S., Atkinson, M., et al. (2007). Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the IL2RA region in type 1 diabetes. *Nat Genet* 39, 1074–1082.

Lundmark, F., Duvefelt, K., Iacobaeus, E., Kockum, I., Wallstrom, E., Khademi, M., Oturai, A., Ryder, L.P., Saarela, J., Harbo, H.F., et al. (2007). Variation in interleukin 7 receptor alpha chain (IL7R) influences risk of multiple sclerosis. *Nat Genet* 39, 1108–1113.

Malek, T.R. (2008). The biology of interleukin-2. *Annu Rev Immunol* 26, 453–479.

Mazzucchelli, R. and Durum, S.K. (2007). Interleukin-7 receptor expression: Intelligent design. *Nat Rev Immunol* 7, 144-154.

Quintana, F.J., Basso, A.S., Iglesias, A.H., Korn, T., Farez, M.F., Bettelli, E., Caccamo, M., Oukka, M. and Weiner, H.L. (2008a). Control of T(reg) and T(H)17 cell differentiation by the aryl hydrocarbon receptor. *Nature* 453, 65–71.

Quintana, F.J., Farez, M.F., Viglietta, V., Iglesias, A.H., Merbl, Y., Izquierdo, G., Lucas, M., Basso, A.S., Khoury, S.J., Lucchinetti, C.F., et al. (2008b). Antigen microarrays identify unique serum autoantibody signatures in clinical and pathologic subtypes of multiple sclerosis. *Proc Natl Acad Sci USA* 105, 18889–18894.

Rose, J.W., Watt, H.E., White, A.T. and Carlson, N.G. (2004). Treatment of multiple sclerosis with an anti-interleukin-2 receptor monoclonal antibody. *Ann Neurol* 56, 864–867.

Suzuki, H., Kundig, T.M., Furlonger, C., Wakeham, A., Timms, E., Matsuyama, T., Schmits, R., Simard, J.J., Ohashi, P.S., Griesser, H., et al. (1995). Deregulated T cell activation and autoimmunity in mice lacking interleukin-2 receptor beta. *Science* 268, 1472–1476.

Viglietta, V., Baecher-Allan, C., Weiner, H.L. and Hafler, D.A. (2004). Loss of functional suppression by CD^4 + CD^{25} + regulatory T cells in patients with multiple sclerosis. *J Exp Med* 199, 971–979.

REFERENCES

Abdeen, H., Heggarty, S., Hawkins, S.A., Hutchinson, M., McDonnell, G.V. and Graham, C.A. (2006). Mapping candidate non-MHC susceptibility regions to multiple sclerosis. *Genes Immun* 7, 494–502. Epub 2006 Jul 13.

Achiron, A., Gurevich, M., Friedman, N., Kaminski, N. and Mandel, M. (2004). Blood transcriptional signatures of multiple sclerosis: Unique gene expression of disease activity. *Ann Neurol* 55, 410–417.

Almeras, L., Lefranc, D., Drobecq, H., de Seze, J., Dubucquoi, S., Vermersch, P. and Prin, L. (2004). New antigenic candidates in multiple sclerosis: Identification by serological proteome analysis. *Proteomics* 4, 2184–2194.

Aloisi, F. and Pujol-Borrell, R. (2006). Lymphoid neogenesis in chronic inflammatory diseases. *Nat Rev Immunol* 6, 205–217.

Aune, T.M., Parker, J.S., Maas, K., Liu, Z., Olsen, N.J. and Moore, J.H. (2004). Co-localization of differentially expressed genes and shared susceptibility loci in human autoimmunity. *Genet Epidemiol* 27, 162–172.

Baranzini, S.E., Mousavi, P., Rio, J., Caillier, S.J., Stillman, A., Villoslada, P., Wyatt, M.M., Comabella, M., Greller, L.D., Somogyi, R. et al. (2005). Transcription-based prediction of response to IFNbeta using supervised computational methods. *PLoS Biol* 3, e2.

Bielekova, B. and Martin, R. (2004). Development of biomarkers in multiple sclerosis. *Brain* 127, 1463–1478.

Blom, T., Franzen, A., Heinegard, D. and Holmdahl, R. (2003). Comment on "The influence of the proinflammatory cytokine, osteopontin, on autoimmune demyelinating disease". *Science* 299, 1845. author reply 1845.

Brinkmeier, H., Aulkemeyer, P., Wollinsky, K.H. and Rudel, R. (2000). An endogenous pentapeptide acting as a sodium channel blocker in inflammatory autoimmune disorders of the central nervous system. *Nat Med* 6, 808–811.

Caillier, S., Barcellos, L.F., Baranzini, S.E., Swerdlin, A., Lincoln, R.R., Steinman, L., Martin, E., Haines, J.L., Pericak-Vance, M., Hauser, S.L. et al. (2003). Osteopontin polymorphisms and disease course in multiple sclerosis. *Genes Immun* 4, 312–315.

Cepok, S., Zhou, D., Srivastava, R., Nessler, S., Stei, S., Bussow, K., Sommer, N. and Hemmer, B. (2005). Identification of Epstein-Barr virus proteins as putative targets of the immune response in multiple sclerosis. *J Clin Invest* 115, 1352–1360.

Chabas, D., Baranzini, S.E., Mitchell, D., Bernard, C.C., Rittling, S.R., Denhardt, D.T., Sobel, R.A., Lock, C., Karpuj, M., Pedotti, R. et al. (2001). The influence of the proinflammatory cytokine, osteopontin, on autoimmune demyelinating disease. *Science* 294, 1731–1735.

Chakraborty, R. and Weiss, K.M. (1988). Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proc Natl Acad Sci USA* 85, 9119–9123.

Chan, S.M., Ermann, J., Su, L., Fathman, C.G. and Utz, P.J. (2004). Protein microarrays for multiplex analysis of signal transduction pathways. *Nat Med* 10, 1390–1396.

Chattopadhyay, P.K., Yu, J. and Roederer, M. (2005). A live-cell assay to detect antigen-specific $CD4^+$ T cells with diverse cytokine profiles. *Nat Med* 11, 1113–1117. Epub 2005 Sep 25.

Chen, D.S., Soen, Y., Stuge, T.B., Lee, P.P., Weber, J.S., Brown, P.O. and Davis, M.M. (2005). Marked differences in human melanoma antigen-specific T cell responsiveness after vaccination using a functional microarray. *PLoS Med* 2, e265.

Chiocchetti, A., Comi, C., Indelicato, M., Castelli, L., Mesturini, R., Bensi, T., Mazzarino, M.C., Giordano, M., D'Alfonso, S., Momigliano-Richiardi, P. et al. (2005). Osteopontin gene haplotypes correlate with multiple sclerosis development and progression. *J Neuroimmunol* 163, 172–178. Epub 2005 Apr 25.

Comabella, M., Pericot, I., Goertsches, R., Nos, C., Castillo, M., Blas Navarro, J., Rio, J. and Montalban, X. (2005). Plasma osteopontin levels in multiple sclerosis. *J Neuroimmunol* 158, 231–239.

Corcione, A., Aloisi, F., Serafini, B., Capello, E., Mancardi, G.L., Pistoia, V. and Uccelli, A. (2005). B-cell differentiation in the CNS of patients with multiple sclerosis. *Autoimmun Rev* 4, 549–554.

Cummins, T.R., Renganathan, M., Stys, P.K., Herzog, R.I., Scarfo, K., Horn, R., Dib-Hajj, S.D. and Waxman, S.G. (2003). The pentapeptide QYNAD does not block voltage-gated sodium channels. *Neurology* 60, 224–229.

Devos, D., Forzy, G., de Seze, J., Caillez, S., Louchart, P., Gallois, P. and Hautecoeur, P. (2001). Silver stained isoelectrophoresis of tears and cerebrospinal fluid in multiple sclerosis. *J Neurol* 248, 672–675.

Dyment, D.A., Ebers, G.C. and Sadovnick, A.D. (2004). Genetics of multiple sclerosis. *Lancet Neurol* 3, 104–110.

Fernald, G.H., Yeh, R.F., Hauser, S.L., Oksenberg, J.R. and Baranzini, S.E. (2005). Mapping gene activity in complex disorders: Integration of expression and genomic scans for multiple sclerosis. *J Neuroimmunol* 167, 157–169.

Games, (2003). A meta-analysis of whole genome linkage screens in multiple sclerosis. *J Neuroimmunol* 143, 39–46.

Genain, C.P., Nguyen, M.H., Letvin, N.L., Pearl, R., Davis, R.L., Adelman, M., Lees, M.B., Linington, C. and Hauser, S.L. (1995). Antibody facilitation of multiple sclerosis-like lesions in a nonhuman primate. *J Clin Invest* 96, 2966–2974.

Hafler, D.A. and De Jager, P.L. (2005). Applying a new generation of genetic maps to understand human inflammatory disease. *Nat Rev Immunol* 5, 83–91.

Hueber, W., Kidd, B.A., Tomooka, B.H., Lee, B.J., Bruce, B., Fries, J.F., Sonderstrup, G., Monach, P., Drijfhout, J.W., van Venrooij, W.J. et al. (2005). Antigen microarray profiling of autoantibodies in rheumatoid arthritis. *Arthritis Rheum* 52, 2645–2655.

Jeffreys, A.J., Kauppi, L. and Neumann, R. (2001). Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat Genet* 29, 217–222.

Jersild, C., Svejgaard, A. and Fog, T. (1972). HL-A antigens and multiple sclerosis. *Lancet* 1, 1240–1241.

Kanter, J.L., Narayana, S., Ho, P.P., Catz, I., Warren, K.G., Sobel, R.A., Steinman, L. and Robinson, W.H. (2006). Lipid microarrays identify key mediators of autoimmune brain inflammation. *Nat Med* 12, 138–143.

Kappos, L. and Hartung, H.P. (2005). 10 years of interferon beta-1b (Beta feron therapy. *J Neurol* 252(Suppl 3), iii1–iii2.

Kuschel, C., Steuer, H., Maurer, A.N., Kanzok, B., Stoop, R. and Angres, B. (2006). Cell adhesion profiling using extracellular matrix protein microarrays. *Biotechniques* 40, 523–531.

Lassmann, H. (1998). Neuropathology in multiple sclerosis: New concepts. *Mult Scler* 4, 93–98.

Lassmann, H., Bruck, W. and Lucchinetti, C. (2001). Heterogeneity of multiple sclerosis pathogenesis: Implications for diagnosis and therapy. *Trends Mol Med* 7, 115–121.

Li, Q.Z., Xie, C., Wu, T., Mackay, M., Aranow, C., Putterman, C. and Mohan, C. (2005). Identification of autoantibody clusters that best predict lupus disease activity using glomerular proteome arrays. *J Clin Invest* 115, 3428–3439.

Lincoln, M.R., Montpetit, A., Cader, M.Z., Saarela, J., Dyment, D.A., Tiislar, M., Ferretti, V., Tienari, P.J., Sadovnick, A.D., Peltonen, L. et al. (2005). A predominant role for the HLA class II region in the association of the MHC region with multiple sclerosis. *Nat Genet* 37, 1108–1112. Epub 2005 Sep 25.

Lindberg, R.L., De Groot, C.J., Certa, U., Ravid, R., Hoffmann, F., Kappos, L. and Leppert, D. (2004). Multiple sclerosis as a generalized CNS disease – comparative microarray analysis of normal appearing white matter and lesions in secondary progressive MS. *J Neuroimmunol* 152, 154–167.

Lock, C., Hermans, G., Pedotti, R., Brendolan, A., Schadt, E., Garren, H., Langer-Gould, A., Strober, S., Cannella, B., Allard, J. et al. (2002). Gene-microarray analysis of multiple sclerosis lesions yields new targets validated in autoimmune encephalomyelitis. *Nat Med* 8, 500–508.

Luster, A.D., Alon, R. and von Andrian, U.H. (2005). Immune cell migration in inflammation: Present and future therapeutic targets. *Nat Immunol* 6, 1182–1190.

Mackay, R.P. (1950). The familial occurrence of multiple sclerosis and its implications. *Ann Intern Med* 33, 298–320.

Miller, D.H. (2004). Biomarkers and surrogate outcomes in neurodegenerative disease: Lessons from multiple sclerosis. *NeuroRx* 1, 284–294.

Oksenberg, J.R. and Barcellos, L.F. (2005). Multiple sclerosis genetics: Leaving no stone unturned. *Genes Immun* 6, 375–387.

Perfetto, S.P., Chattopadhyay, P.K. and Roederer, M. (2004). Seventeen-colour flow cytometry: Unravelling the immune system. *Nat Rev Immunol* 4, 648–655.

Polman, C.H., O'Connor, P.W., Havrdova, E., Hutchinson, M., Kappos, L., Miller, D.H., Phillips, J.T., Lublin, F.D., Giovannoni, G., Wajgt, A.

et al. (2006). A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis. *N Engl J Med* 354, 899–910.

Quintana, F.J., Hagedorn, P.H., Elizur, G., Merbl, Y., Domany, E. and Cohen, I.R. (2004). Functional immunomics: Microarray analysis of IgG autoantibody repertoires predicts the future response of mice to induced diabetes. *Proc Natl Acad Sci USA* 101(Suppl 2), 14615–14621.

Reich, D., Patterson, N., De Jager, P.L., McDonald, G.J., Waliszewska, A., Tandon, A., Lincoln, R.R., DeLoa, C., Fruhan, S.A., Cabre, P. et al. (2005). A whole-genome admixture scan finds a candidate locus for multiple sclerosis susceptibility. *Nat Genet* 37, 1113–1118. Epub 2005 Sep 25.

Robinson, W.H., DiGennaro, C., Hueber, W., Haab, B.B., Kamachi, M., Dean, E.J., Fournel, S., Fong, D., Genovese, M.C., de Vegvar, H.E. et al. (2002a). Autoantigen microarrays for multiplex characterization of autoantibody responses. *Nat Med* 8, 295–301.

Robinson, W.H., Garren, H., Utz, P.J. and Steinman, L. (2002b). Millennium Award. Proteomics for the development of DNA tolerizing vaccines to treat autoimmune disease. *Clin Immunol* 103, 7–12.

Robinson, W.H., Fontoura, P., Lee, B.J., de Vegvar, H.E., Tom, J., Pedotti, R., DiGennaro, C.D., Mitchell, D.J., Fong, D., Ho, P.P. et al. (2003). Protein microarrays guide tolerizing DNA vaccine treatment of autoimmune encephalomyelitis. *Nat Biotechnol* 21, 1033–1039.

Sachidanandam, R., Weissman, D., Schmidt, S.C., Kakol, J.M., Stein, L.D., Marth, G., Sherry, S., Mullikin, J.C., Mortimore, B.J., Willey, D.L. et al. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409, 928–933.

Sawcer, S., Ban, M., Maranian, M., Yeo, T.W., Compston, A., Kirby, A., Daly, M.J., De Jager, P.L., Walsh, E., Lander, E.S. et al. (2005). A high-density screen for linkage in multiple sclerosis. *Am J Hum Genet* 77, 454–467. Epub 2005 Jul 29.

Shendure, J., Mitra, R.D., Varma, C. and Church, G.M. (2004). Advanced sequencing technologies: Methods and goals. *Nat Rev Genet* 5, 335–344.

Smith, M.W. and O'Brien, S.J. (2005). Mapping by admixture linkage disequilibrium: Advances, limitations and guidelines. *Nat Rev Genet* 6, 623–632.

Soen, Y., Chen, D.S., Kraft, D.L., Davis, M.M. and Brown, P.O. (2003). Detection and characterization of cellular immune responses using peptide-MHC microarrays. *PLoS Biol* 1, E65.

Sturzebecher, S., Wandinger, K.P., Rosenwald, A., Sathyamoorthy, M., Tzou, A., Mattar, P., Frank, J.A., Staudt, L., Martin, R. and McFarland, H.F. (2003). Expression profiling identifies responder and non-responder phenotypes to interferon-beta in multiple sclerosis. *Brain* 126, 1419–1429.

Sundstrom, P., Juto, P., Wadell, G., Hallmans, G., Svenningsson, A., Nystrom, L., Dillner, J. and Forsgren, L. (2004). An altered immune response to Epstein-Barr virus in multiple sclerosis: A prospective study. *Neurology* 62, 2277–2282.

Tajouri, L., Mellick, A.S., Tourtellotte, A., Nagra, R.M. and Griffiths, L.R. (2005). An examination of MS candidate genes identified as differentially regulated in multiple sclerosis plaque tissue, using absolute and comparative real-time Q-PCR analysis. *Brain Res Brain Res Protoc* 15, 79–91.

Teunissen, C.E., Dijkstra, C. and Polman, C. (2005). Biological markers in CSF and blood for axonal degeneration in multiple sclerosis. *Lancet Neurol* 4, 32–41.

Viglietta, V., Baecher-Allan, C., Weiner, H.L. and Hafler, D.A. (2004). Loss of functional suppression by $CD4^+$ $CD25^+$ regulatory T cells in patients with multiple sclerosis. *J Exp Med* 199, 971–979.

Vogt, M.H., Lopatinskaya, L., Smits, M., Polman, C.H. and Nagelkerken, L. (2003). Elevated osteopontin levels in active relapsing-remitting multiple sclerosis. *Ann Neurol* 53, 819–822.

Vogt, M.H., Floris, S., Killestein, J., Knol, D.L., Smits, M., Barkhof, F., Polman, C.H. and Nagelkerken, L. (2004). Osteopontin levels and increased disease activity in relapsing-remitting multiple sclerosis patients. *J Neuroimmunol* 155, 155–160.

Weiner, H.L. (2004). Multiple sclerosis is an inflammatory T-cell-mediated autoimmune disease. *Arch Neurol* 61, 1613–1615.

Weinstock-Guttman, B., Badgett, D., Patrick, K., Hartrich, L., Santos, R., Hall, D., Baier, M., Feichter, J. and Ramanathan, M. (2003). Genomic effects of IFN-beta in multiple sclerosis patients. *J Immunol* 171, 2694–2702.

Whitney, A.R., Diehn, M., Popper, S.J., Alizadeh, A.A., Boldrick, J.C., Relman, D.A. and Brown, P.O. (2003). Individuality and variation in gene expression patterns in human blood. *Proc Natl Acad Sci USA* 100, 1896–1901.

Yednock, T.A., Cannon, C., Fritz, L.C., Sanchez-Madrid, F., Steinman, L. and Karin, N. (1992). Prevention of experimental autoimmune encephalomyelitis by antibodies against alpha 4 beta 1 integrin. *Nature* 356, 63–66.

Zephir, H., Almeras, L., El Behi, M., Dussart, P., de Seze, J., Steibel, J., Trifilieff, E., Dubucquoi, S., Dessaint, J.P., Vermersch, P. et al. (2006). Diversified serum IgG response involving non-myelin CNS proteins during experimental autoimmune encephalomyelitis. *J Neuroimmunol* 179, 53–64. Epub 2006 Aug 8.

Zhang, H., Cheng, X., Richter, M. and Greene, M.I. (2006). A sensitive and high-throughput assay to detect low-abundance proteins in serum. *Nat Med* 12, 473–477.

CHAPTER

Genomic Assessment of Inflammatory Bowel Disease

Ad A. van Bodegraven and Cisca Wijmenga

INTRODUCTION

Inflammatory bowel diseases (IBD) are a group of chronic inflammatory disorders of the gastrointestinal tract of unknown origin; they comprise three main entities: ulcerative colitis (UC), Crohn's disease (CD), and indeterminate colitis. Indeterminate colitis represents a patient group that cannot be classified as either UC or CD, but which constitutes about 10% of IBD patients. Follow-up studies have shown that the majority of these patients eventually develop the features of UC. In addition, collagenous colitis and microscopic colitis have been described, although they will not be discussed in this chapter. Celiac disease is also a chronic inflammatory disorder of the gastrointestinal tract, but this disease is not usually included in IBD.

The incidence of CD has increased 8- to 10-fold over the past 50 years (Bach, 2002), whereas the incidence of UC has remained stable. CD is mainly seen in urbanized, developed countries and the change in incidence is associated with improved hygiene or social standards in Western societies. In contrast, UC is a more global disease (Karlinger et al., 2000). The median age of patients with IBD at diagnosis has increased over time, partly due to a higher proportion of elderly patients (>65 years of age), especially in UC. However, the incidence of CD is also increasing in patients younger than 17 years of age. The lifetime risk for Caucasians to develop UC is twice that of developing CD (0.15%) (Schreiber et al., 2005). Hence, the incidence of IBD varies with geographic location, industrialization and the availability of diagnostic means. It ranges from 0.5 to 24.5 per 100,000 people per year for UC, and from 0.1 to 10 per 100,000 per year for CD (Gallop et al., 1988; Moum et al., 1996; Russel et al., 1998). There is an evident north–south gradient, with the highest incidence of IBD occurring in northern countries, such as Scandinavia and Scotland, followed by Northwest Europe (the Netherlands, United Kingdom) and the United States, and with the lowest reported incidences in southern, Mediterranean Europe (Greece, Crete, Italy). Recent data from developing countries show similar trends (Jiang et al., 2006). Apart from the differences in incidence rates of IBD around the world, national, and regional differences have also been observed (Ekbom et al., 1991; Latour et al., 1998; Timmer and Goebell, 1999; Timmer et al., 1999). Reported prevalence rates of UC and CD also vary widely in different studies, ranging from 21 to 234 per 100,000 for UC and from 12 to 146 per 100,000 for CD. Colorectal CD seems to be increasing, and in UC, proctitis and left-sided UC seem to be becoming more prevalent.

Both diseases show variation in the individual clinical presentation and outcome that is most likely due to differences in genetic susceptibility, exposure to environmental factors, the commensal bacteria in the intestine, and the intestinal immune system. Even though CD and UC share some features, they also show major differences and it is still not clear whether both disorders are related at the molecular level, which would justify lumping them together as IBD. However, more and more evidence is emerging that a dysregulated mucosal immune response underlies the chronic intestinal inflammation, potentially initiated by a dysfunctional (immunological) intestinal barrier.

Copyright © 2010, Elsevier Inc.
All rights reserved.

PREDISPOSITION (GENETIC AND NON-GENETIC)

Current theories suggest that multiple factors, each of relatively weak effect, may act together to influence disease risk in IBD. Apart from genetic factors, there is evidence that supports an association between IBD and a large number of seemingly unrelated environmental factors, including smoking, oral contraception, diet, breastfeeding, drugs, geographical and social status, stress, microbial agents, intestinal permeability, and appendectomy (Corrao et al., 1998; Danese et al., 2004; Garcia Rodriguez et al., 2005; Loftus, 2004). However, only smoking has been confirmed as a clear environmental risk factor for IBD, protecting against UC but increasing the risk for CD (Cosnes, 2004).

In addition, IBD has a strong genetic component, although this seems to be more pronounced in CD than in UC. The derived heritability in CD is higher than for many other common complex diseases. Evidence for a strong inherited predisposition to IBD susceptibility comes from twin studies and studies of familial aggregation (Peeters et al., 1996; Satsangi et al., 1994; Tysk et al., 1988). There is, however, no discernable Mendelian inheritance pattern; thus we must assume a complex pattern of inheritance in which both susceptibility genes and environmental factors may contribute to IBD pathogenesis.

Twin studies provide a powerful means of assessing the contribution of both genetic and environmental factors to disease susceptibility (Martin et al., 1997). There are different measures of concordance used in such studies, but those based on National Twin Registries allow the individual twins to be selected independently of disease, so that they can then be assessed separately. In a recent study based on 38,507 identified twins born in Denmark from 1953 to 1982, the proband-wise concordance rates for UC and CD among monozygotic and dizygotic twins were estimated and found to be 58.3% for CD and 18.2% for UC among the monozygotic pairs, and 0% and 4.5%, respectively among the dizygotic pairs (Halfvarson et al., 2003; Orholm et al., 2000).

The sibling relative risk provides a further measure of the heritability of IBD and is defined as the risk to an affected patient's sibling divided by the population risk (prevalence). Estimates from German (Kuster et al., 1989) and northern France/Belgium (Laharie et al., 2001) populations suggest sibling relative risks for CD of 15–35. For a sibling of an UC patient, this risk is substantially lower: a Danish (Orholm et al., 1991) and Italian (Meucci et al., 1992) study showed relative sibling risks of 6–9 for UC. Another argument for a genetic contribution to IBD pathogenesis stems from differences in prevalence among different ethnic groups, with the highest rates found amongst Caucasians and people with an Ashkenazi Jewish background (Roth et al., 1989; Yang et al., 1993).

Over the past 10 years substantial progress has been made in identifying susceptibility genes for IBD, using both genome-wide linkage scans in affected sibling pairs and genetic association studies. To date, 14 genome-wide linkage scans have been conducted that have resulted in the identification of at least 9 loci (*IBD1-IBD9*). A recent meta-analysis on data from 10 genome-wide scans also revealed the *IBD1*, *IBD3*, *IBD5*, and *IBD6* loci, as well as many novel loci (Van Heel et al., 2004). Unfortunately not all IBD susceptibility loci have been replicated consistently. Silverberg and colleagues showed that IBD is frequently misdiagnosed, thereby reducing the ability to detect linkage (Silverberg et al., 2001). A recent study using a phenotyped cohort of 904 affected relative pairs uncovered novel loci and also demonstrated that the *IBD2* locus is an extensive UC locus (Achkar et al., 2006). The next step, to go from a linkage region to a disease susceptibility gene, has also proved to be difficult. The usual course is to switch from a family-based linkage design to a population-based genetic association design (Wild and Rioux, 2004). The latter requires large numbers of case/control pairs and polymorphic markers able to capture all the genetic variation. The identification of the vast majority of common (>1%) single nucleotide polymorphisms (SNP) and the correlation between them ("HapMap") (Altshuler et al., 2005) will revolutionize such studies in future. From the established IBD linkage regions, only the disease susceptibility genes from *IBD1* (*CARD15/NOD2*) and *IBD5* (*SLC22A4/SLC22A5*) have been uncovered (Table 45.1). Both these genes confer risk of CD. The current status of genetic research in IBD has been discussed in excellent reviews by Bonen and Cho (2003), Wild and Rioux (2004), Newman and Siminovitch (2005), and Vermeire and Rutgeerts (2005).

The *CARD15/NOD2* gene was identified through both a classical positional cloning strategy (Hugot et al., 2001) and a positional and functional candidate gene approach (Hampe et al., 2001; Ogura et al., 2001). Mutations in *CARD15/NOD2* are found in approximately 30% of CD patients. Although multiple variants in the gene can increase susceptibility only to CD, the most commonly identified mutations are the R702W, G908R, and L1007fsinsC. The increased risk conferred by mutations in the *CARD15/NOD2* gene depends on the number of mutant alleles: heterozygosity increases the risk 2- to 3-fold, whereas homozygosity is associated with a 20- to 40-fold higher risk of developing CD. However, the frequencies of mutant *CARD15/NOD2* alleles vary considerably across European populations (Arnott et al., 2004) and show low frequencies in northern European countries (Arnott et al., 2004; Helio et al., 2003; Medici et al., 2006; Thjodleifsson et al., 2003). *CARD15/NOD2* mutations are associated with a more severe form of the disease, an early age at onset, and ileal lesions (Abreu et al., 2002; Hampe et al., 2002; Lesage et al., 2002). *CARD15*, the caspase recruitment domain (CARD) family member 15 protein, is a member of the Nod1/Apaf-1 family and encodes a protein with two CARDs and six leucine-rich repeats (LRRs). The protein is primarily expressed in the peripheral blood leukocytes, but also in the intestinal epithelial layer, in particular in Paneth cells and associated with mucosal alpha-defensin expression (Ogura et al., 2003; Wehkamp et al., 2004). It plays also a role in the immune response to intracellular bacterial lipopolysaccharides (LPS) by recognizing the muramyl dipeptide (MDP) derived from them and activating the NF-kB protein. Although its role in CD is

Table 45.1 Genes known or suspected of being involved in IBD pathogenesis

Replicated IBD genes/loci	Chromosomal location	Disease	Relative risk heterozygote/ homozygote	Alleged mechanism of action	Genotype/ phenotype correlation
CARD15/NOD2 (IBD1)	16q13	CD	2–4 versus 20–40	Innate immunity, balance of immunological response	Ileal involvement and earlier clinical presentation
DLG5	10q23	CD	~1.5	Maintenance of epithelial barrier integrity	Ileal (combined with peri-anal)
SLC22A4/OCTN1, SLC22A5/OCTN2 (IBD5)	5q31–q33	CD	2.5 versus 4	Epithelial barrier	Multi-site, more severe phenotype
MYO9B	19p13	UC	1.2	Epithelial barrier	
IL23R	1p31	IBD	0.26	Innate immunity	Both small intestinal (ileal CD) and large intestinal (UC) inflammation
ATG16L1	2q37.1	CD	1.45	Mediating innate immune response to bacteria, interaction with NOD2	
Suggested IBD genes/loci					
IBD2	12q14	UC		?	Extensive UC
IBD3	6p21			MHC associated immunological responses	Both small intestine and colonic disease locations
IBD4	14q11–12	CD			Interaction with smoking
MDR1/ABCB1	7q22	IBD?		Membrane transporter (xenobiotic handling)	Refractory (and extensive) disease
PXR		IBD		Response to xenobiotics	
NOD1/CARD4	7p14	IBD		Innate immunity	Early age at onset (<25 y)
TLR-9	3p21			Mediating innate response, interaction with NOD2	
TLR-4		IBD		Mediating innate immune response to gram-negative bacteria. Interaction with NOD2	
TLR-5				Mediating innate immune response to flagellin	
NF-kB1	4q	IBD		Pivotal mediator of generation pro-inflammatory cytokines	
ICAM-1 (IBD6?)	19p13	IBD		Function of pivotal adhesion molecule	In CD: fistulizing disease. In UC: limited disease extent
PPARg (IBD9?)		IBD		Innate response to intestinal bacteria	
COX2		IBD		Inflammatory response	
TNFSF15				Pivotal pro-inflammatory cytokine	

not entirely elucidated, many leads point at involvement with the innate immune response.

Fine-mapping of the *IBD5* locus initially led to the identification of a unique haplotype of 250 kb (Rioux et al., 2001). Since the SNPs within the haplotype were in strong linkage disequilibrium, they all conferred risk to CD. The association to the 250 kb haplotype has been replicated extensively, thereby confirming the importance of the *IBD5* locus. Subsequently, extensive sequence analysis revealed two variants within the organic cation transport genes *SLC22A4/OCTN1* and *SLC22A5/OCTN2,* which have been proposed to alter the gene function and expression and to form a haplotype associated with CD risk (Peltekova et al., 2004). Because of the strong linkage disequilibrium across the entire 250 kb haplotype, it remains to be proven if the two identified variants act independently of the surrounding SNPs (reviewed in Reinhard and Rioux, 2006). *SLC22A4/OCTN1* and *SLC22A5/OCTN2* are involved in the active cellular uptake of carnitine across the intestinal epithelial layer. The role of the *SLC22A4/OCTN1* and *SLC22A5/OCTN2* genes in CD pathology is not clear, but they are thought to act through an inappropriate inflammatory host response to commensal flora.

The fine-mapping of an IBD locus on 10q23 (Hampe et al., 1999) resulted in the identification of the *DLG5* gene, a member of the family of discs large (DLG) homologs (Stoll et al., 2004) and the membrane-associated guanylate kinase (MAGUK) superfamily. DLG5 localizes to the plasma membrane and cytoplasm and interacts with components of adherens junctions and the cytoskeleton. A function in the transmission of extracellular signals to the cytoskeleton and in maintaining epithelial cell structure has been proposed. The association with *DLG5* varies significantly across populations (Buning et al., 2006; Daly et al., 2005; Ferraris et al., 2006; Gazouli et al., 2005; Medici et al., 2006; Newman et al., 2006; Noble et al., 2000a; Tenesa et al., 2006; Torok et al., 2005; Tremelling et al., 2006; Vermeire et al., 2005; Yamazaki et al., 2004).

More recently, the celiac disease-associated gene *MYO9B* (Monsuur et al., 2005), which localizes to the *IBD6* locus, was shown to be associated with IBD in four different populations (Van Bodegraven et al., 2006). Table 45.1 gives an overview of the different genes known or suspected of being involved in IBD pathogenesis. The picture that is emerging suggests that genes involved in IBD are either important in the innate immune response to commensal bacteria or involved in the maintenance of the epithelial barrier.

More recently it has become possible to perform genome-wide genetic association studies. Using 300,000 SNPs a highly significant association was found between CD and SNPs in the gene encoding a subunit of the proinflammatory cytokine IL23 (*IL-23R)* (Duerr et al., 2006). An uncommon coding variant within the *IL-23R* gene strongly protects against CD. The proinflammatory cytokine IL23 is considered to be a driver of innate immune pathology in the intestine. A second genome-wide study using nearly 20,000 nonsynonymous SNPs, found a strong association between CD and the autophagy-related 16-like 1 gene (*ATG16L1*) encoding a protein in the autophagosome pathway that processes intracellular bacteria (Hampe et al., 2007). The association of both IL23R and *ATG16L1* to CD fits in the emerging picture that genes involved in IBD are important in the innate immune response to commensal bacteria.

SCREENING

Screening for IBD can either be performed in high-risk families, or in the population at large. Apart from the general remarks on screening in a specific cohort, in particular for a disease with no known cure, risk factors have to be available. In affected families we can use the genetic risk factors to identify people at risk of developing IBD. A number of potential genetic risk factors have already been described that predispose to IBD, with *CARD15/NOD2* being the most commonly-associated gene. Unfortunately, the three common *CARD15/NOD2* risk variants have low sensitivity (38.5%) and rather low specificity (88.6%) for preventive screening of CD (Chamaillard et al., 2006). As there are no strategies for influencing the risk profile to attract disease, screening for this or any other risk gene does not yet contribute to diagnostic or treatment strategies for a (potential) patient.

In addition, several biomarkers have been described that are present in both IBD patients and in unaffected family members; these include increased intestinal permeability, anti-*Saccharomyces cerevisiae* antibodies (ASCA), and perinuclear antineutrophil cytoplasmatic antibodies (pANCA). In addition to these three markers, antibodies against microbial substances such as anaerobic coccoid rods, *Pseudomonas fluorescens* (I2), flagellins, and the outer membrane porin C of *Escherichia coli* (OmpC) have also been described in affected IBD patients (Bouma et al., 1999; Oudkerk Pool et al., 1993; Shanahan et al., 1992; Vermeire et al., 2001). It is interesting that these markers all point to the underlying etiopathogenetic hypothesis of IBD, namely a dysfunctional intestinal barrier function, leading to a dysfunctional handling of macromolecules and (intestinal) bacterial antigens, finally leading to an exaggerated immunological intestinal mucosal response, whether at the level of antigen-presenting cells or of cells involved in acquired immunological reactions (Fasano et al., 2005; Schreiber et al., 2005). Whereas serum pANCAs are mainly associated with UC (Oudkerk Pool et al., 1993), ASCA have particularly been associated with CD. *S. cerevisiae*, present in baker's yeast, is found in a wide variety of foods, and thus are an antigenic challenge to most people. Although it has been hypothesized that individuals characterized by an increased leakage of *S. cerevisiae* antigens generate ASCA, this hypothesis does not hold true for other commonly present, similar antigens, such as *Candida albicans*, which are not regularly found in CD patients.

Although these markers can be used in combination for some degree of differentiation between UC and CD (see also the Diagnosis section below), the presence of these types of markers has not been associated with a particular course of the disease, nor do they have enough sensitivity and specificity to help delineate non-affected people at risk of developing IBD (Yang et al., 1995).

DIAGNOSIS

UC and CD are clinically distinct disorders that are also differentially diagnosed. Recent research is aimed at genotype–phenotype associations in order to improve clinical assessment and treatment of patients. To interpret these genotype–phenotype data, phenotypical assessment has to be of outstanding and reproducible quality. Efforts to clearly define different subtypes of IBD have been undertaken by a group of IBD-experts at the World Congress of Gastroenterology 2005, Montreal, leading to the Montreal classification (Silverberg et al., 2005).

Ulcerative Colitis

UC is clinically characterized by urgency, bloody stools, a high-defecation frequency and abdominal discomfort, and it is commonly associated with fever and weight loss. In general, the severity of symptoms parallels the intensity and extent of inflammation. The inflammatory reaction primarily involves the mucosa of the large bowel, characteristically appearing as a continuous inflammation from the anal verge to the more proximal colon with a sharp demarcation to apparently normal mucosa. Microscopic examination reveals erosions, ulcerations, cryptitis, crypt abscesses, and infiltration of polymorphonuclear granulocytes, plasma cells, lymphocytes, and eosinophils in the lamina propria, accompanied by mucin depletion.

UC can be diagnosed by a combination of characteristic clinical, endoscopic, and pathohistological findings. Corroborating findings may be observed in blood chemistry (inflammatory response, iron-deficiency or chronic-disease associated anemia), in feces analysis (markers of inflammation, such as calprotectin), in the absence of microbial pathogens, and in serological findings, in particular the presence of pANCA. Endoscopic findings are invaluable for the subclassification of the disease, because this depends on disease extent. The role of radiological examination is limited in UC patients (Hanauer, 2004).

Classification of UC depends on disease extent, with subclasses of distal disease confined to the rectum (proctitis) or rectum and sigmoid colon (proctosigmoiditis), left-sided disease (from anal verge to splenic flexure), and universal or pancolitis, referring to disease proximal to the splenic flexure and usually involving the complete colon (pancolitis), in the new Montreal classification referred as E1-3. Severity of disease is classically assessed by the Truelove and Witts score, but several other combined scores are available.

Crohn's Disease

CD is an inflammatory disease of the whole intestinal tract, thus leading to a broad spectrum of clinical presentations. Its most classical course is a localized inflammation of the terminal ileum, causing periods of relapsing abdominal pain with gradual weight loss, followed after months to years by postprandial pain with colitis due to intestinal obstruction. Involvement is limited to the colon, which was only acknowledged in the middle of the 20th century, although it presents with the same symptoms as UC, except that rectal bleeding is less pronounced. Increasingly, colorectal disease has been reported over time in several epidemiological studies. Furthermore, oral (granulomatous) ulceration or cheilitis, esophageal inflammation with difficulty of food passage and retrosternal pain, epigastrial pain and nausea due to gastritis or duodenal inflammation may all occur. Inflammation due to CD may extend over the whole bowel wall, but is frequently discontinuous, with alternating normal and affected areas. The transmural inflammation may lead to fistula formation between bowel segments, or between bowel and skin, bladder, or vagina. Peri-anal fistulas are relatively common. The macroscopic appearance encompasses areas of inflammation with erythema, exudate and aphtoid ulceration that may confluence and form large, deep tramline ulcers, or – when ulcers fuse transversely – cobblestones. Subsequent healing may lead to strictures. Characteristically, these areas of inflammation are patchy, whereas continuous inflammation, especially of the colon, may occur. The hallmark early microscopic lesion is an aphtoid ulcer; histopathologic ulcer features in a surrounding with unaffected, adjacent crypts. Focal mucosal disease is a pivotal histopathologic finding: focal cryptitis, focal chronic inflammation, isolated (acute) terminal ileitis all fit in with CD. Granulomas, localized well-formed aggregates of epitheloid histiocytes preferably in the presence of giant cells or a surrounding cuff of lymphocytes, are regularly found.

CD can be diagnosed by the combination of characteristic clinical, endoscopic, and pathohistological findings. Again, additional findings may be found in blood chemistry (inflammatory response, iron deficiency or chronic disease-associated anemia), in feces (calprotectin), in the absence of microbial pathogens, and in serological findings, in particular the presence of ASCA. The role of radiological examination is complimentary to endoscopic findings in CD patients. Small bowel diagnostics still rely on classical radiological methods such as small bowel follow-through and enteroclysis. Recently, newer techniques have begun replacing the old ones. These new techniques include CT scans and MRI scans, preferentially enhanced by intra-intestinal and vascular contrast. In addition, a video capsule has been developed, introducing visual assessment of the small bowel and leading to a more accurate small bowel assessment, while double-balloon endoscopy has provided endoscopic access to the complete small bowel since 2001, allowing sampling for biopsy specimens as well as visual assessment.

CD is usually classified by a combination of age, disease localization, and disease behavior, detailed in the Vienna classification (Gasche et al., 2000). In 2005 the Vienna classification was updated to become the Montreal classification (Silverberg et al., 2005). Severity of CD is most commonly assessed by the CD activity index (CDAI), a score that calculates disease activity for eight components, some of which are monitored over 7 days: stool frequency, general well-being, abdominal discomfort, erythrocyte sedimentation rate (ESR), weight, extra intestinal symptoms, use of antidiarrheal drugs, and fever. The validity of such a subjective, 7 days' score has been discussed extensively. An instant bedside or bureau-side method is the Harvey-Bradshaw (HB) score that is highly related with CDAI, comprising frequency of liquid stools,

general well-being, abdominal pain, complications (extra-intestinal symptoms), and presence of an abdominal mass. This handy score is less widely used than CDAI in clinical (scientific) practice, partly due to EMEA and FDA regulations on drug registration. In order to avoid any subjective measure in a severity assessment score, a Disease activity score (DAI) of laboratory values and calculable physical signs was drawn up. The main problem with this score is its cumbersome calculation. CDAI and DAI are only related to a small extent. It is of interest that in recent clinical trials with biologicals patients were stratified for elevated CDAI score as well as an increased C-reactive protein (CRP) concentration as an indicator the acute phase response. Apparently, CDAI calculation underscores inflammatory variables. A solely subjective score reflecting the quality of life of IBD patients has been validated with the same purpose; assessment of disease activity and response to therapy. The most regularly used score is the Inflammatory Bowel Disease Questionnaire (IBDQ) a 32-item questionnaire that evaluates quality of life using four dimensions: bowel complaints, general malaise, social impact, and emotional burden.

Serological markers may be helpful in identifying patients with IBD. pANCA is expressed in the majority of UC patients, varying from 30–85%, whereas only 4–30% of CD patients show this profile, and then, particularly, in patients with CD confined to the colon, making differentiation between UC and CD more difficult. Patients suffering from concurrent primary sclerosing cholangitis (PSC) have a higher prevalence of pANCA. Positive reactions against ASCA, identifying antibodies against yeast cell wall phosphopeptidomains, are more commonly found in CD patients (50–70%) than in UC patients (6–14%), and healthy controls (0–5%). IgA type antibodies appear to have a higher specificity for CD, but ASCA is also present in celiac disease (up to 30%), ankylosing spondylitis, autoimmune hepatitis, and others.

The presence in various IBD populations – and their family members – of other detected microbial antibodies, such as against the outer membrane porin C of *E. coli* (OmpC), antibodies against *P. fluorescens* (I2) and several more has not been studied sufficiently to be of use in diagnosing disease. The origin and significance of these antibodies are yet unknown.

The genetic background of IBD patients varies greatly, making the presence of the susceptibility genes already described insufficient to diagnose IBD, even in cohorts of people at risk, such as family members. Two well-described examples of these risk genes that are (as yet) unsuitable for establishing a diagnosis are *HLA-DRB1*0103* for UC in Caucasians, *HLA-DRB1*1502* in Japanese and *CARD15/NOD2* in CD. Genetic markers are therefore not currently used in the diagnostic process of IBD.

Despite the differences in clinical characteristics between CD and UC, some 10% of patients are diagnosed as unclassified or indeterminate colitis (depending on whether surgery, i.e., proctocolectomy, has been performed) as they cannot be classified as either CD or UC. Up till now, no serological, genetic or other characteristics can distinguish with enough accuracy between the three IBD-diagnoses, UC, CD, and unclassified colitis. A recent study showed that gene expression on peripheral blood monocytes can provide biomarkers to distinguish CD and UC patients (Burczynski et al., 2006). Although these results need to be validated and replicated in independent and prospective cohorts, this type of diagnostic tools holds great promise for the future.

PROGNOSIS

The course of UC is usually relapsing, a disease pattern occurring in about 90% (Langholz et al., 1994). About 10% will have only one period of active disease in a 25-year time span and the probability of chronic active UC is also low (<20%) (Langholz et al., 1994; Witte et al., 2000). Risk factors precipitating relapses are use of aspirin or NSAID, enteric infection, and seasonal variation, possibly related to stimulation of the (eosinophils-mediated) immune system. Smoking and appendectomy appear to be protective (Andres and Friedman, 1999). About 25–40% of UC patients will require surgery in the short- or long-term (Turner et al., 2007). Chronic active, medical-therapy refractory UC is considered to be a proper indication for surgery, usually proctocolectomy in combination with an ileo-anal pouch reconstruction. Prior to surgery, high-dose intravenous corticosteroid therapy up to 60 mg prednisolone-equivalents daily is usual and well-documented to be effective in about two-third of patients (Turner et al., 2007). Predictors of surgical intervention are disease extent, clinical signs of severe systemic disease involvement (such as stool frequency, temperature, heart rate), and inflammatory variables like serum CRP and albumin concentration, whereas fecal calprotectin concentration holds promise for the future (Turner et al., 2007). In addition, fulminant active disease, concurrent development of colonic adenocarcinoma, and complications due to active disease such as intractable colonic bleeding or stenosis are other indications for surgery. Although observed in large cohorts, the increased risk of developing colonic adenocarcinoma is relatively limited. Risk factors for colonic adenocarcinoma comprise the presence of primary sclerosing cholangitis (PSC), and non-responsive disease of the universal colon. This increased risk comes on top of general markers of an increased risk for the development of colonic carcinoma, such as a positive family history. The debate of an increased risk of lymphoma in IBD patients is still ongoing, with confounding factors of chronic disease, use of immunosuppressive and other methodological difficulties.

Markers for the course of UC comprise both serological and genetic markers. Although the presence of pANCA has neither been associated with disease localization, disease severity, nor response to therapy, it is more commonly found in patients with concurrent PSC. The latter has been associated with development of colonic (and cholangio-) adenocarcinoma. Genetic markers of disease severity include *HLA DRB1*0103*, *IKBL+738*, and *hMLH1* 655A > G. The *HLA DRB1*0103*-allele has been associated with a severe course of UC, including non-response to corticosteroids and necessity of surgery (Satsangi et al., 1996; Roussomoustakaki et al., 1997), and similar findings were reported in a Spanish UC cohort for the presence of *IKBL+738* (de la Concha et al., 2000). The *hMLH1* gene

might be involved in genetic susceptibility to refractory UC as the GG genotype at position 655 of the *hMLH1* gene was almost 5 times more frequent in refractory UC patients compared with nonrefractory patients (Bagnoli et al., 2004). These and other combinations of genes have been associated with clinical course; however most studies could not be reproduced or concerned rather small patient cohorts, limiting the clinical significance in daily practice.

The course and prognosis of CD has been described extensively by Munkholm et al., based on large population-based cohorts from Denmark (Munkholm et al., 1994, 1997; Langholz et al., 1997). These population studies are excellent for extracting data on disease course, but refer mainly to this particular Nordic population so that extrapolation to other populations remains difficult. Nevertheless, these data reveal that the course of disease during the first year after diagnosis predicts the course in the following 5 years. Being quiescent in the first year gives rise to a quiescent disease course in subsequent years in 44% of CD patients; only 8% have an active course of disease. The remaining 48% suffers from a relapsing disease course. In patients with an active disease in the first year following diagnosis, these numbers are 5% and 45%, with 50% having a relapsing course. A second approach is to subclassify according to the proposals from a working group of IBD experts in Vienna, the so-called Vienna classification (Gasche et al., 2000), later updated in Montreal (Silverberg et al., 2005). In particular, patients with a diagnosis at young age and those suffering from penetrating disease have a worse disease outcome. Surgery, as a measure of failure of medical therapy, and thus of a disadvantageous course of CD, is still necessary in about 70% of patients. This high percentage has not yet dropped, notwithstanding new therapies, showing that truly disease-modifying treatment approaches still need to be developed.

Many efforts have been undertaken to identify predictors of disease course allowing for adapted therapeutic approaches. Most, however, have restricted clinical value. ASCA positivity is associated with small intestinal localization of disease (terminal ileum), and correlated with young age of diagnosis of CD. Presence of IgG and IgA antibodies is associated with a stenotizing or penetrating type of disease behavior and less often with a chronic inflammatory colonic type of CD (Abreu et al., 2002; Mow et al., 2004).

In phenotype-genotype studies, the course of CD is usually classified according to anatomic localization, disease behavior, and age of onset (according to the Vienna classification). There is no evidence, however, that this subclassification constitutes nosologic entities. In addition, the Vienna classification varies over time due to the developing course of disease (Louis et al., 2001). However, the validity and rigor of phenotyping are major factors in the adequate interpretation of correlations with an alleged genetic background. Thus interpretation of genotype studies to predict disease course is proving difficult, and fixed time spans have been suggested in the Montreal classification before a phenotype can be reliably assessed. Immediately following the description of *CARD15/NOD2*, several groups reported a phenotypic correlation of one of the major mutated *CARD15/NOD2* alleles with localization of CD in the ileum, a finding that has been corroborated repeatedly. Surgery and a stricturing phenotype of CD have also been reported in several series (Alvarez-Lobos et al., 2005; Russell et al., 2005), whereas, interestingly, the *CARD15/NOD2* genotype did not predict response to the potent anti-TNF-alpha drug infliximab. Other genes reported to be associated with CD, including the infrequent *HLA-DRB1*03*, *SLC22A4/OTCN1* and *SLC22A5/OCTN2* (Noble et al., 2005; Vermeire et al., 2005), and the *IBD 5* (Latiano et al., 2006) locus, have been related to a variety of phenotypes, including earlier age of disease onset, peri-anal (fistulizing) disease, ileal or colonic disease, or reduced need of surgery; there are no corroborating data from large cohorts.

Extra-intestinal manifestations have been reported since the very beginning that IBD has been recognized as a disease entity. Extra-intestinal complications comprise inflammatory changes of joints, eyes, cutis, and liver, but also thrombo-embolic manifestations and osteoporosis. Several studies have reported an overall incidence in 25–35% of IBD patients. Symptoms may be confined to any single organ system, although combinations of organ systems can also occur. A wide variety of rare manifestations have been reported. Most frequently the joints are affected with arthralgia and peripheral or axial arthritis. Other regular manifestations involve the skin, showing erythema nodosum or pyoderma gangrenosum, the eyes with inflammatory reactions such as (epi)scleritis and uveitis, the bone with osteopenia or even osteoporosis, the liver with PSC, leading to an increased risk of cholangiocarcinoma, the hematological tract with anemia and thrombocytopenia, and finally the vascular system with thrombo-embolic events. In addition, many autoimmune disorders may be associated with IBD. Some of these manifestations or concurrent diseases may reflect a common intestinal pathology, with a common pathway immunological response. Hence, IBD might be considered a systemic disorder disease, which may have implications for the proper recognition of patients at risk for IBD, especially when the extra-intestinal manifestations or concurrent diseases precede the development of IBD.

It is likely that genetic factors contribute to expression of this highly variable clinical picture, but few studies have been published. A correlation between the phenotype of IBD-associated arthropathy and HLA alleles has been reported, with *HLA-B27* and *HLA-B35* related to complaints of large joints and *HLA-B44* to small joints (Orchard et al., 2000). The same group of investigators ascribed a role for HLA alleles in the occurrence of ocular and cutaneous symptoms (Orchard et al., 2002). Other genetic factors have been associated with bone loss in IBD patients. Non-carriage of the 240-base pair allele of the *IL-1ra* gene and carriage of the 130-base pair allele of *IL-6* were independent of clinical severity of disease and application of corticosteroids associated with increased bone loss in a group of 83 IBD patients (Schulte et al., 2000).

It is likely that larger, well-phenotyped cohorts of IBD patients will allow for better study of genetic markers for extraintestinal manifestations, as may be deduced from efforts with specific IBD-associated containing chips, the so-called "IBDChip" (Sans et al., 2006). The IBDchip is the world's first diagnostic

DNA chip and is currently being validated in the European community. The chip contains 61 known polymorphic alleles, with alleged association with diagnosis and course of IBD and is aimed at a better prediction of prognosis and response to therapy of patients suffering from IBD.

PHARMACOGENOMICS

A broader understanding of the IBD disease pathology will deliver novel drug targets for disease intervention. It is anticipated that genomics and proteomics technologies will assist in compound identification in drug discovery. Infliximab, a chimeric monoclonal antibody against TNF-alpha, is such a target that neutralizes one of the critical inflammatory mediators in IBD. Nevertheless, the drug is only effective in 40–66% of patients after 2 years of treatment (Hanauer et al., 2002; Present et al., 1999; Sands et al., 2004). It is clear that drug response, in general, is partly genetically determined. Carriage of *CARD15/NOD2* seems to play no role in response to infliximab (Mascheretti et al., 2002; Vermeire et al., 2002), whereas another study on CD patients who had received infliximab showed that patients homozygous for the V allele of the *FcgammaRIIIa-158* polymorphism had a better biological and possibly better clinical response to infliximab (Louis et al., 2004). In addition, there has been a suggestion that the *IBD5* locus is involved in response to infliximab (Urcelay et al., 2005). Recently there has also been a report suggesting that the lack of response to infliximab can, in part, be attributed to certain genetic polymorphisms in two apoptosis genes, Fas ligand and caspase-9 (Hlavaty et al., 2005). Additional studies are needed to corroborate these findings.

IBD patients homozygous for the methylenetetrahydrofolate reductase (*MTHFR*) 1298C allele are more likely to experience side effects than patients homozygous for the wild-type A allele (21.0% versus 6.3%, $P < 0.05$) when treated with the immunosuppressor methotrexate (MTX) (Herrlinger et al., 2005). This is in contrast to what was found in patients with rheumatoid arthritis where the *MTHFR* 1298A allele was associated with MTX-related adverse events in both Caucasians and African-Americans (Odds Ratio [OR] 15.86, 95% CI 1.51-167.01; $p = 0.021$) (Hughes et al., 2006). Similar results were observed in a Japanese study on rheumatoid arthritis, which showed that subjects with the 1298A allele had a higher frequency of side effects from MTX ($P < 0.05$, RR = 1.42, 95% CI 1.11-1.82) (Urano et al., 2002). These conflicting results raise some concern and require additional validation studies as well as a better understanding of how MTX works.

A more well-established polymorphism influencing drug response is the thiopurine methyl transferase gene (*TPMT*) in relation to the outcome of therapy with thiopurines, of which azathioprine and 6-mercaptopurine are frequently used in IBD patients. The metabolism of azathioprine or 6-mercaptopurine is rather complicated, involving various enzymatic steps with interindividual differences in genotype and phenotype. Essentially, the therapeutic activity of thiopurines is related to the concentration of the metabolite 6-thioguanine nucleosides (6-TGN), probably mediated by the phosphorylated forms of this metabolite (Neurath et al., 2005; Poppe et al., 2006). The pivotal enzyme TPMT methylates 6-mercaptopurine and its metabolite 6-thioinosinemonophosphate (6-TIMP) into 6-methylmercaptopurine (6-MMP) and 6-methyl-thioinosinemonophosphate (6-MTIMP), respectively. The activity of TPMT is genetically regulated (Krynetski et al., 1996; Yates et al., 1997). Approximately 1 in 9 patients is heterozygous for the common polymorphisms, and 1 in 300 patients is homozygous. Patients with one mutant (dysfunctional) TPMT allele have a diminished TPMT activity and patients with two non-functional mutant alleles have no TPMT activity. Low TPMT activity will result in an increased amount of azathioprine or 6-mercaptopurine being metabolized by hypoxanthine phosphoribosyl transferase (HGPRT), resulting in high levels of 6TGN. Hence, TPMT deficiency leads to a potentially life-threatening myelo-suppression as 6-TGN accumulates. TPMT with a higher than average activity leads to generation of high levels of 6-MMP, which has been associated with hepatotoxicity in one study in a specific group of children suffering from CD (Dubinsky et al., 2000), and 6-MTIMP, and low levels of 6TGN. This is associated with therapeutic inefficacy. Taken together, determination of the TPMT alleles may be helpful in predicting response to thiopurine therapy and in assessment of risk of myelotoxicity or hepatotoxicity. Its determination prior to treatment with thiopurines has therefore been advocated. Nevertheless, other factors contribute to adverse events of thiopurines, as has been shown by retrospective analysis of more than 40 IBD patients in whom thiopurines had to be withdrawn due to myelotoxicity. Of these the majority had wild-type TPMT alleles, although most patients with myelodepression in the first weeks of treatment were homozygous for mutant alleles (Colombel et al., 2000).

The response to glucocorticosteroids also seems to be partly genetically determined. Polymorphisms in the multidrug-resistance gene 1 (*MDR-1*) are excellent candidates to affect the absorption and concentrations of MDR-1 substrates (Farrell et al., 2000; Hoffmeyer et al., 2000). In a study in Hungarian patients it has been suggested that the *DLG5* 113A allele – which was shown not to be associated with disease susceptibility – may confer resistance to steroids (Lakatos et al., 2006).

With the advanced understanding of human genetic variation, it is now possible to perform genome-wide association studies (GWAS) in large cohorts of responders and non-responders to identify genetic risk and efficacy factors that are related to the effects of such drugs; this is similar to what is being done to identify disease susceptibility genes. Such research may have broad implications since it is likely that the response to a particular drug will not be restricted to the IBD phenotype. Hence, in future, these genetic profiles may help tailor disease treatment and patient care by increasing efficacy and preventing adverse drug reactions (Egan et al., 2006).

2009 UPDATE

With improved genotyping technologies and the completion of the human genome sequence, GWAS have recently become possible and were declared the breakthrough of the year in 2007 (Pennisi, 2007). Not only has the first successful GWAS been performed in CD (Duerr et al., 2006), but inflammatory bowel disease (IBD) has been studied very extensively and also very successfully with nine GWAS for CD (Duerr et al., 2006; Franke et al., 2007; Hampe et al., 2007; Kugathasan et al., 2008; Libioulle et al., 2007; Raelson et al., 2007; Rioux et al., 2007; Wellcome Trust Case Control Consortium, 2007; Yamazaki et al., 2005) and a subsequent meta-analysis of three of these studies (Barrett et al., 2008). There are currently more than 30 CD-associated loci. To date three GWAS for UC have been performed in addition, revealing more than 10 UC-associated loci (Fisher et al., 2008; Franke et al., 2008; Silverberg et al., 2009). This has put IBD at the forefront of complex disease genetic research.

The first GWAS (Duerr et al., 2006) identified the gene encoding a subunit of the interleukin 23 receptor (*IL23R*) on chromosome 1p31 as a susceptibility gene for CD in a population of non-Jewish patients of European ancestry with ileal CD. An uncommon variant, Arg381Gln, was shown to have a strong protective effect for CD, and other noncoding *IL23R* variants were independently associated. This association was confirmed in independent cohorts and, although less strongly, is also associated with UC susceptibility (Weersma et al., 2008c). Since the proinflammatory cytokine IL23 is increasingly being recognized as involved in gut inflammation, this finding is of great importance in elucidating the pathogenetic background of IBD.

The second study (Hampe et al, 2007) found a strong association between CD and the autophagy-related 16-like 1 gene (*ATG16L1*) encoding a protein in the autophagosome pathway, which processes intracellular bacteria. A coding SNP (Thr300Ala) was strongly associated with susceptibility for CD. A third GWAS by the US NIDDK consortium in CD identified the same SNP in *ATG16L1* to be associated with CD and demonstrated that functional knockdown of the gene abrogates autophagy of *Salmonella typhimurium,* suggesting that autophagy and host cell responses to intracellular microbes are involved in the pathogenesis of CD (Rioux et al., 2007). A large GWAS including 14,000 cases (including 2000 CD patients) and 3000 controls by the Wellcome Trust Case Control Consortium (WTCCC, 2007) and a follow-up study (Parkes et al., 2007) identified variants in *IRGM*, which, like *ATG16L1*, is also involved in autophagy, to be associated with CD susceptibility. The finding that autophagy is involved in CD pathogenesis is compatible with the reported therapeutically beneficial administration of the autophagy-inducing mTOR inhibitors sirolimus and everolimus to refractory IBD-patients (Dumortier et al., 2008; Massey et al., 2008).

The WTCCC study and the follow-up study identified 12 loci, including *IRGM* that met the criteria for genome-wide significance. These included SNPs at 1q24, 1q31, 3p21 (harboring the macrophage stimulating-1 *MST-1* gene), *IL12B*, *NKX2-3*, and 21q22. Furthermore, the previously known *IBD5* locus at 5q31 and the gene desert on 5p13, identified by a French–Belgian GWAS, were confirmed (Libioulle, et al., 2007; Parkes et al., 2007; Rioux, 2001; WTCCC, 2007). Of interest is that the Protein Tyrosine Phosphatase 2 (*PTPN2*) gene is not only associated with CD, but also with type 1 diabetes, which reflects the partly shared pathogenesis of these distinct immune-related diseases (Zhernakova et al., 2009). Furthermore it is interesting that in the WTCCC study, CD was by far the most informative disease with regard to associated genes when compared to the other diseases that were studied, such as hypertension, bipolar disorder, and coronary heart disease.

Additional loci that were identified by GWAS were the nel-like precursor encoding gene (*NELL1*), paired-like homeobox 2B (*PHOX2B*), 10q21, and 16q24 (Franke et al., 2007; Rioux et al., 2007). Furthermore, an extensive three-stage analysis of genes involved in innate immunity identified two loci harboring *CARD9* and *IL18RAP* to be associated with both CD and UC (Zhernakova et al., 2008). Two loci on 20q13 and 21q22 were found to be associated with pediatric onset IBD in yet another GWAS (Kugathasan et al., 2008). A candidate gene study identified SNPs located in a predicted regulatory region on chromosome 1q44 downstream of *NLRP3* (Villani et al., 2008).

In 2008 a meta-analysis was performed, combining data from the UK–WTCCC, French–Belgian, and US–NIDDK studies. This has been a tremendously successful effort with 19 newly identified loci, which were replicated in an independent cohort, and confirmation of 11 previously published loci. These new loci included, amongst others, *PTPN22, JAK2, STAT3*, and *CDKAL1* (Barrett et al, 2008).

Despite the fact that many novel genes were identified to be involved in CD pathogenesis, until recently relatively little was known about UC. Most studies so far in UC focused on known CD risk variants. These studies showed that several risk loci are common to both CD and UC, while others – such as *ATG16L1*, *IRGM*, and *CARD15* – were specific for CD. Confirmed common CD–UC risk factors include variants in *MST1*, *NKX2-3*, *IL23R*, *IL12B*, and *CCNY* (Fisher et al., 2008; Franke et al., 2008). A genome-wide candidate gene experiment investigating 10,886 nonsynonymous SNPs in 905 British UC patients and 1465 controls identified the Extracellular Matrix Protein 1 (*ECM1*) on 1q21.2 as a UC-specific susceptibility gene (Fisher et al., 2008). ECM1 is known to strongly activate NF-κ B signaling which is a recognized key mediator in IBD. Another comprehensive genetic association analysis of 41

genes from the tight junction pathway, represented by 197 tag SNPs, showed association with *PARD3* (involved in cell polarity) and *MAGI2* (a scaffolding protein) for both UC and celiac disease, delineating the apparently shared genetic background of these diseases (Festen et al., 2008; Van Bodegraven et al., 2006; Wapenaar et al., 2008; Zhernakova et al., 2009).

In 2008, a European GWAS in UC was performed that identified several novel genetic variants that contribute to UC susceptibility. The most strongly replicated signals showed association to *IL10, ARPC2*, and the *HLA-BTNL2* region. IL10 is an immune-response inhibiting cytokine which has been investigated for many years in IBD. Notably, *IL10-/-* mice are known to develop a spontaneous colitis, indicating the probable important role of IL10 in UC (Franke et al., 2008). Another GWAS in UC from the United States confirmed *IL23R* and the *HLA-BTNL2* region and identified two additional loci on chromosomes 1p36 and 12q15 (Silverberg et al., 2009).

Although the genetics of IBD is only scratching the surface and many more variants remain to be identified, the first studies have been initiated to look at predictive value of molecular markers for disease diagnosis and prognosis (Weersma et al., 2008a, b).

2009 UPDATE REFERENCES

Barrett, J.C., Hansoul, S., Nicolae, D.L., Cho, J.H., Duerr, R.H., Rioux, J.D., Brant, S.R., Silverberg, M.S., Taylor, K.D., Barmada, M.M. et al. (2008). Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* 40(8), 955–962.

Duerr, R.H., Taylor, K.D., Brant, S.R., Rioux, J.D., Silverberg, M.S., Daly, M.J., Steinhart, A.H., Abraham, C., Regueiro, M., Griffiths, A. et al. (2006). A genome-wide association study identifies IL23R as an inflammatory bowel disease gene *Science* 314(5804), 1461–1463.

Dumortier, J., Lapalus, M.G., Guillaud, O., Poncet, G., Gagnieu, M.C., Partensky, C. and Scoazec, J.Y. (2008). Everolimus for refractory Crohn's disease: A case report. *Inflamm Bowel Dis* 14(6), 874–877.

Festen, E.A., Zhernakova, A., Wijmenga, C. and Weersma, R.K. (2008). Association of DLG5 variants with gluten-sensitive enteropathy. *Gut* 57(7), 1027–1028.

Fisher, S.A., Tremelling, M., Anderson, C.A., Gwilliam, R., Bumpstead, S., Prescott, N.J., Nimmo, E.R., Massey, D., Berzuini, C., Johnson, C. et al. (2008). Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease. *Nat Genet* 40(6), 710–712.

Franke, A., Balschun, T., Karlsen, T.H., Sventoraityte, J., Nikolaus, S., Mayr, G., Domingues, F.S., Albrecht, M., Nothnagel, M., Ellinghaus, D. et al. (2008). Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility. *Nat Genet* 40(11), 1319–1323.

Franke, A., Hampe, J., Rosenstiel, P., Becker, C., Wagner, F., Häsler, R., Little, R.D., Huse, K., Ruether, A., Balschun, T. et al. (2007). Systematic association mapping identifies NELL1 as a novel IBD disease gene. *PLoS ONE* 2(1), e691.

Hampe, J., Franke, A., Rosenstiel, P., Till, A., Teuber, M., Huse, K., Albrecht, M., Mayr, G., De La Vega, F.M., Briggs, J. et al. (2007). A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat Genet* 39(2), 207–211.

Kugathasan, S., Baldassano, R.N., Bradfield, J.P., Sleiman, P.M., Imielinski, M., Guthery, S.L., Cucchiara, S., Kim, C.E., Frackelton, E.C., Annaiah, K. et al. (2008). Loci on 20q13 and 21q22 are associated with pediatric-onset inflammatory bowel disease. *Nat Genet* 40(10), 1211–1215.

Libioulle, C., Louis, E., Hansoul, S., Sandor, C., Farnir, F., Franchimont, D., Vermeire, S., Dewit, O., de Vos, M., Dixon, A. et al. (2007). Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. *PLoS Genet* 3(4), e58.

Massey, D.C., Bredin, F. and Parkes, M. (2008). Use of sirolimus (rapamycin) to treat refractory Crohn's disease. *Gut* 57(9), 1294–1296.

Parkes, M., Barrett, J.C., Prescott, N.J., Tremelling, M., Anderson, C.A., Fisher, S.A., Roberts, R.G., Nimmo, E.R., Cummings, F.R., Soars, D., et al. (2007). Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat Genet* 39(7), 830–832.

Pennisi, E. (2007). Breakthrough of the year. Human genetic variation. *Science* 318(5858), 1842–1843.

Raelson, J.V., Little, R.D., Ruether, A., Fournier, H., Paquin, B., Van Eerdewegh, P., Bradley, W.E., Croteau, P., Nguyen-Huu, Q., Segal, J. et al. (2007). Genome-wide association study for Crohn's disease in the Quebec Founder Population identifies multiple validated disease loci. *Proc Natl Acad Sci U S A* 104(37), 14747–14752.

Rioux, J.D., Daly, M.J., Silverberg, M.S., Lindblad, K., Steinhart, H., Cohen, Z., Delmonte, T., Kocher, K., Miller, K., Guschwan, S. et al. (2001). Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat Genet* 29(2), 223–228.

Rioux, J.D., Xavier, R.J., Taylor, K.D., Silverberg, M.S., Goyette, P., Huett, A., Green, T., Kuballa, P., Barmada, M.M., Datta, L.W. et al. (2007). Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat Genet* 39(5), 596–604.

Silverberg, M.S., Cho, J.H., Rioux, J.D., McGovern, D.P., Wu, J., Annese, V., Achkar, J.P., Goyette, P., Scott, R., Xu, W. et al. Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study. *Nat Genet* 2009 January 4; [Epub ahead of print].

van Bodegraven, A.A., Curley, C.R., Hunt, K.A., Monsuur, A.J., Linskens, R.K., Onnie, C.M., Crusius, J.B., Annese, V., Latiano, A., Silverberg, M.S. et al. (2006). Genetic variation in myosin IXB is associated with ulcerative colitis. *Gastroenterology* 131, 1768–1774.

Villani, A.C., Lemire, M., Fortin, G., Louis, E., Silverberg, M.S., Collette, C., Baba, N., Libioulle, C., Belaiche, J., Bitton, A. et al. (2009). Common variants in the NLRP3 region contribute to Crohn's disease susceptibility. *Nat Genet* 41(1), 71–76.

Wapenaar, M.C., Monsuur, A.J., van Bodegraven, A.A., Weersma, R.K., Bevova, M.R., Linskens, R.K., Howdle, P., Holmes, G., Mulder, C.J., Dijkstra, G. et al. (2008). Associations with tight junction genes PARD3 and MAGI2 in Dutch patients point to a common barrier defect for celiac disease and ulcerative colitis. *Gut* 57(4), 463–467.

Weersma, R.K., Stokkers, P.C., van Bodegraven, A.A., van Hogezand, R.A., Verspaget, H.W., de Jong, D.J., van der Woude, C.J., Oldenburg, B., Linskens, R.K., Festen, E.A.M. et al. (2008a). Molecular prediction of disease risk and severity in a large Dutch Crohn's disease cohort. *Gut* 58(3), 388–395.

Weersma, R.K. and Wijmenga, C. (2008b). Using genetic information in the identification, classification and treatment of Crohn's disease patients: Are we there yet? *Expert Rev Gastroenterol Hepatol* 2(6), 719–721.

Weersma, R.K., Zhernakova, A., Nolte, I.M., Lefebvre, C., Rioux, J.D., Mulder, F., van Dullemen, H.M., Kleibeuker, J.H., Wijmenga, C. and Dijkstra, G. (2008c). ATG16L1 and IL23R are associated with inflammatory bowel diseases but not with celiac disease in the Netherlands. *Am J Gastroenterol* 103(3), 621–627.

Wellcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661–678.

Yamazaki, K., McGovern, D., Ragoussis, J., Paolucci, M., Butler, H., Jewell, D., Cardon, L., Takazoe, M., Tanaka, T., Ichimori, T. et al. (2005). Single nucleotide polymorphisms in TNFSF15 confer susceptibility to Crohn's disease. *Hum Mol Genet* 14(22), 3499–3506.

Zhernakova, A., Festen, E.M., Franke, L., Trynka, G., van Diemen, C.C., Monsuur, A.J., Bevova, M., Nijmeijer, R.M., van't Slot, R., Heijmans, R. et al. (2008). Genetic analysis of innate immunity in Crohn's disease and ulcerative colitis identifies two susceptibility loci harboring CARD9 and IL18RAP. *Am J Hum Genet* 82(5), 1202–1210.

Zhernakova, A., van Diemen, C.C. and Wijmenga, C. (2009). Detecting shared pathogenesis from the shared genetics of immune-related diseases. *Nat Rev Genet* 10(1), 43–55.

REFERENCES

Abreu, M.T., Taylor, K.D., Lin, Y.C., Hang, T., Gaiennie, J., Landers, C.J., Vasiliauskas, E.A., Kam, L.Y., Rojany, M., Papadakis, K.A. et al. (2002). Mutations in NOD2 are associated with fibrostenosing disease in patients with Crohn's disease. *Gastroenterology* 123(3), 679–688.

Achkar, J.P., Dassopoulos, T., Silverberg, M.S., Tuvlin, J.A., Duerr, R.H., Brant, S.R., Siminovitch, K., Reddy, D., Datta, L.W., Bayless, T.M. et al. (2006). Phenotype-stratified genetic linkage study demonstrates that IBD2 is an extensive ulcerative colitis locus. *Am J Gastroenterol* 101(3), 572–580.

Altshuler, D., Brooks, L.D., Chakravarti, A., Collins, F.S., Daly, M.J. and Donnelly, P. (2005 Oct 27). International HapMap Consortium. A haplotype map of the human genome. *Nature* 437(7063), 1299–1320.

Alvarez-Lobos, M., Arostegui, J.I., Sans, M., Tassies, D., Plaza, S., Delgado, S., Lacy, A.M., Pique, J.M., Yague, J. and Panes, J. (2005). Crohn's disease patients carrying Nod2/CARD15 gene variants have an increased and early need for first surgery due to stricturing disease and higher rate of surgical recurrence. *Ann Surg* 242(5), 693–700.

Andres, P.G. and Friedman, L.S. (1999). Epidemiology and the natural course of inflammatory bowel disease. *Gastroenterol Clin North Am* 28(2), 255–281. vii

Arnott, I.D., Nimmo, E.R., Drummond, H.E., Fennell, J., Smith, B. R., MacKinlay, E., Morecroft, J., Anderson, N., Kelleher, D., O'Sullivan, M. et al. (2004). NOD2/CARD15, TLR4 and CD14 mutations in Scottish and Irish Crohn's disease patients: Evidence for genetic heterogeneity within Europe?. *Genes Immun* 5(5), 417–425.

Bach, J.F. (2002 Sep 19). The effect of infections on susceptibility to autoimmune and allergic diseases. *N Engl J Med* 347(12), 911–920.

Bagnoli, S., Putignano, A.L., Melean, G., Baglioni, S., Sestini, R., Milla, M., d'Albasio, G., Genuardi, M., Pacini, F., Trallori, G. et al. (2004). Susceptibility to refractory ulcerative colitis is associated with polymorphism in the hMLH1 mismatch repair gene. *Inflamm Bowel Dis* 10(6), 705–708.

Bonen, D.K. and Cho, J.H. (2003). The genetics of inflammatory bowel disease. *Gastroenterology* 124(2), 521–536.

Bouma, G., Crusius, J.B., Garcia-Gonzalez, M.A., Meijer, B.U., Hellemans, H.P., Hakvoort, R.J., Schreuder, G.M., Kostense, P.J., Meuwissen, S.G. and Pena, A.S. (1999). Genetic markers in clinically well defined patients with ulcerative colitis (UC). *Clin Exp Immunol* 115(2), 294–300.

Buning, C., Geerdts, L., Fiedler, T., Gentz, E., Pitre, G., Reuter, W., Luck, W., Buhner, S., Molnar, T., Nagy, F. et al. (2006). DLG5 variants in inflammatory bowel disease. *Am J Gastroenterol* 101(4), 786–792.

Burczynski, M.E., Peterson, R.L., Twine, N.C., Zuberek, K.A., Brodeur, B.J., Casciotti, L., Maganti, V., Reddy, P.S., Strahs, A., Immermann, F. et al. (2006). Molecular classification of Crohn's disease and ulcerative colitis patients using transcriptional profiles in peripheral blood mononuclear cells. *J Mol Diagn* 8(1), 51–61.

Chamaillard, M., Iacob, R., Desreumaux, P. and Colombel, J.F. (2006). Advances and perspectives in the genetics of inflammatory bowel diseases. *Clin Gastroenterol Hepatol* 4(2), 143–151.

Colombel, J.F., Ferrari, N., Debuysere, H., Marteau, P., Gendre, J.P., Bonaz, B., Soule, J.C., Modigliani, R., Touze, Y., Catala, P. et al. (2000). Genotypic analysis of thiopurine S-methyltransferase in patients with Crohn's disease and severe myelosuppression during azathioprine therapy. *Gastroenterology* 118(6), 1025–1030.

Corrao, G., Tragnone, A., Caprilli, R., Trallori, G., Papi, C., Andreoli, A., Di Paolo, M., Riegler, G., Rigo, G.P., Ferrau, O. et al. (1998). Risk of inflammatory bowel disease attributable to smoking, oral contraception and breastfeeding in Italy: A nationwide case-control study. Cooperative Investigators of the Italian Group for the Study of the Colon and the Rectum (GISC). *Int J Epidemiol* 27(3), 397–404.

Cosnes, J. (2004). Tobacco and IBD: Relevance in the understanding of disease mechanisms and clinical practice. *Best Pract Res Clin Gastroenterol* 18(3), 481–496.

Daly, M.J., Pearce, A.V., Farwell, L., Fisher, S.A., Latiano, A., Prescott, N.J., Forbes, A., Mansfield, J., Sanderson, J., Langelier, D.

et al. (2005). Association of DLG5 R30Q variant with inflammatory bowel disease. *Eur J Hum Genet* 13(7), 835–839.

Danese, S., Sans, M. and Fiocchi, C. (2004). Inflammatory bowel disease: The role of environmental factors. *Autoimmun Rev* 3(5), 394–400.

de la Concha, E.G., Fernandez-Arquero, M., Lopez-Nava, G., Martin, E., Allcock, R.J., Conejero, L., Paredes, J.G. and Diaz-Rubio, M. (2000). Susceptibility to severe ulcerative colitis is associated with polymorphism in the central MHC gene IKBL. *Gastroenterology* 119(6), 1491–1495.

Dubinsky, M.C., Lamothe, S., Yang, H.Y., Targan, S.R., Sinnett, D., Theoret, Y. and Seidman, E.G. (2000). Pharmacogenomics and metabolite measurement for 6-mercaptopurine therapy in inflammatory bowel disease. *Gastroenterology* 118(4), 705–713.

Duerr, R.H., Taylor, K.D., Brant, S.R., Rioux, J.D., Silverberg, M.S., Daly, M.J., Steinhart, A.H., Abraham, C., Regueiro, M., Griffiths, A. et al. (2006 Dec 1). genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* 314(5804), 1461–1463.

Egan, L.J., Derijks, L.J. and Hommes, D.W. (2006). Pharmacogenomics in inflammatory bowel disease. *Clin Gastroenterol Hepatol* 4(1), 21–28.

Ekbom, A., Helmick, C., Zack, M. and Adami, H.O. (1991). The epidemiology of inflammatory bowel disease: A large, population-based study in Sweden. *Gastroenterology* 100(2), 350–358.

Farrell, R.J., Murphy, A., Long, A., Donnelly, S., Cherikuri, A., O'Toole, D., Mahmud, N., Keeling, P.W., Weir, D.G. and Kelleher, D. (2000). High multidrug resistance (P-glycoprotein 170) expression in inflammatory bowel disease patients who fail medical therapy. *Gastroenterology* 118(2), 279–288.

Fasano, A. and Shea-Donohue, T. (2005). Mechanisms of disease: The role of intestinal barrier function in the pathogenesis of gastrointestinal autoimmune diseases. *Nat Clin Pract Gastroenterol Hepatol* 2(9), 416–422.

Ferraris, A., Torres, B., Knafelz, D., Barabino, A., Lionetti, P., de Angelis, G.L., Iacono, G., Papadatou, B., D'Amato, G., Di Ciommo, V. et al. (2006). Relationship between CARD15, SLC22A4/5, and DLG5 polymorphisms and early-onset inflammatory bowel diseases: An Italian Multicentric Study. *Inflamm Bowel Dis* 12(5), 355–361.

Gallop, J.H., Phillips, S.F., Melton, L.J. and Zinsmeister, A.R. (1998). Epidemiologic aspects of Crohn's disease: A population based study in Olmsted County, Minnesota. *Gut* 29, 49–56.

Garcia Rodriguez, L.A., Gonzalez-Perez, A., Johansson, S. and Wallander, M.A. (2005 Aug 15). Risk factors for inflammatory bowel disease in the general population. *Aliment Pharmacol Ther* 22(4), 309–315.

Gasche, C., Scholmerich, J., Brynskov, J., D'Haens, G., Hanauer, S.B., Irvine, E.J., Jewell, D.P., Rachmilewitz, D., Sachar, D.B., Sandborn, W.J. et al. (2000). A simple classification of Crohn's disease: Report of the Working Party for the World Congresses of Gastroenterology, Vienna 1998. *Inflamm Bowel Dis* 6(1), 8–15.

Gazouli, M., Mantzaris, G., Archimandritis, A.J., Nasioulas, G. and Anagnou, N.P. (2005 Dec 21). Single nucleotide polymorphisms of OCTN1, OCTN2, and DLG5 genes in Greek patients with Crohn's disease. *World J Gastroenterol* 11(47), 7525–7530.

Halfvarson, J., Bodin, L., Tysk, C., Lindberg, E. and Jarnerot, G. (2003). Inflammatory bowel disease in a Swedish twin cohort: A long-term follow-up of concordance and clinical characteristics. *Gastroenterology* 124(7), 1767–1773.

Hampe, J., Shaw, S.H., Saiz, R., Leysens, N., Lantermann, A., Mascheretti, S., Lynch, N.J., MacPherson, A.J., Bridger, S., van Deventer, S. et al. (1999). Linkage of inflammatory bowel disease to human chromosome 6p. *Am J Hum Genet* 65(6), 1647–1655.

Hampe, J., Cuthbert, A., Croucher, P.J., Mirza, M.M., Mascheretti, S., Fisher, S., Frenzel, H., King, K., Hasselmeyer, A., MacPherson, A.J. et al. (2001 Jun 16). Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. *Lancet* 357(9272), 1925–1928.

Hampe, J., Grebe, J., Nikolaus, S., Solberg, C., Croucher, P.J., Mascheretti, S., Jahnsen, J., Moum, B., Klump, B., Krawczak, M. et al. (2002 May 11). Association of NOD2 (CARD 15) genotype with clinical course of Crohn's disease: A cohort study. *Lancet* 359(9318), 1661–1665.

Hampe, J., Franke, A., Rosenstiel, P., Till, A., Teuber, M., Huse, K., Albrecht, M., Mayr, G., De La Vega, F.M., Briggs, J. et al. (2007). A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat Genet* 39(2), 207–211.

Hanauer, S.B., Feagan, B.G., Lichtenstein, G.R., Mayer, L.F., Schreiber, S., Colombel, J.F., Rachmilewitz, D., Wolf, D.C., Olson, A., Bao, W. et al. (2002 May 4). ACCENT I Study Group. Maintenance infliximab for Crohn's disease: The ACCENT I randomised trial. *Lancet* 359(9317), 1541–1549.

Hanauer, S.B. (2004). Update on the etiology, pathogenesis and diagnosis of ulcerative colitis. *Nat Clin Pract Gastroenterol Hepatol* 1(1), 26–31.

Helio, T., Halme, L., Lappalainen, M., Fodstad, H., Paavola-Sakki, P., Turunen, U., Farkkila, M., Krusius, T. and Kontula, K. (2003). CARD15/NOD2 gene variants are associated with familially occurring and complicated forms of Crohn's disease. *Gut* 52(4), 558–562.

Herrlinger, K.R., Cummings, J.R., Barnardo, M.C., Schwab, M., Ahmad, T. and Jewell, D.P. (2005). The pharmacogenetics of methotrexate in inflammatory bowel disease. *Pharmacogenet Genomics* 15(10), 705–711.

Hoffmeyer, S., Burk, O., von Richter, O., Arnold, H.P., Brockmoller, J., Johne, A., Cascorbi, I., Gerloff, T., Roots, I., Eichelbaum, M. et al. (2000 Mar 28). Functional polymorphisms of the human multidrug-resistance gene: Multiple sequence variations and correlation of one allele with P-glycoprotein expression and activity *in vivo*. *Proc Natl Acad Sci USA* 97(7), 3473–3478.

Hlavaty, T., Pierik, M., Henckaerts, L., Ferrante, M., Joossens, S., van Schuerbeek, N., Noman, M., Rutgeerts, P. and Vermeire, S. (2005 Oct 1). Polymorphisms in apoptosis genes predict response to infliximab therapy in luminal and fistulizing Crohn's disease. *Aliment Pharmacol Ther* 22(7), 613–626.

Hughes, L.B., Beasley, T.M., Patel, H., Tiwari, H.K., Morgan, S. L., Baggott, J.E., Saag, K.G., McNicholl, J., Moreland, L.W., Alarcon, G.S. et al. (2006). Racial/ethnic differences in allele frequencies of single nucleotide polymorphisms in the methylenetetrahydrofolate reductase gene and their influence on response to methotrexate in rheumatoid arthritis. *Ann Rheum Dis* 65, 1213–1218.

Hugot, J.P., Chamaillard, M., Zouali, H., Lesage, S., Cezard, J.P., Belaiche, J., Almer, S., Tysk, C., O'Morain, C.A., Gassull, M. et al. (2001 May 31). Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 411(6837), 599–603.

Jiang, L., Xia, B., Li, J., Ye, M., Yan, W., Deng, C., Ding, Y., Luo, H., Hou, W., Zhao, Q. et al. (2006). Retrospective survey of 452 patients with inflammatory bowel disease in Wuhan city, central China. *Inflamm Bowel Dis* 12(3), 212–217.

Karlinger, K., Gyorke, T., Mako, E., Mester, A. and Tarjan, Z. (2000). The epidemiology and the pathogenesis of inflammatory bowel disease. *Eur J Radiol* 35(3), 154–167.

Krynetski, E.Y., Tai, H.L., Yates, C.R., Fessing, M.Y., Loennechen, T., Schuetz, J.D., Relling, M.V. and Evans, W.E. (1996). Genetic polymorphism of thiopurine S-methyltransferase: Clinical importance and molecular mechanisms. *Pharmacogenetics* 6(4), 279–290.

Kuster, W., Pascoe, L., Purrmann, J., Funk, S. and Majewski, F. (1989). The genetics of Crohn disease: Complex segregation analysis of a family study with 265 patients with Crohn disease and 5,387 relatives. *Am J Med Genet* 32(1), 105–108.

Laharie, D., Debeugny, S., Peeters, M., Van Gossum, A., Gower-Rousseau, C., Belaiche, J., Fiasse, R., Dupas, J.L., Lerebours, E., Piotte, S. et al. (2001). Inflammatory bowel disease in spouses and their offspring. *Gastroenterology* 120(4), 816–819.

Lakatos, P.L., Fischer, S., Claes, K., Kovacs, A., Molnar, T., Altorjay, I., Demeter, P., Tulassay, Z., Palatka, K. and Papp, M. (2006). Hungarian IBD Study Group DLG5 R30Q is not associated with IBD in Hungarian IBD patients but predicts clinical response to steroids in Crohn's disease. *Inflamm Bowel Dis* 12(5), 362–368.

Langholz, E., Munkholm, P., Davidsen, M. and Binder, V. (1994). Course of ulcerative colitis: Analysis of changes in disease activity over years. *Gastroenterology* 107(1), 3–11.

Langholz, E., Munkholm, P., Krasilnikoff, P.A. and Binder, V. (1997). Inflammatory bowel diseases with onset in childhood. Clinical features, morbidity, and mortality in a regional cohort. *Scand J Gastroenterol* 32(2), 139–147.

Latiano, A., Palmieri, O., Valvano, R.M., D'Inca, R., Vecchi, M., Ferraris, A., Sturniolo, G.C., Spina, L., Lombardi, G., Dallapiccola, B. et al. (2006). Contribution of IBD5 locus to clinical features of IBD patients. *Am J Gastroenterol* 101(2), 318–325.

Latour, P., Louis, E. and Belaiche, J. (1998). Incidence of inflammatory bowel disease in the area of Liege: A 3 years prospective study (1993-1996). *Acta Gastroenterol Belg* 61(4), 410–413.

Lesage, S., Zouali, H., Cezard, J.P., Colombel, J.F., Belaiche, J., Almer, S., Tysk, C., O'Morain, C., Gassull, M., Binder, V. et al. (2002). EPWG-IBD Group; EPIMAD Group; GETAID Group CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. *Am J Hum Genet* 70(4), 845–857.

Loftus, E.V., Jr (2004). Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences. *Gastroenterology* 126(6), 1504–1517.

Louis, E., Collard, A., Oger, A.F., Degroote, E., Aboul Nasr, E.L., Yafi, F.A. and Belaiche, J. (2001). Behaviour of Crohn's disease according to the Vienna classification: Changing pattern over the course of the disease. *Gut* 49(6), 777–782.

Louis, E., El Ghoul, Z., Vermeire, S., Dall'Ozzo, S., Rutgeerts, P., Paintaud, G., Belaiche, J., De Vos, M., Van Gossum, A., Colombel, J.F. et al. (2004). Association between polymorphism in IgG Fc receptor IIIa coding gene and biological response to infliximab in Crohn's disease. *Aliment Pharmacol Ther* 19(5), 511–519.

Martin, N., Boomsma, D. and Machin, G. (1997). A twin-pronged attack on complex traits. *Nat Genet* 17(4), 387–392.

Mascheretti, S., Hampe, J., Croucher, P.J., Nikolaus, S., Andus, T., Schubert, S., Olson, A., Bao, W., Folsch, U.R. and Schreiber, S. (2002). Response to infliximab treatment in Crohn's disease is not associated with mutations in the CARD15 (NOD2) gene: An analysis in 534 patients from two multicenter, prospective GCP-level trials. *Pharmacogenetics* 12(7), 509–515.

Medici, V., Mascheretti, S., Croucher, P.J., Stoll, M., Hampe, J., Grebe, J., Sturniolo, G.C., Solberg, C., Jahnsen, J., Moum, B. et al. (2006). Extreme heterogeneity in CARD15 and DLG5 Crohn disease-associated polymorphisms between German and Norwegian populations. *Eur J Hum Genet* 14(4), 459–468.

Meucci, G., Vecchi, M., Torgano, G., Arrigoni, M., Prada, A., Rocca, F., Curzio, M., Pera, A. and de Franchis, R. (1992). Familial aggregation of inflammatory bowel disease in northern Italy: A multicenter study. The Gruppo di Studio per le Malattie Infiammatorie Intestinali (IBD Study Group). *Gastroenterology* 103(2), 514–519.

Monsuur, A.J., de Bakker, P.I., Alizadeh, B.Z., Zhernakova, A., Bevova, M.R., Strengman, E., Franke, L., van't Slot, R., van Belzen, M.J., Lavrijsen, I.C. et al. (2005). Myosin IXB variant increases the risk of celiac disease and points toward a primary intestinal barrier defect. *Nat Genet* 37(12), 1341–1344.

Moum, B., Vatn, M.H., Ekbom, A., Aadland, E., Fausa, O., Lygren, I., Sauar, J., Schulz, T. and Stray, N. (1996). Incidence of ulcerative colitis and indeterminate colitis in four counties of southeastern Norway, 1990-93. A prospective population-based study. The Inflammatory Bowel South-Eastern Norway (IBSEN) Study Group of Gastroenterologists. *Scand J Gastroenterol* 31(4), 362–366.

Mow, W.S., Vasiliauskas, E.A., Lin, Y.C., Fleshner, P.R., Papadakis, K.A., Taylor, K.D., Landers, C.J., Abreu-Martin, M.T., Rotter, J.I., Yang, H. et al. (2004). Association of antibody responses to microbial antigens and complications of small bowel Crohn's disease. *Gastroenterology* 126(2), 414–424.

Munkholm, P., Langholz, E., Davidsen, M. and Binder, V. (1994). Frequency of glucocorticoid resistance and dependency in Crohn's disease. *Gut* 35(3), 360–362.

Munkholm, P. (1997). Crohn's disease–occurrence, course and prognosis. An epidemiologic cohort-study. *Dan Med Bull* 44(3), 287–302.

Neurath, M.F., Kiesslich, R., Teichgraber, U., Fischer, C., Hofmann, U., Eichelbaum, M., Galle, P.R. and Schwab, M. (2005). 6-thioguanosine diphosphate and triphosphate levels in red blood cells and response to azathioprine therapy in Crohn's disease. *Clin Gastroenterol Hepatol* 3(10), 1007–1014.

Newman, B. and Siminovitch, K.A. (2005). Recent advances in the genetics of inflammatory bowel disease. *Curr Opin Gastroenterol* 21(4), 401–407.

Newman, W.G., Gu, X., Wintle, R.F., Liu, X., van Oene, M., Amos, C.I. and Siminovitch, K.A. (2006). DLG5 variants contribute to Crohn disease risk in a Canadian population. *Hum Mutat* 27(4), 353–358.

Noble, C.L., Nimmo, E.R., Drummond, H., Smith, L., Arnott, I.D. and Satsangi, J. (2000a). DLG5 variants do not influence susceptibility to inflammatory bowel disease in the Scottish population. *Gut* 54(10), 1416–1420.

Noble, C.L., Nimmo, E.R., Drummond, H., Ho, G.T., Tenesa, A., Smith, L., Anderson, N., Arnott, I.D. and Satsangi, J. (2005b). The contribution of OCTN1/2 variants within the IBD5 locus to disease susceptibility and severity in Crohn's disease. *Gastroenterology* 129(6), 1854–1864.

Orchard, T.R., Thiyagaraja, S., Welsh, K.I., Wordsworth, B.P., Hill Gaston, J.S. and Jewell, D.P. (2000). Clinical phenotype is related to HLA genotype in the peripheral arthropathies of inflammatory bowel disease. *Gastroenterology* 118(2), 274–278.

Orchard, T.R., Chua, C.N., Ahmad, T., Cheng, H., Welsh, K.I. and Jewell, D.P. (2002). Uveitis and erythema nodosum in inflammatory bowel disease: Clinical features and the role of HLA genes. *Gastroenterology* 123(3), 714–718.

Orholm, M., Munkholm, P., Langholz, E., Nielsen, O.H., Sorensen, I.A. and Binder, V. (1991). Familial occurrence of inflammatory bowel disease. *N Engl J Med* 324(2), 84–88.

Orholm, M., Binder, V., Sorensen, T.I., Rasmussen, L.P. and Kyvik, K.O. (2000). Concordance of inflammatory bowel disease among Danish twins Results of a nationwide study. *Scand J Gastroenterol* 35(10), 1075–1081.

Ogura, Y., Bonen, D.K., Inohara, N., Nicolae, D.L., Chen, F.F., Ramos, R., Britton, H., Moran, T., Karaliuskas, R., Duerr, R.H. et al. (2001). A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 411(6837), 603–606.

Ogura, Y., Lala, S., Xin, W., Smith, E., Dowds, T.A., Chen, F.F., Zimmermann, E., Tretiakova, M., Cho, J.H., Hart, J. et al. (2003). Expression of NOD2 in Paneth cells: A possible link to Crohn's ileitis. *Gut* 52(11), 1591–1597.

Oudkerk Pool, M., Ellerbroek, P.M., Ridwan, B.U., Goldschmeding, R., von Blomberg, B.M., Pena, A.S., Dolman, K.M., Bril, H., Dekker, W., Nauta, J.J. et al. (1993). Serum antineutrophil cytoplasmic autoantibodies in inflammatory bowel disease are mainly associated with ulcerative colitis. A correlation study between perinuclear antineutrophil cytoplasmic autoantibodies and clinical parameters, medical, and surgical treatment. *Gut* 34(1), 46–50.

Peeters, M., Nevens, H., Baert, F., Hiele, M., de Meyer, A.M., Vlietinck, R. and Rutgeerts, P. (1996). Familial aggregation in Crohn's disease: Increased age-adjusted risk and concordance in clinical characteristics. *Gastroenterology* 111(3), 597–603.

Peltekova, V.D., Wintle, R.F., Rubin, L.A., Amos, C.I., Huang, Q., Gu, X., Newman, B., Van Oene, M., Cescon, D., Greenberg, G. et al. (2004). Functional variants of OCTN cation transporter genes are associated with Crohn disease. *Nat Genet* 36(5), 471–475.

Poppe, D., Tiede, I., Fritz, G., Becker, C., Bartsch, B., Wirtz, S., Strand, D., Tanaka, S., Galle, P.R. and Bustelo, X.R. (2006). Azathioprine suppresses ezrin-radixin-moesin-dependent T cell-APC conjugation through inhibition of Vav guanosine exchange activity on Rac proteins. *J Immunol* 176(1), 640–651.

Present, D.H., Rutgeerts, P., Targan, S., Hanauer, S.B., Mayer, L., van Hogezand, R.A., Podolsky, D.K., Sands, B.E., Braakman, T., DeWoody, K.L. et al. (1999 May 6). Infliximab for the treatment of fistulas in patients with Crohn's disease. *N Engl J Med* 340(18), 1398–1405.

Reinhard, C. and Rioux, J.D. (2006). Role of the IBD5 susceptibility locus in the inflammatory bowel diseases. *Inflamm Bowel Dis* 12(3), 227–238.

Rioux, J.D., Xavier, R.J., Taylor, K.D., Silverberg, M.S., Goyette, P., Huett, A., Green, T., Kuballa, P., Barmada, M.M., Datta, L.W. et al. (2007). Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat. Genet* 39(5), 596–604

Roth, M.P., Petersen, G.M., McElree, C., Feldman, E. and Rotter, J.I. (1989). Geographic origins of Jewish patients with inflammatory bowel disease. *Gastroenterology* 97(4), 900–904.

Roussomoustakaki, M., Satsangi, J., Welsh, K., Louis, E., Fanning, G., Targan, S., Landers, C. and Jewell, D.P. (1997). Genetic markers may predict disease behavior in patients with ulcerative colitis. *Gastroenterology* 112(6), 1845–1853.

Russel, M.G., Dorant, E., Volovics, A., Brummer, R.J., Pop, P., Muris, J.W., Bos, L.P., Limonard, C.B. and Stockbrugger, R.W. (1998). High incidence of inflammatory bowel disease in The Netherlands: Results of a prospective study. The South Limburg IBD Study Group. *Dis Colon Rectum* 41(1), 33–40.

Russell, R.K., Drummond, H.E., Nimmo, E.E., Anderson, N., Smith, L., Wilson, D.C., Gillett, P.M., McGrogan, P., Hassan, K., Weaver, L.T. et al. (2005). Genotype-phenotype analysis in childhood-onset Crohn's disease: NOD2/CARD15 variants consistently predict phenotypic characteristics of severe disease. *Inflamm Bowel Dis* 11(11), 955–964.

Sands, B.E., Blank, M.A., Patel, K. and van Deventer, S.J. (2004). ACCENT II study long-term treatment of rectovaginal fistulas in Crohn's disease: Response to infliximab in the ACCENT II Study. *Clin Gastroenterol Hepatol* 2(10), 912–920.

Sans, M., Artieda, M., Diego, T., Cara, C., Rodrigo, L., Barreiro, M., Arroyo, M.T., Aceituno, M., Alvarez-Lobos, M., Yague, J. et al. (2006). IBDChip: A new strategy to predict clinical course and development of complications in patients with inflammatory bowel disease (IBD). *Gastroenterology* 130(4Suppl 2), A52–A53.

Satsangi, J., Rosenberg, W.M.C. and Jewell, D.P. (1994). The prevalence of inflammatory bowel disease in relatives of patients with Crohn's disease. *Eur J Gastroenterol Hepatol* 6, 413–416.

Satsangi, J., Welsh, K.I., Bunce, M., Julier, C., Farrant, J.M., Bell, J.I. and Jewell, D.P. (1996 May 4). Contribution of genes of the major histocompatibility complex to susceptibility and disease phenotype in inflammatory bowel disease. *Lancet* 347(9010), 1212–1217.

Schreiber, S., Rosenstiel, P., Albrecht, M., Hampe, J. and Krawczak, M. (2005). Genetics of Crohn disease, an archetypal inflammatory barrier disease. *Nat Rev Genet* 6(5), 376–388.

Schulte, C.M., Dignass, A.U., Goebell, H., Roher, H.D. and Schulte, K.M. (2000). Genetic factors determine extent of bone loss in inflammatory bowel disease. *Gastroenterology* 119(4), 909–920.

Shanahan, F., Duerr, R.H., Rotter, J.I., Yang, H., Sutherland, L.R., McElree, C., Landers, C.J. and Targan, S.R. (1992). Neutrophil autoantibodies in ulcerative colitis: Familial aggregation and genetic heterogeneity. *Gastroenterology* 103(2), 456–461.

Silverberg, M.S., Daly, M.J., Moskovitz, D.N., Rioux, J.D., McLeod, R.S., Cohen, Z., Greenberg, G.R., Hudson, T.J., Siminovitch, K.A. and Steinhart, A.H. (2001). Diagnostic misclassification reduces the ability to detect linkage in inflammatory bowel disease genetic studies. *Gut* 49(6), 773–776.

Silverberg, M.S., Satsangi, J., Ahmad, T., Arnott, I.D., Bernstein, C.N., Brant, S.R., Caprilli, R., Colombel, J.F., Gasche, C., Geboes, K. et al. (2005). Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: Report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. *Can J Gastroenterol* 19(Suppl A), 5–36.

Stoll, M., Corneliussen, B., Costello, C.M., Waetzig, G.H., Mellgard, B., Koch, W.A., Rosenstiel, P., Albrecht, M., Croucher, P.J., Seegert, D. et al. (2004). Genetic variation in DLG5 is associated with inflammatory bowel disease. *Nat Genet* 36(5), 476–480.

Tenesa, A., Noble, C., Satsangi, J. and Dunlop, M. (2006). Association of DLG5 and inflammatory bowel disease across populations. *Eur J Hum Genet* 14(3), 259–260.

Thjodleifsson, B., Sigthorsson, G., Cariglia, N., Reynisdottir, I., Gudbjartsson, D.F., Kristjansson, K., Meddings, J.B., Gudnason, V., Wandall, J.H., Andersen, L.P. et al. (2003). Subclinical intestinal inflammation: An inherited abnormality in Crohn's disease relatives?. *Gastroenterology* 124(7), 1728–1737.

Timmer, A. and Goebell, H. (1999a). Incidence of ulcerative colitis, 1980–1995. A prospective study in an urban population in Germany. *Z Gastroenterol* 37(11), 1079–1084.

Timmer, A., Breuer-Katschinski, B. and Goebell, H. (1999b). Time trends in the incidence and disease location of Crohn's disease

1980-1995: A prospective analysis in an urban population in Germany. *Inflamm Bowel Dis* 5(2), 79–84.

Torok, H.P., Glas, J., Tonenchi, L., Lohse, P., Muller-Myhsok, B., Limbersky, O., Neugebauer, C., Schnitzler, F., Seiderer, J., Tillack, C. et al. (2005). Polymorphisms in the DLG5 and OCTN cation transporter genes in Crohn's disease. *Gut* 54(10), 1421–1427.

Tremelling, M., Waller, S., Bredin, F., Greenfield, S. and Parkes, M. (2006). Genetic variants in TNF-alpha but not DLG5 are associated with inflammatory bowel disease in a large United Kingdom cohort. *Inflamm Bowel Dis* 12(3), 178–184.

Turner, D., Walsh, C.M., Steinhart, A.H. and Griffiths, A.M. (2007). Response to corticosteroids in severe ulcerative colitis: A systematic review of the literature and a meta-regression. *Clin Gastroenterol Hepatol* 5(1), 103–110.

Tysk, C., Lindberg, E., Jarnerot, G. and Floderus-Myrhed, B. (1988). Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking. *Gut* 29(7), 990–996.

Urano, W., Taniguchi, A., Yamanaka, H., Tanaka, E., Nakajima, H., Matsuda, Y., Akama, H., Kitamura, Y. and Kamatani, N. (2002). Polymorphisms in the methylenetetrahydrofolate reductase gene were associated with both the efficacy and the toxicity of methotrexate used for the treatment of rheumatoid arthritis, as evidenced by single locus and haplotype analyses. *Pharmacogenetics* 12(3), 183–190.

Urcelay, E., Mendoza, J.L., Martinez, A., Fernandez, L., Taxonera, C., Diaz-Rubio, M. and de la Concha, E.G. (2005). IBD5 polymorphisms in inflammatory bowel disease: Association with response to infliximab. *World J Gastroenterol* 11(8), 1187–1192.

van Bodegraven, A.A., Curley, C.R., Hunt, K.A., Monsuur, A.J., Linskens, R.K., Onnie, C.M., Crusius, J.B., Annese, V., Latiano, A., Silverberg, M.S. et al. (2006). Genetic variation in myosin IXB is associated with ulcerative colitis. *Gastroenterology* 131(6), 1768–1774.

van Bodegraven, A.A., Curley, C.R., Hunt, K.A., Monsuur, A.J., Linskens, R.K., Onnie, C.M., Crusius, J.B., Annese, V., Latiano, A., Silverberg, M.S. et al. (2006). Genetic variation in myosin IXB is associated with ulcerative colitis. *Gastroenterology* 131, 1768–1774.

Van Heel, D.A., Fisher, S.A., Kirby, A., Daly, M.J., Rioux, J.D. and Lewis, C.M. (2004 Apr 1). Genome Scan Meta-Analysis Group of the IBD International Genetics Consortium Inflammatory bowel disease susceptibility loci defined by genome scan meta-analysis of 1952 affected relative pairs. *Hum Mol Genet* 13(7), 763–770.

Vermeire, S., Peeters, M., Vlietinck, R., Joossens, S., Den Hond, E., Bulteel, V., Bossuyt, X., Geypens, B. and Rutgeerts, P. (2001). Anti-Saccharomyces cerevisiae antibodies (ASCA), phenotypes of IBD, and intestinal permeability: A study in IBD families. *Inflamm Bowel Dis* 7(1), 8–15.

Vermeire, S., Louis, E., Rutgeerts, P., De Vos, M., Van Gossum, A., Belaiche, J., Pescatore, P., Fiasse, R., Pelckmans, P., Vlietinck, R. et al. (2002). Belgian Group of Infliximab Expanded Access Program and Fondation Jean Dausset CEPH, Paris, France. NOD2/CARD15 does not influence response to infliximab in Crohn's disease. *Gastroenterology* 123(1), 106–111.

Vermeire, S., Pierik, M., Hlavaty, T., Claessens, G., van Schuerbeeck, N., Joossens, S., Ferrante, M., Henckaerts, L., Bueno de Mesquita, M., Vlietinck, R. et al. (2005). Association of organic cation transporter risk haplotype with perianal penetrating Crohn's disease but not with susceptibility to IBD. *Gastroenterology* 129(6), 1845–1853.

Vermeire, S. and Rutgeerts, P. (2005). Current status of genetics research in inflammatory bowel disease. *Genes Immun* 6(8), 637–645.

Wehkamp, J., Harder, J., Weichenthal, M., Schwab, M., Schaffeler, E., Schlee, M., Herrlinger, K.R., Stallmach, A., Noack, F., Fritz, P. et al. (2004). NOD2 (CARD15) mutations in Crohn's disease are associated with diminished mucosal alpha-defensin expression. *Gut* 53(11), 1658–1664.

Wild, G.E. and Rioux, J.D. (2004). Genome scan analyses and positional cloning strategy in IBD: Successes and limitations. *Best Pract Res Clin Gastroenterol* 18(3), 541–553.

Witte, J., Shivananda, S., Lennard-Jones, J.E., Beltrami, M., Politi, P., Bonanomi, A., Tsianos, E.V., Mouzas, I., Schulz, T.B., Monteiro, E. et al. (2000). Disease outcome in inflammatory bowel disease: Mortality, morbidity and therapeutic management of a 796-person inception cohort in the European Collaborative Study on Inflammatory Bowel Disease (EC-IBD). *Scand J Gastroenterol* 35(12), 1272–1277.

Yamazaki, K., Takazoe, M., Tanaka, T., Ichimori, T., Saito, S., Iida, A., Onouchi, Y., Hata, A. and Nakamura, Y. (2004). Association analysis of SLC22A4, SLC22A5 and DLG5 in Japanese patients with Crohn disease. *J Hum Genet* 49(12), 664–668.

Yang, H., McElree, C., Roth, M.P., Shanahan, F., Targan, S.R. and Rotter, J.I. (1993). Familial empirical risks for inflammatory bowel disease: Differences between Jews and non-Jews. *Gut* 34(4), 517–524.

Yang, P., Jarnerot, G., Danielsson, D., Tysk, C. and Lindberg, E. (1995). P-ANCA in monozygotic twins with inflammatory bowel disease. *Gut* 36(6), 887–890.

Yates, C.R., Krynetski, E.Y., Loennechen, T., Fessing, M.Y., Tai, H.L., Pui, C.H., Relling, M.V. and Evans, W.E. (1997 Aug 15). Molecular diagnosis of thiopurine S-methyltransferase deficiency: Genetic basis for azathioprine and mercaptopurine intolerance. *Ann Intern Med* 126(8), 608–614.

CHAPTER

Asthma Genomics

Scott T. Weiss, Benjamin A. Raby and Juan C. Celedón

INTRODUCTION

Asthma is a complex disease affecting over 300 million individuals in the developed world (Palmer and Cookson, 2000). Ninety percent of all asthma cases, including asthma in adults, have their origin in childhood. Of concern are the increases in asthma prevalence (CDC, 1995) and hospitalization rates (Weiss et al., 1993). Between 1980 and 1994, the self-reported prevalence of asthma in the United States of America increased from 30.7 to 53.8 per 1000, an increase of 75% (CDC, 1995). This increase has been accompanied by a similar increase in health care utilization and mortality over the same time period (CDC, 1995). An estimated 12.6 billion dollars were spent on the diagnosis and management of asthma in the United States in 1998, of which 58% were direct medical expenditures (Sullivan and Strassels, 2002). Despite the availability of several classes of therapeutic agents for asthma, it has been estimated that as many as one-half of asthmatic patients do not respond to treatment with β_2-agonists, leukotriene antagonists, or inhaled corticosteroids (Drazen et al., 2000; Liggett, 2001; Silverman et al., 2002). It is quite likely that genetics and genomics will significantly impact asthma in the next 5 years particularly in the area of prediction of clinical events.

ASTHMA: BASIC PATHOBIOLOGY

Asthma is a clinical syndrome of unknown etiology characterized by reversible episodes of airflow obstruction, airway hyperresponsiveness, and a chronic inflammatory process of the airways of which mast cells, eosinophils, T-lymphocytes, epithelial cells, and airway smooth muscle cells play a prominent role (Elias et al., 2003). Figure 46.1 shows important pathobiologic features of asthma. $CD4^+$ lymphocytes produce IL-3, IL-4, IL-5, IL-13, and GM-CSF and thereby promote the synthesis of IgE, an important allergic effector molecule. Chemokines, such as eotaxin, RANTES, and IL-8 produced by epithelial and inflammatory cells, serve to amplify and perpetuate the inflammatory events. Several bronchoactive mediators, such as histamine, leukotrienes, and neuropeptides are released into the airways and precipitate an asthma attack by causing airway smooth muscle constriction, mucus secretion, and edema. Over time, there is smooth muscle growth and the deposition of subepithelial connective tissue, a process referred to as airway remodeling. Clinically, asthmatics have difficulty exhaling air because of an increase in airway resistance that is a consequence of smooth muscle contraction, inflammation, and remodeling (Figure 46.2). Physiological impairment is quantitated most commonly by the forced expiratory volume in 1 s (FEV_1) (Figure 46.2). FEV_1 is the volume of air a person can "blow out" in 1 s and is very useful as a measurement of lung function because it is easily obtained, reproducible, and correlated with asthma severity and therapeutic responses (ATS, 1987). While it is difficult to be precise, it is reasonable to estimate that although there is general agreement on the physiology described above, the actual genetic and genomic architecture of these events is just beginning to be described. The above description of events, while true, is likely

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

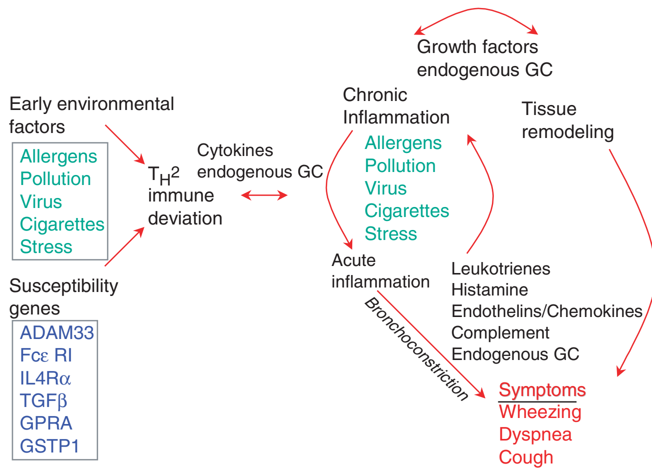
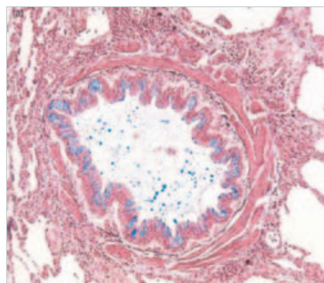


Figure 46.1 Asthma pathobiology 1.



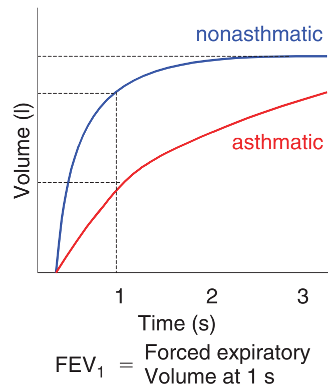


Figure 46.2 Asthma pathobiology 2.

to only represent 15–20% of the total pathobiology that will be ultimately elucidated by genomic methods.

PREDISPOSITION (GENETIC AND NON-GENETIC) TO ASTHMA

Non-Genetic Predisposition

Eighty percent of childhood asthmatics exhibit hypersensitivity to indoor aeroallergens (Sears et al., 1991), and high concentrations of indoor allergens can worsen asthma symptoms in sensitized individuals (Rosenstreich et al., 1997). In a small follow-up study of atopic children of atopic parents, exposure to high concentrations of dust mite allergen in the first year of life was associated with higher risk for sensitization and asthma at age 11 (Sporik et al., 1990). Although certain home characteristics are significant predictors of dust mite, cat, and cockroach allergen exposure, they may be poor at predicting whether low levels of allergen are present in a home (Chew et al., 1998). Since "low levels" of indoor allergens may lead to sensitization in susceptible individuals, direct measurement of allergen concentrations in house dust have been recommended for use in epidemiologic studies (Chew et al., 1998).Viral respiratory infections are a cause of asthma exacerbations but not a proven cause of disease onset. Passive (primarily *in utero*) and active cigarette smoking are the other major environmental cause of the disease. It is likely that the response to these environmental exposures has a genetic and or genomic basis; however, virtually nothing is known about the important genetic predictors of adverse response to these environmental agents. Polymorphisms in IL13 and ADRB2 seem to predict response to passive and active cigarette smoking, respectively, but nothing is known about the relationship of genetic polymorphisms to other important environmental exposures.

Genetic Predisposition

Data from a significant number of studies indicate familial aggregation of asthma (Jenkins et al., 1997; Van Arsdel and Motulsky, 1959). A method frequently used to establish the existence of familial aggregation of a trait is to calculate the recurrence risk to relatives of type R (λ). Using data from Sibbald et al. (1980) resulted in a λ of 3.3 for asthma. Familial aggregation, however, could result from genetic factors or a shared environment. Estimates of the heritability of asthma in several twin studies conducted around the world have ranged from 36% to 79% (Duffy et al., 1990; Edfors-Lubs, 1971; Hopp et al., 1984; Lubs, 1972; Nieminen et al., 1991; Sibbald et al., 1980), with the highest values coming from studies that had a more comprehensive phenotypic assessment of asthma (Sandford et al., 1996).

GENOME-WIDE LINKAGE ANALYSES OF ASTHMA AND ITS INTERMEDIATE PHENOTYPES

To date, 16 groups have reported results of genome-wide linkage analysis for asthma phenotypes (Table 46.1) (Bleecker et al., 1999; Celedon et al., 2007; CSGA, 1997; Daniels et al., 1996; Dizier et al., 2000; Evans et al., 2004; Ferreira et al., 2005; Haagerup et al., 2002; Hakonarson et al., 2002; Hersh et al., 2007b; Koppelman et al., 2002; Laitinen et al., 2001; Ober et al., 1998; Pillai et al., 2006; Wjst et al., 1999; Xu et al., 2000, 2001a, b; Yokouchi et al., 2000). Several of these groups have published second-generation surveys with larger numbers of subjects and/or markers (Blumenthal et al., 2004; Brasch-Andersen et al., 2006; Dizier et al., 2005; Huang et al., 2003; Meyers et al., 2005; Ober et al., 2000a, b; Xu et al., 2001a, b). In spite of obvious heterogeneity in the design of these studies, 13 regions have shown significant evidence of linkage to asthma or its intermediate phenotypes: chromosomes 2p16 (to airway hyperresponsiveness (AHR), in families from Europe, Australia, and the United States) (Pillai et al., 2006), 2p25 (to AHR, in Chinese families) (Xu et al., 2001a, b), 2q33 (to eosinophil count in Australian

Table 46.1 Results of genome-wide linkage analyses of asthma and/or its intermediate phenotypes[a]

Reference	Population	Number of families (subjects)	Linkage to asthma**	Linkage to intermediate phenotypes of asthma
Daniels et al. (1996)	Australian	80 (363)	Not assessed	Skin test index=**11q12**; AHR[b] = 4q, 7p; eosinophil count = 6p; atopy = 6p, 13q; total serum IgE= 16q
Xu et al. (2001a, b); Huang et al. (2003); Blumenthal et al. (2004)	North American (white, Hispanic, African American)	266–287 (885–1931)	6p (whites), 1p (Hispanics), **11q13** (African Americans)	Dust mite allergy=**19p13** (whites); atopy = 21q (Hispanics)
Ober et al. (2000)	Hutterites	1 (693)[c]	1p, 3p, 5q, 8p	AHR = 5p, 19p; dust mite allergy = 2q; mold allergy = 3q, 11p
Wjst et al. (1999)	German and Swedish	97 (415)		
Xu et al. (2000); Koppelman et al. 2002; Meyers et al. (2005)	Dutch	200 (1174)	3p, 5q	AHR = 3p, 5q; total serum IgE: 2q, 3q, 5q, **7q21**, 12q, 13q; eosinophil count: 2q, 15q, 17q
Yokouchi et al. (2000)	Japanese	47 (197)	4q, **5q31-q33**, 6p, 12q, 13q	Not assessed
Dizier M et al. (2000, 2005)	French	107–295 (not available)	1p (asthma and allergic rhinitis)	Eosinophil count: 12q
Laitinen et al. (2001)	Finnish	86 (443)	4q, 7p14-p15 (asthma and increased total serum IgE)	Total serum IgE=**7p14-p15**
Xu et al. (2001a, b)	Chinese	533 (2551)	Not assessed	AHR=**2p25**, 19q; total serum IgE: 1q; cockroach allergy = 4q
Haagerup et al. (2002); Brasch-Andersen et al. (2006)	Danish	100 (424)	1p, 5q, 6p, **12q24**	Total serum IgE: 3q, 5q, 6p; atopy = 6p
Hakonarson et al. (2002)	Icelandic	175 (1134)	**14q24**	Not assessed
Van Eerdewegh et al. (2002)	North American and British	460 (920)	**20p13**	Not assessed (results for AHR not presented separately from asthma)
Ferreira et al. (2005)	Australian	202 (591)	1q, 4p, 11p, 17q, 18p, 19p	AHR = 6p, 20q; atopy = 2q, 3q, 6p, 17q, 20q; dust mite allergy=**20q13**; total serum IgE = 10q
Evans et al. (2004)	Australian	539 (2360)	Not assessed	Eosinophil count=**2q33**, 8q
Pillai et al. (2006)	North American, European, Australian	414 (1555)		AHR=**2p16**, 4p
Celedon et al. (2007); Raby et al. (2007)	Costa Rican	8 (638)	12q	AHR=**12q24** (nonsmokers); total serum IgE=**20p12** (males)

** Complete linkage results for all chromosomes not published.

[a] Only regions showing suggestive or significant evidence of linkage (as defined by Lander and Kruglyak) to asthma and/or its intermediate phenotypes (other than measures of lung function) are shown. Regions with significant evidence of linkage to asthma and/or its intermediate phenotypes are **bold**.

[b] AHR = airway hyperresponsiveness.

[c] Large pedigree divided into 10–20 sub-pedigrees for data analysis.

twins) (Evans et al., 2004), 5q31-q33 (to mite-sensitive asthma in a Japanese population) (Yokouchi et al., 2000), 7p14-p15 (to total serum IgE in a Finnish population) (Laitinen et al., 2004), 7q21 (to total serum IgE in a Dutch population) (Xu et al., 2000), 11q12-q13 (to atopy in Australians and asthma in African Americans) (Huang et al., 2003; Laitinen et al., 2004), 12q24 (to asthma in Denmark (Brasch-Andersen et al., 2006) and AHR in Costa Rica (Celedon et al., 2007), 14q24 (to asthma in an Icelandic population) (Hakonarson et al., 2002), 19p13 (to dust mite allergy in whites in the United States) (Blumenthal et al., 2004), 20p12 (to total IgE among Costa Rican males) (Raby et al., 2007), 20p13 (to asthma in families from the United States and United Kingdom) (Van Eerdewegh et al., 2002), and 20q13 (to dust mite allergy in Australians) (Ferreira et al., 2005).

In spite of the relatively large number of genomic regions linked to asthma-related phenotypes, only five potential asthma-susceptibility genes (PDH finger protein 11 [*PHF11*], dipeptidylpeptidase 10 [*DPP10*], disintegrin and metalloprotease 33 [*ADAM 33*], G protein-coupled receptor for asthma susceptibility [*GPR154* or *NPSR1*], and human leukocyte antigen G [*HLA-G*]) have been identified by a positional cloning approach (Allen et al., 2003; Laitinen et al., 2004; Nicolae et al., 2005; Van Eerdewegh et al., 2002; Zhang et al., 2003). In all cases, significant linkage between a chromosomal region and asthma phenotypes in a genome-wide linkage analysis was followed by fine-mapping studies of linkage and association in the linked region. Replication studies of the association between asthma phenotypes and these potential asthma-susceptibility genes have yielded inconsistent results, and no specific asthma-susceptibility variants have been identified within these genes (Hersh et al., 2007a, b). However, results of multiple association studies (Hersh et al., 2007a, b), as well as functional data in rodents and humans provide strong support for a significant role of GPR154 in asthma pathogenesis (Laitinen et al., 2004).

CANDIDATE-GENE ASSOCIATION STUDIES OF ASTHMA

Although there are more than 500 genetic association studies for asthma, most suffer from methodological problems including small sample size, non-comprehensive coverage of the gene(s) of interest, failure to correct for multiple testing, and (for case-control studies) lack of adjustment for population stratification. To date, there have been reports of positive associations between variants in over 100 genes and asthma phenotypes (Ober and Hoffjan, 2006). Using the gene as the unit of replication (a very liberal criterion), more than 40 associations were replicated in at least two populations and 17 associations were replicated in at least 5 populations (Allen et al., 2003; Hersh et al., 2007a, b; Ober and Hoffjan, 2006). On the other hand, few associations have been replicated for specific polymorphisms in genes for a particular phenotype and in the same direction, suggesting that LD with adjacent variants and/or bias explains a significant proportion of these findings.

Variants in nine positional candidate genes (IL4, IL4RA, IL13, ADRB2, CD14, TNFA, HLA-DRB1, HLA-DQB1, and FCERIB) have been associated with asthma phenotypes in at least ten populations (Ober and Hoffjan, 2006). Through activation of its receptor, interleukin 4 (IL4) stimulates production of total serum IgE. The IL4 and IL4 receptor α chain (IL4RA) loci are on genomic regions linked to asthma phenotypes (5q31 for IL4 and 16p12 for ILRA). A functional SNP in the promoter of IL4 has been associated with total serum IgE (Rosenwasser et al., 1995), asthma (Rosenwasser et al., 1995), rhinitis (Zhu et al., 2000), asthma severity (Sandford et al., 2000), and atopic dermatitis (Novak et al., 2002). SNPs in exons of IL4RA have been associated with asthma (Mitsuyasu et al., 1998), total serum IgE (Howard et al., 2002), atopic dermatitis (Hershey et al., 1997), and asthma severity (Rosa-Rosa et al., 1999). The interleukin 13 (IL13) and monocyte differentiation antigen CD14 (CD14) genes are on chromosome 5q31-33, a region linked to asthma phenotypes. Experiments in rodents and humans support a critical role of interleukin 13 in asthma pathogenesis (Grunig et al., 1998). Functional SNPs in the promoter and coding regions of IL13 have been associated with asthma (Heinzmann et al., 2003; Howard et al., 2001; van der Pouw Kraan et al., 1999), airway responsiveness (Howard et al., 2001), atopy (DeMeo et al.,2002; Howard et al., 2001), eosinophilia (Hunninghake et al., 2007), and total serum IgE (Graves et al., 2000; Hunninghake et al., 2007). CD14 is a receptor for bacterial cell wall components that may influence immune responses (Guerra et al., 2004). A functional SNP in the promoter of CD14 has been associated with total serum IgE (Baldini et al., 1999), atopy (Ober et al., 2000a, b), food allergy and non-atopic asthma (Woo et al., 2003), and airway responsiveness (O'Donnell et al., 2004). Tumor necrosis factor alpha (TNFA) is a proinflammatory cytokine (Wang et al., 2004), and HLA class II molecules are candidates for controlling immune responses to allergens. The TNFA and human leukocyte antigen DRB1 (HLA-DRB1) genes are on chromosome 6p21, a genomic region linked to asthma phenotypes. A functional variant in the promoter of TNFA has been associated with asthma (Moffatt and Cookson, 1997), airway responsiveness (Moffatt and Cookson, 1997), atopy (Castro et al., 2000), and total serum IgE (Shin et al., 2004). Functional SNPs and/or haplotypes in HLA-DRB1 have been associated with sensitization to specific allergens (Ansari et al., 1989), atopy (Aron et al., 1996), asthma (Di Somma et al., 2003), and total serum IgE (Moffatt et al., 2001). The gene for the beta chain of the high affinity receptor for IgE (FCERIB) is on chromosome 11q13, a genomic region linked to atopy. SNPs and/or haplotypes in FCERIB have been associated with asthma (Cox et al., 1998; Green et al., 1998), total serum IgE (Li and Hopkin, 1997; Shirakawa et al., 1994), and atopy (Shirakawa et al., 1994). Although there are studies refuting the association between each of these nine positional candidate genes and asthma phenotypes, many of these negative studies lacked adequate statistical power to exclude weak-to-moderate genetic effects. In addition, multiple studies have found an association between known functional SNPs in seven

of these genes (IL4, IL4RA, CD14, IL13, TNFA, ADRB2, and HLA-DRB1) and asthma phenotypes. Thus, current evidence suggests that these genes influence asthma pathogenesis.

GENOME-WIDE ASSOCIATION STUDIES OF ASTHMA

To date, only one genome-wide association study of asthma has been published (Moffatt et al., 2007) After testing for association between a panel of 317,000 SNPs and asthma in subjects with and without asthma, Cookson and coworkers identified a gene on chromosome 17q21 (ORMDL3) as a potential susceptibility gene for asthma. These findings were then replicated in two independent cohorts. Other genome-wide association studies of asthma and its intermediate phenotypes are ongoing.

ASTHMA GENOMICS

Complementing these genetic surveys are a handful of studies examining global gene expression patterns in experimental models of asthma and allergic inflammation, in addition to several clinical studies (Brutsche et al., 2002; Guajardo et al., 2005; Hansel et al., 2005; Karp et al., 2000; Laprise et al., 2004; Lilly et al., 2005; Zimmermann et al., 2003; Zou et al., 2002a, b). Akin to genome-wide SNP association studies, the primary strength of genome-wide microarray surveys of gene expression is that they are hypothesis-free in that no assumptions regarding the underlying pathobiology of the conditions under study are placed on data gathering, analysis or interpretation. As such, these approaches have the potential to reveal previously unsuspected molecules or pathways relating to asthma and allergy. Table 46.2 summarizes microarray studies in asthma assessing *in vivo* gene expression (studies of cultured asthma-relevant cells, transgenic mouse models of asthma, or pharmacogenetics are not included). In one of the first studies, Karp and colleagues combined pulmonary tissue gene expression data with SNP genotypes (an example of integrative genomics) in segregating backcross mice to identify C5 as a primary mediator of allergic response (Karp et al., 2000). They demonstrated that: (1) C5 was the only differentially expressed gene following allergen challenge among high- (A/J) and low- (C3H/HeJ) responder mice that also mapped to an AHR locus on chromosome 2; (2) C5 expression is strain dependent and correlates with a 2-bp missense mutation; and (3) blockade of C5 signaling inhibits allergen-mediated responses. Human studies suggest that C5 expression is increased in allergic airway inflammation (Krug et al., 2001) and that C5 genetic variation influences asthma susceptibility (Hasegawa et al., 2004). In a second study, Zimmermann and colleagues (2003) studied patterns of differential gene expression in two models of allergic pulmonary inflammation and identified a common set of differentially expressed genes relating to arginine metabolism, thus implicating this previously unsuspected pathway in the pathogenesis of allergic inflammation. It is interesting that initial genetic association studies of arginase I and II suggest that genetic variants in this pathway are also related to asthma susceptibility (Li et al., 2006).

In contrast to the studies in mouse models, the results from six published studies of gene expression in human populations of asthmatics have been somewhat less exciting in that few truly novel biologic pathways have been reported. We note that these studies evaluated relatively small number of individuals and varied greatly in the tissues studied, analytic methods used, and definitions of significant differential gene expression. It is, therefore, not surprising that the gene signatures identified have little in common with each other. However, it is notable that each of these studies identified genes previously implicated in asthma pathobiology, and several of these genes harbor genetic variation previously associated with asthma-susceptibility and severity (see Table 46.2). It is therefore likely that many of the other genes identified in these studies play previously unrecognized roles in the pathogenesis of asthma. The challenge remains separating the wheat from the chaff. Given the size of these differentially expressed asthma gene sets – as many as 324 genes in one study of nasal epithelium (Guajardo et al., 2005) – experimental validation using animal models is not feasible on a large scale. One approach would be to test all of the identified genes for evidence of genetic association in asthma cohorts. As an example, Laprise and colleagues have demonstrated association to asthma with five polymorphisms (including one non-synonymous variant) in the CX3CR1 fractalkine receptor (Tremblay et al., 2006). This gene was among those demonstrating greatest differential expression in bronchial epithelium derived from asthmatics (Laprise et al., 2004).

The most important cumulative result of these early studies is that together they demonstrate the feasibility of using expression profiling for the study of asthma, and should prompt a second generation of genomic epidemiologic studies. We suggest that the first step include repeating some of these earlier study designs but using considerably larger sample size, with a focus on identifying gene expression signatures and global patterns of gene expression (rather than specific genes). These studies should be large enough to enable splitting of the cohort in two to allow follow-up validation studies of the predictive accuracy of any identified signatures – such as assessment of the predictive power of identified sets in differentiating asthma affection status, atopic status, or disease severity. In addition, these studies should include collection of DNA from all subjects to enable integration of gene expression data with SNP genotype data for the identification of asthma-associated regulatory variation. In this way, investigators will not only have a catalog of differentially expressed asthma genes, but also a set of putative functional variants for follow-up in well-designed genetic association studies.

PHARMACOGENETICS

The two main types of asthma drugs are the so-called "reliever" drugs that target the acute bronchoconstriction and the so-called

Table 46.2 Published expression microarray studies in asthma**

Study	Organism	Tissue/cell type	Phenotype assessed	Sample size	Major findings
Karp et al. (2000)	A/J and C3H/HeJ mice	Lung tissue	Ovalbumin challenge	4 A/J, C3H/HeJ, F1 8 F1xA/J backcross	21 genes overlapping all cross-strain comparisons, including C5, on chromosome 2 (airways responsiveness QTL)
Zou et al. (2002a, b)	Cynomolgus monkeys	Lung tissue	IL4 or Ascaris suum inhalational challenge	Controls = 6 IL4 = 1, Ascaris = 6	149 genes (2.5-fold change), including eotaxin, MCP-1, PARC, COL2A1
Zimmermann et al. (2003); Munitz et al. (2007)	Balb/c mice	Lung tissue	Allergen challenge with ovalbumin or A. Fumigatus	Ova = 3 A. Fumigatus = 3 Control = 5	291 genes overlapping. l-arginine metabolism genes identified: CAT2, Arginase I and Arginase II. CD48 antigen later implicated following allergen challenge
Brutsche et al. (2002)	Human	Peripheral blood mononuclear cells	Atopic status	Atopy and Asthma = 18 Atopy, no asthma = 8 Controls = 14	Developed composite score (10 genes) with 96% sensitivity and 92% specificity for differentiating atopy from controls. Included IL1RA1, IL6, RET, CD71
Laprise et al. (2004)	Human	Bronchial wall	Asthma versus normal pre- and post-ICS	Asthma = 4 Normal = 4	79 genes, including 21 previously implicated in asthma, including ALOX15, NOS2A, TRAa, MUC5A, COL2A1
Hansel et al. (2005)	Human	Peripheral blood $CD4^+$ lymphocytes	Asthma severity	Mild, no ICS = 5 Mild, ICS = 5 Severe = 5	40 genes, including TGFB1, UTS2, TRDa@, JUND
Guajardo et al. (2005)	Human	Nasal epithelium	Asthma exacerbation	Control = 10 Stable = 10 Exacerbation = 10	324 genes (2-fold change) including IL1b, Defensinb1, IL1R1, retinoic acid receptor
Lilly et al. (2005)	Human	Bronchial epithelium	Segmental airway allergen challenge	5 asthmatics pre- and post- challenge	149 genes (2-fold change) including IL1RN, IL1b, PTGFR, MCP-1, NFKB, MIP1b
Youssef et al. (2007)	Human	Peripheral blood basophils	Response to IgE receptor cross-linking	3 nonreleasers and 5 releasers, pre- and post- cross-linking	253 differentially expressed genes following “releaser” basophil stimulation, including RANTES, MIP1β, VEGF, heat-shock proteins

**Studies of cultured asthma-relevant cells, transgenic mouse models of asthma, or pharmacogenetics are not included.

"controller" drugs that are used to reduce the severity of airway inflammation and the severity and frequency of obstruction (CSGA, 1997). The main reliever drugs are rapid-acting β_2-agonists (e.g., albuterol, metaproterenol, pirbuterol, levalbuterol), which relax the bronchial smooth muscle by activating β_2-adrenergic receptors. This is the treatment of choice for very mild asthma. For moderate and severe asthma, the reliever treatment is usually combined with controller treatment. The two commonly used classes of controller agents are the inhaled glucocorticoids and the leukotriene inhibitors. Inhaled glucocorticoids (e.g., budesonide, beclomethasone, flunisolide, and fluticasone) and leukotriene inhibitors (e.g., montelukast and zafirlukast) modify the inflammatory micro-environment of the airway to reduce airway obstruction and hyperresponsiveness.

There is large interindividual variation in the treatment response to asthma medications (Drazen et al., 2000; Szefler et al., 2002). For example, in a study by Malmstrom and colleagues comparing the efficacy of the inhaled steroid beclomethasone (200ug bid) with the leukotriene antagonist montelukast (10 mg/qd), it is clear that both drugs are effective over a 12-week course of treatment with a mean increase in FEV_1 of 13.1% for the beclomethasone group and 7.4 % for the montelukast group (Figure 46.3a) (Malmstrom et al., 1999). However, when these same data are viewed from a different perspective, focusing on the number of individuals as a function of percent change in FEV_1 from baseline, it is clear that many patients had little response (Figure 46.3b). In fact, 22% of patients appear to have had an adverse response to treatment with a decline in FEV_1 at 12 weeks compared with baseline. The mean therapeutic improvement in FEV_1 for all patients is driven by a dramatic increase in FEV_1 in a minority of trial subjects. In a similar study, 38% of patients randomized to inhaled budesonide or fluticasone demonstrated improvements in FEV_1 of under 5% over the course of 24 weeks (Szefler et al., 2002). A unique subset of up to 25% of the non-responders can be classified as having glucocorticoid-resistant asthma (Chan et al., 1998). These patients are non-responsive to even high doses of oral steroids. Furthermore, a careful analysis of these studies indicates that individuals responsive to one class of asthma drug may not necessarily be the individuals responsive to a different class of asthma drug. Despite these studies, there is no universally accepted definition of a steroid (or β_2-agonist) non-responder. These types of data illustrating variable drug efficacy are not limited to asthma drug trials but can be found for almost all classes of drugs. The degree to which this interindividual variability is genetic remains uncertain and probably differs with each class of asthma drug. Moreover, interindividual variability may depend on the particular preparation within a class of asthma drug. Calculations of the repeatability (r) of the treatment response for all three classes of asthma drugs, defined as the fraction of the total population variance that results from within individual differences, show values for r between 60% and 80% indicating the upper limit of the genetic component and suggesting that a substantial fraction of the variance of the treatment response could be genetic (Drazen et al., 2000).

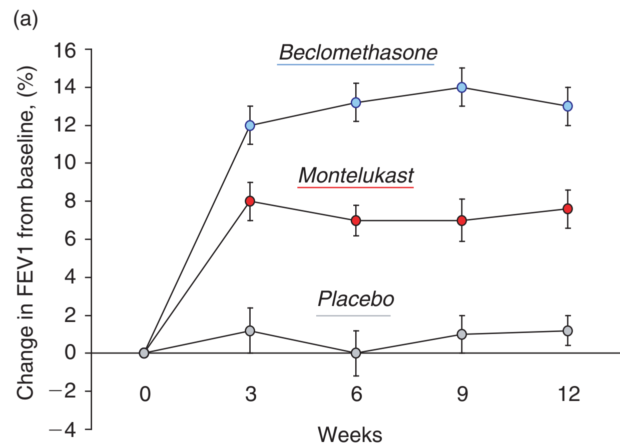


Figure 46.3a Standard approach to clinical trials data showing mean effect by treatment group.

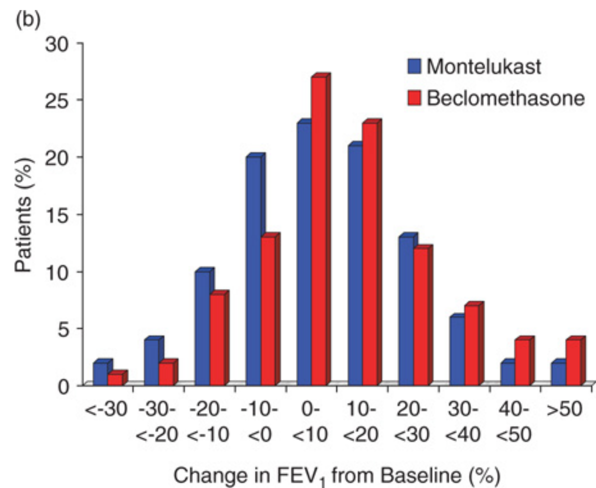


Figure 46.3b Histogram of individual response to beclomethasone subjects in first arm of 3a.

The leukotrienes (LT) are a family of lipoxygenated eicosatetraenoic acids derived from arachidonic acid and produced in the airways of asthmatics that are potent bronchoconstrictors. Of the many enzymes involved in the formation of the leukotrienes, polymorphisms in two, ALOX5 and LTC4S have been associated with altered response to LT inhibitors. ALOX5 catalyzes the conversion of arachidonic acid to LTA4, which is a rate limiting step involved in the synthesis of all leukotrienes (Silverman and Drazen, 1999). In 221 patients with asthma who received either high-dose ABT-761, an ALOX5 inhibitor similar to zileuton, (n=114) or placebo (n=107) treatment, we found that approximately 6% of asthmatic patients had no wild-type allele at the ALOX5 promoter locus and had a diminished response to ABT-761 treatment (Drazen et al., 1999). These findings were consistent with the hypothesis that repeats of the

–GGGCGG- sequence other than the wild type are associated with decreased gene transcription and ALOX5 product production. This has recently been confirmed in cells obtained from patients with asthma (Kalayci et al., 2003, 2006).

Another enzyme of the leukotriene pathway, LTC4S, is responsible for the adduction of glutathione at the C-6 position of the arachidonic acid backbone to form LTC4, a potent bronchoconstrictor. There is a known SNP in the LTC4S promoter, A-444C, with a C allele frequency of 0.19 in normal subjects and of 0.27 in patients with severe asthma (Sayers et al., 2003). The −444C allele creates an activator protein-2 binding sequence that appears to be functional (Sayers et al., 2003), suggesting that the −444C variant is associated with enhanced cysteinyl leukotriene production. Sampson and colleagues found that, among asthmatic subjects treated with zafirlukast (20 mg bid), those homozygous for the A allele ($n=10$) at the −444 locus had a lower FEV_1 response than those with the C/C or C/A ($n=13$) genotype (Sampson et al., 2000). These findings provide possible evidence of a second pharmacogenetic locus in addition to the ALOX5 promoter locus modulating the leukotriene pathway.

Studies evaluating the β_2-agonist pathway have focused largely on the β_2-adrenergic receptor gene (ADRB2). Numerous clinical (Israel et al., 2000, 2001; Reihsaus et al., 1993) and cellular (Liggett, 2000) studies, including one recent prospective, genotype-stratified, clinical trial (Israel et al., 2004), support the association of variation in this gene with response to inhaled β_2-agonist therapy. Other genes in the β_2-agonist pathway are now being actively investigated (Tantisira et al., 2005).

Genetic associations have also been reported for response to corticosteroids in asthma with the following phenotypes: lung function change (Tantisira et al., 2004b), airway constriction (Tantisira et al., 2004a) and relaxation (Tantisira et al., 2005), severe exacerbations (Tantisira et al., 2005), and steroid resistance.

2009 UPDATE

In 2008, substantial additional evidence emerged implicating filaggrin as an important gene in atipic eczema and in asthma associated with this disease. There is, however, no evidence that filaggrin is a risk factor for asthma in the absence of its association with atopic dermatitis (Brown et al., 2008; de Jongh et al., 2008; Henderson et al., 2008; Weidinger et al., 2008a, b).

Confirmation emerged for *ORMDL3* as an asthma locus, although there is controversy as to the importance of its nearest neighbor *GSDML* (Gasmin-like), which is in linkage disequilibrium with *ORMDL3* and may be coexpressed with it (Galanter et al., 2008; Wu et al., 2009). There is likely more to this story, and it will take some time to sort out.

A number of candidate genes have been initially associated with asthma including *IL17RB*, *LKX1*, *IL1RL1*, *IL18R1*, and *IL18RAP*.

Arginase 1 was identified as a gene determining the pharmacogenetic trait of short-acting response to bronchodilator, moving the focus away from *ADRB2* (Litonjua et al., 2008).

2009 UPDATE REFERENCES

Brown, S.J., Sandilands, A., Zhao, Y., Liao, H., Relton, C.L., Meggitt, S.J., Trembath. R.C., Barker, J.N., Reynolds, N.J., Cordell, H.J. et al. (2008). Prevalent and low-frequency null mutations in the filaggrin gene are associated with early-onset and persistent atopic eczema. *J Invest Dermatol* 128, 1591–1594.

de Jongh, C.M., Khrenova, L., Verberk, M.M., Calkoen, F., van Dijk, F.J., Voss, H., John, S.M. and Kezic, S. (2008). Loss-of-function polymorphisms in the filaggrin gene are associated with an increased susceptibility to chronic irritant contact dermatitis: A case-control study. *Br J Dermatol* 159, 621–627.

Galanter, J., Choudhry, S., Eng, C., Nazario, S., Rodriguez-Santana, J.R., Casal, J., Torres-Palacios, A., Salas, J., Chapela, R., Watson, H.G. et al. (2008). ORMDL3 gene is associated with asthma in three ethnically diverse populations. *Am J Respir Crit Care Med* 177, 1194–1200.

Henderson, J., Northstone, K., Lee, S.P., Liao, H., Zhao, Y., Pembrey, M., Mukhopadhyay, S., Smith, G.D., Palmer, C.N., McLean, W.H. and Irvine, A.D. (2008). The burden of disease associated with filaggrin mutations: A population-based, longitudinal birth cohort study. *J Allergy Clin Immunol* 121, 872–877e9.

Litonjua, A.A., Lasky-Su, J., Schneiter, K., Tantisira, K.G., Lazarus, R., Klanderman, B., Lima, J.J., Irvin, C.G., Peters, S.P., Hanrahan, J.P. et al. (2008). ARG1 is a novel bronchodilator response gene: Screening and replication in four asthma cohorts. *Am J Respir Crit Care Med* 178, 688–694.

Weidinger, S., Bauecht, H., Wagenpfeil, S., Henderson, J., Novak, N., Sandilands, A., Chen, H., Rodriguez, E., O'Regan, G.M., Watson, R. et al. (2008a). Analysis of the individuals and aggregate genetic contributions of previously identified serine peptidase inhibitor Kazal type 5 (SPINK5), kallikrein-related peptidase 7 (KLK7), and filaggrin (FLG) polymorphisms to eczema risk. *J Allergy Clin Immunol* 122, 976.

Weidinger, S., O'Sullivan, M., Illig, T., Baurecht, H., Depner, M., Rodriguez, E., Ruether, A., Klopp, N., Vogelberg, C., Weiland, S.K. et al. (2008b). Filaggrin mutations, atopic eczema, hay fever, and asthma in children. *J Allergy Clin Immunol* 121, 1203–1209.

Wu, H., Romieu, I., Sienra-Monge, J.J., Li, H., Del Rio-Navarro, B.E. and London, S.J. (2009). Genetic variation in ORM1-like 3 (ORMDL3) and gasdermin-like (GSDML) and childhood asthma. *Allergy* 64(4), 629–635.

ACKNOWLEDGEMENTS

The authors wish to acknowledge their colleagues in the Channing Laboratory Respiratory Genetics Group, and the Center for Genomic Medicine at Brigham and Women's Hospital.

REFERENCES

Allen, A.S., Rathouz, P.J. and Satten, G.A. (2003). Informative missingness in genetic association studies: Case-parent designs. *Am J Hum Genet* 72, 671–680.

Ansari, A.A., Freidhoff, L.R., Meyers, D.A., Bias, W.B. and Marsh, D.G. (1989). Human immune responsivesness to *Lolium perenne* pollen allergen Lol p III (rye III) is associated with HLA-DR3 and DR5. *Hum Immunol* 25, 59–71.

Aron, Y., Desmazes-Dufeu, N., Matran, R., Polla, B.S., Dusser, D., Lockhart, A. and Swierczewski, E. (1996). Evidence of a strong positive association between atopy and the HLA class II alleles DR4 and DR7. *Clin Exp Aller* 26, 821–826.

Baldini, M., Lohman, I.C., Halonen, M., Erickson, R.P., Holt, P.G. and Martinez, F.D. (1999). A polymorphism in the 5′ flanking region of the CD14 gene is associated with circulating soluble CD14 levels and with total serum immunoglobulin E. *Am J Respir Cell Mol Biol* 20, 976–983.

Bleecker, E.R., Postma, D.S., Howard, T.D., Koppelman, G.H., Meijer, G.G., Xu, J., Stine, O.C. and Meyers, D.A. (1999). Genome screen for susceptibility loci in a genetically homogeneous Dutch population. *Am J Respir Crit Care Med* 159, A645. Abstract

Blumenthal, M.N., Ober, C., Beaty, T.H., Bleecker, E.R., Langefeld, C.D., King, R.A., Lester, L., Cox, N., Barnes, K., Togias, A. et al. (2004). Genome scan for locu linked to mite sensitivity: The Collaborative Study on the Genetics of Asthma (CSGA). *Genes Immun* 5, 226–231.

Brasch-Andersen, C., Tan, Q., Borglum, A.D., Haagerup, A., Larsen, T.R., Vestbo, J. and Kruse, T.A. (2006). Significant linkage to chromosome 12q24.32-q24.33 and identification of SFRS8 as a possible asthma susceptibility gene. *Thorax* 61, 874–879.

Brutsche, M.H., Joos, L., Carlen Brutsche, I.E., Bissinger, R., Tamm, M., Custovic, A. and Woodcock, A. (2002). Array-based diagnostic gene-expression score for atopy and asthma. *J Aller Clin Immunol* 109, 271–273.

Castro, J., Telleria, J.J., Linares, P. and Blanco-Quiros, A. (2000). Increased TNFA2, but not TNFB1, allele frequency in Spanish atopic patients. *J Investig Allergol Clin Immunol* 10, 149–154.

Celedon, J.C., Soto-Quiros, M.E., Avila, L., Lake, S., Liang, C., Fournier, E., Spesny, N., Hersh, C.P., Sylvia, J.S., Hudson, T.J. et al. (2007). Significant linkage to airway responsiveness on chromosome 12q24 in families of children with asthma in Costa Rica. *Hum Genet* 120, 691–699.

Centers for Disease Control. (1998). Surveillance for Asthma – United States, 1960–1995. Morbidity Mortality Weekly Rep, 47, 1–27.

Chan, M.T., Leung, D.Y., Szefler, S.J. and Spahn, J.D. (1998). Difficult-to-control asthma: Clinical characteristics of steroid-insensitive asthma. *J Aller Clin Immunol* 101, 594–601.

Chew, G., Burge, H., Dockery, D., Muilenberg, M., Weiss, S. and Gold, D. (1998). Limitations of a home characteristics questionnaire as a predictor of indoor allergen levels. *Am J Respir Crit Care Med* 157, 1536–1541.

The Collaborative Study on the Genetics of Asthma (CSGA) (1997). A genome-wide search for asthma susceptibility loci in ethnically diverse populations. *Nat Genet* 15, 389–392.

Cox, H.E., Moffatt, M.F., Faux, J.A., Walley, A.J., Coleman, R., Trembath, R.C., Cookson, W.O. and Harper, J.I. (1998). Association of atopic dermatitis to the beta subunit of the high affinity immunoglobulin E receptor. *Br J Dermatol* 138, 182–187.

Daniels, S.E., Bhattacharrya, S., James, A., Leaves, N.I., Young, A., Hill, M.R., Faux, J.A., Ryan, G.F., LeSouef, P.N., Lathrop, G.M. et al. (1996). A genome-wide search for quantitative trait loci underlying asthma. *Nature* 383, 247–250.

DeMeo, D.L., Lange, C., Silverman, E.K., Senter, J.M., Drazen, J.M., Barth, M.J., Laird, N. and Weiss, S.T. (2002). Univariate and multivariate family-based association analysis of the IL-13 ARG130GLN polymorphism in the Childhood Asthma Management Program. *Genet Epidemiol* 23, 335–348.

Di Somma, C., Charron, D., Deichmann, K., Buono, C. and Ruffilli, A. (2003). Atopic asthma and TNF-308 alleles: Linkage disequilibrium and association analyses. *Hum Immunol* 64, 359–365.

Dizier, M.H., Besse-Schmittler, C., Guilloud-Bataille, M., Annesi-Maesano, I., Boussaha, M., Bousquet, J., Charpin, D., Degioanni, A., Gormand, F., Grimfeld, A. et al. (2000). Genome screen for asthma and related phenotypes in the French EGEA study. *Am J Respir Crit Care Med* 162, 1812–1818.

Dizier, M.H., Bouzigon, E., Guilloud-Bataille, M., Betard, C., Bousquet, J., Charpin, D., Gormand, F., Hochez, J., Just, J., Lemainque, A. et al. (2005). Genome screen in the French EGEA study: Detection of linked regions shares or not shared by allergic rhinitis and asthma. *Genes Immun* 6, 95–102.

Drazen, J.M., Yandava, C.N., Dube, L., Szczerback, N., Hippensteel, R., Pillari, A., Israel, E., Schork, N., Silverman, E.S., Katz, D.A. et al. (1999). Pharmacogenetic association between ALOX5 promoter genotype and the response to anti-asthma treatment. *Nat Genet* 22, 168–170.

Drazen, J.M., Silverman, E.K. and Lee, T.H. (2000). Heterogeneity of therapeutic responses in asthma. *Br Med Bull* 56, 1054–1070.

Duffy, D.L., Martin, N.G., Battistutta, D., Hopper, J.L. and Mathews, J.D. (1990). Genetics of asthma and hay fever in Australian twins. *Am Rev Respir Dis* 142, 1351–1358.

Edfors-Lubs, M.L. (1971). Allergy in 7000 twin pairs. *Acta Allergol* 26, 249–285.

Elias, J.A., Lee, C.G., Zheng, T., Ma, B., Homer, R.J. and Zhu, Z. (2003). New insights into the pathogenesis of asthma. *J Clin Invest* 111, 291–297.

Evans, D.M., Zhu, G., Duffy, D.L., Montgomery, G.W., Frazer, I.H. and Martin, N.G. (2004). Major quantitative trait locus for eosinophil count is located on chromosome 2q. *J Aller Clin Immunol* 114, 826–830.

Ferreira, M.A., O'Gorman, L., LeSouef, P., Burton, P.R., Toelle, B.G., Robertson, C.F., Visscher, P.M., Martin, N.G. and Duffy, D.L.

(2005). Robust estimation of experimentwise P values applied to a genome scan of multiple asthma traits identifies a new region of significant linkage on chromosome 20q13. *Am J Hum Genet* 77, 1075–1085.

Graves, P.E., Kabesch, M., Halonen, M., Holberg, C.J., Baldini, M., Fritzsch, C., Weiland, S.K., Erickson, R.P., von Mutius, E. and Martinez, F.D. (2000). A cluster of seven tightly linked polymorphisms in the IL-13 gene is associated with total serum IgE levels in three populations of white children. *J Aller Clin Immunol* 105, 506–513.

Green, S.L., Gaillard, M.C., Song, E., Dewar, J.B. and Halkas, A. (1998). Polymorphisms of the beta chain of the high-affinity immunoglobulin E receptor (Fc epsilon Rl-beta) in South African black and white asthmatic and nonasthmatic individuals. *Am J Respir Crit Care Med* 158, 1487–1492.

Grunig, G., Warnock, M., Wakil, A.E., Venkayya, R., Brombacker, F., Rennick, D.M., Sheppard, D., Mohrs, M., Donaldson, D.D., Locksley, R.M. et al. (1998). Requirement for IL-13 independently of IL-14 in experimental asthma (see comments). *Science* 282, 2261–2263.

Guajardo, J.R., Schleifer, K.W., Daines, M.O., Ruddy, R.M., Aronow, B.J., Wills-Karp, M. and Hershey, G.K. (2005). Altered gene expression profiles in nasal respiratory epithelium reflect stable versus acute childhood asthma. *J Aller Clin Immunol* 115, 243–251.

Guerra, S., Lohman, I.C., Halonen, M., Martinez, F.D. and Wright, A.L. (2004). Reduced interferon gamma production and soluble CD14 levels in early life predict recurrent wheezing by 1 year of age. *Am J Respir Crit Care Med* 169, 70–76.

Haagerup, A., Bjerke, T., Schiotz, P.O., Binderup, H.G., Dahl, R. and Kruse, T.A. (2002). Asthma and atopy – a total genome scan for susceptibility genes. *Allergy* 57, 680–686.

Hakonarson, H., Bjornsdottir, U.S., Halapi, E., Palsson, S., Adalsteinsdottir, E., Gislason, D., Finnbogason, G., Gislason, T., Kristjansson, K., Arnason, T. et al. (2002). A major susceptibility gene for asthma maps to chromosome 14q24. *Am J Hum Genet* 71, 483–491.

Hansel, N.N., Hilmer, S.C., Georas, S.N., Cope, L.M., Guo, J., Irizarry, R.A. and DIette, G.B. (2005). Oligonucleotide-microarray analysis of peripheral-blood lymphocytes in severe asthma. *J Lab Clin Med* 145, 263–274.

Hasegawa, K., Tamari, M., Shao, C., Shimizu, M., Takahashi, N., Mao, X.Q., Yamasaki, A., Kamada, F., Doi, S. and Fujiwara, H. (2004). Variations in the C3, C3a receptor, and C5 genes affect susceptibility to bronchial asthma. *Hum Genet*.

Heinzmann, A., Dietrich, H., Jerkic, S.P., Kurz, T. and Deichmann, K.A. (2003). Promoter polymorphisms of the CD14 gene are not associated with bronchial asthma in Caucasian children. *Eur J Immunogenet* 30, 345–348.

Hersh, C., Raby, B., Soto-Quiros, M.E., Murphy, A., Avila, L., Lasky-Su, J., Sylvia, J.S., Klanderman, B.J., Lange, C., Weiss, S.T. et al. (2007). Comprehensive testing of positionally cloned asthma genes in two populations. *Am J Respir Crit Care Med* Aug 16. Epub ahead of print

Hersh, C.P., Soto-Quiros, M.E., Avila, L., Lake, S., Liang, C., Fournier, E., Spesny, M., Sylvia, J.S., Lazarus, R., Hudson, T. et al. (2007). Genome-wide linkage analysis of pulmonary function in families of children with asthma in Costa Rica. *Thorax* 62, 224–230.

Hershey, G.K., Friedrich, M.L., Esswein, L.A., Thomas, M.L. and Chatila, T.A. (1997). The association of atopy with a gain-of-function mutation in the alpha subunit of the interleukin-4 receptor. *N Engl J Med* 337, 1720–1725.

Hopp, R.J., Bewtra, A.K., Nair, N.M. and Townley, R.G. (1984). Specificity and sensitivity of methacholine inhalation challenge in normal and asthmatic children. *J Aller Clin Immunol* 74, 154–158.

Howard, T.D., Whittaker, P.A., Zaiman, A.L., Koppelman, G.H., Xu, J., Hanley, M.T., Meyers, D.A., Postma, D.S. and Bleecker, E.R. (2001). Identification and association of polymorphisms in the interleukin-13 gene with asthma and atopy in a Dutch population. *Am J Respir Cell Mol Biol* 25, 377–384.

Howard, T.D., Postma, D.S., Hawkins, G.A., Koppelman, G.H., Zheng, S.L., Wysong, A.K., Xu, J., Meyers, D.A. and Bleecker, E.R. (2002). Fine mapping of an IgE-controlling gene on chromosome 2q: Analysis of CTLA4 and CD28. *J Aller Clin Immunol* 110, 743–751.

Huang, S.K., Mathias, R.A., Ehrlich, E., Plunkett, B., Liu, X., Cutting, G.R., Wang, X.J., Li, X.D., Togias, A., Barnes, K.C. et al. (2003). Comparative Study on the Genetics of Asthma. Evidence for asthma susceptibility genes on chromosome 11 in an African-American population. *Hum Genet* 113, 71–75.

Hunninghake, G.M., Soto-Quiros, M.E., Avila, L., Su, J., Murphy, A., DeMeo, D.L., Ly, N.P., Liang, C., Sylvia, J.S., Klanderman, B.J. et al. (2007). Polymorphisms in IL13, total IgE, eosinophilia, and asthma exacerbations in childhood. *J Aller Clin Immunol* 120, 84–90.

Israel, E., Drazen, J.M., Liggett, S.B., Boushey, H.A., Cherniack, R.M., Chinchilli, V.M., Cooper, D.M., Fahy, J.V., Fish, J.E., Ford, J.G. et al. (2000a). The effect of polymorphisms of the beta(2)-adrenergic receptor on the response to regular use of albuterol in asthma. *Am J Respir Crit Care Med* 162, 75–80.

Israel, E., Drazen, J.M., Liggett, S.B., Boushey, H.A., Cherniack, R.M., Chinchilli, V.M., Cooper, D.M., Fahy, J.V., Fish, J.E., Ford, J.G. et al. (2001b). Effect of polymorphism of the beta(2)-adrenergic receptor on response to regular use of albuterol in asthma. *Int Arch Aller Immunol* 124, 183–186.

Israel, E., Chinchilli, V.M., Ford, J.G., Boushey, H.A., Cherniack, R.M., Craig, T.J., Deykin, A., Fagan, J.K. and Fish, J.E. (2004). Genotype stratified prospective cross-over trial of regularly scheduled albuterol treatment in asthma. *Lancet*.

Jenkins, M.A., Hopper, J.L. and Giles, G.G. (1997). Regressive logistic modeling of familial aggregation for asthma in 7,394 population-based nuclear families. *Genet Epidemiol* 14, 317–332.

Kalayci, O., Wechsler, M., Galper, B., Hong, C., Israel, E., Drazen, J.M., Lilly, C.M. (2003) LTC4 production by eosinophils in asthmatic subjects with alternative forms of ALOX-5 core promoter. 11–14.

Kalayci, O., Birben, E., Sackesen, C., Keskin, O., Tahan, F., Wechsler, M.E., Civelek, E., Soyer, O.U., Adalioglu, G., Tuncer, A. et al. (2006). ALOX5 promoter genotype, asthma severity and LTC production by eosinophils. *Allergy* 61, 97–103.

Karp, C.L., Grupe, A., Schadt, E., Ewart, S.L., Keane-Moore, M., Cuomo, P.J., Kohl, J., Wahl, L., Kuperman, D., Germer, S. et al. (2000). Identification of complement factor 5 as a susceptibility locus for experimental allergic asthma. *Nat Immunol* 1, 221–226.

Koppelman, G.H., Stine, O.C., Xu, J., Howard, T.D., Zheng, S.L., Kauffman, H.F., Bleecker, E.R., Meyers, D.A. and Postma, D.S. (2002). Genome-wide search for atopy susceptibility genes in Dutch families with asthma. *J Aller Clin Immunol* 109, 498–506.

Krug, N., Tschernig, T., Erpenbeck, V.J., Hohlfeld, J.M. and Kohl, J. (2001). Complement factors C3a and C5a are increased in bronchoalveolar lavage fluid after segmental allergen provocation in subjects with asthma. *Am J Respir Crit Care Med* 164, 1841–1843.

Laitinen, T., Daly, M.J., Rioux, J.D., Kauppi, P., Laprise, C., Petays, T., Green, T., Cargill, M., Haahtela, T., Lander, E.S. et al. (2001).

A susceptibility locus for asthma-related traits on chromosome 7 revealed by genome-wide scan in a founder population. *Nat Genet* 28, 87–91.

Laitinen, T., Polvi, A., Rydman, P., Vendelin, J., Pulkkinen, V., Salmikangas, P., Makela, S., Rehn, M., Priskanen, A., Rautanen, A. et al. (2004). Characterization of a common susceptibility locus for asthma-related traits. *Science* 304, 300–304.

Laprise, C., Sladek, R., Ponton, A., Bernier, M.C., Hudson, T.J. and Laviolette, M. (2004). Functional classes of bronchial mucosa genes that are differentially expressed in asthma. *BMC Genomics* 5, 21.

Li, A. and Hopkin, J.M. (1997). Atopy phenotype in subjects with variants of the b subunit of the high affinity IgE receptor. *Thorax* 52, 654–655.

Li, H., Romieu, I., Sienra-Monge, J.J., Ramirez-Aguilar, M., Estela Del Rio-Navarro, B., Kistner, E.O., Gjessing, H.K., Lara-Sanchez Idel, C., Chiu, G.Y. and London, S.J. (2006). Genetic polymorphisms in arginase I and II and childhood asthma and atopy. *J Aller Clin Immunol* 117, 119–126.

Liggett, S.B. (2000). Pharmacogenetics of beta-1- and beta-2-adrenergic receptors. *Pharmacology* 61, 167–173.

Liggett, S.B. (2001). Pharmacogenetic applications of the human genome project. *Nat Med* 7, 281–283.

Lilly, C.M., Tateno, H., Oguma, T., Israel, E. and Sonna, L.A. (2005). Effects of allergen challenge on airway epithelial cell gene expression. *Am J Respir Crit Care Med* 171, 579–586.

Lubs, M.L. (1972). Empiric risks for genetic counseling in families with allergy. *J Pediatr* 80, 26–31.

Malmstrom, K., Rodriguez-Gomez, G., Guerra, J., Villaran, C., Pineiro, A., Wei, L.X., Seidenberg, B.C. and Reiss, T.F. (1999). Oral Montelukast, inhaled beclomethasone, and placebo for chronic asthma. A randomized, controlled trial. Montelukast/Beclomethasone Study Group. *Ann Intern Med* 130, 487–495.

Meyers, D.A., Postma, D.S., Stine, O.C., Koppelman, G.H., Ampleford, E.J., Jongepier, H., Howard, T.D. and Bleecker, E.R. (2005). Genome screen for asthma and bronchial hyperresponsiveness: Interactions with passive smoke exposure. *J Aller Clin Immunol* 115, 1169–1175.

Mitsuyasu, H., Izuhara, K., Mao, X.-Q., Gao, P.S., Arinobu, Y., Enomoto, T., Kawai, M., Sasaki, S., Dake, Y., Hamasaki, N. et al. (1998). Ile50Val variant of IL4R alpha upregulates IgE synthesis and associates with atopic asthma. *Nat Genet* 19, 119–120.

Moffatt, M.F. and Cookson, W.O. (1997). Tumour necrosis factor haplotypes and asthma. *Hum Mol Genet* 6, 551–554.

Moffatt, M.F., Kabesch, M., Liang, L., Dixon, A.L., Strachan, D., Heath, S., Depner, M., von Berg, A., Bufe, A., Rietschel, E., Heinzmann, A., Simma, B., Frischer, T., Willis-Owen, S.A., Wong, K.C., Illig, T., Vogelberg, C., Weiland, S.K., von Mutius, E., Abecasis, G.R., Farrall, M., Gut, I.G., Lathrop, G.M. and Cookson, W.O. (2007). Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 448, 470–473.

Moffatt, M.F., Schou, C., Faux, J.A., Abecasis, G.R., Musk, J.A. and Cookson, W.O. (2001). Association between quantitative traits underlying asthma and the HLA-DRB1 locus in a family-based population sample. *Eur J Hum Genet* 9, 341–346.

Munitz, A., Bachelet, I., Finkelman, F.D., Rothenberg, M.E. and Levi-Schaffer, F. (2007). CD48 is critically involved in allergic eosinophilic airway inflammation. *Am J Respir Crit Care Med* 175, 911–918.

Nicolae, D., Cox, N.J., Lester, L.A., Schneider, D., Tan, Z., Billstrand, C., Kuldanek, S., Donfack, J., Kogut, P., Patel, N.M. et al. (2005). Fine mapping and positional candidate studies identify HLA-G as an asthma susceptibility gene on chromosome 6p21. *Am J Hum Genet* 76, 349–357.

Nieminen, M.M., Kaprio, J. and Koskenvuo, M. (1991). A population-based study of bronchial asthma in twin pairs. *Chest* 100, 71–75.

Novak, N., Kruse, S., Kraft, S., Geiger, E., Kluken, H., FImmers, R., Deichmann, K.A. and Bieber, T. (2002). Dichotomic nature of atopic dermatitis reflected by combined analysis of monocyte immunophenotyping and single nucleotide polymorphisms of the interleukin-4/interleukin-13 receptor gene: The dichotomy of extrinsic and intrinsic atopic dermatitis. *J Invest Dermatol* 119, 870–875.

O'Donnell, A.R., Toelle, B.G., Marks, G.B., Hayden, C.M., Laing, I.A., Peat, J.K., Goldblatt, J. and Le Souef, P.N. (2004). Age-specific relationship between CD14 and atopy in a cohort assessed from age 8 to 25 years. *Am J Respir Crit Care Med* 169, 615–622.

Ober, C. and Hoffjan, S. (2006). Asthma genetics 2006: The long and winding road to gene discovery. *Genes Immun*, 1–6.

Ober, C., Cox, N.J., Abney, M., DiRienzi, A., Lander, E.S., Changyaleket, B., Gidley, H., Kurtz, B., Lee, J., Nance, M. et al. (1998). Genome-wide search for asthma susceptibility loci in a founder population. The Collaborative Study on the Genetics of Asthma. *Hum Mol Genet* 7, 1393–1398.

Ober, C., Leavitt, S.A., Tsalenko, A., Howard, T.D., Hoki, D.M., Daniel, R., Newman, D.L., Wu, X., Parry, R., Lester, L.A. et al. (2000). Variation in the interleukin 4-receptor a gene confers susceptibility to asthma and atopy in ethnically diverse populations. *Am J Hum Genet* 66, 517–526.

Ober, C., Tsalenko, A., Parry, R. and Cox, N.J. (2000). A second-generation genomewide screen for asthma-susceptibility alleles in a founder population. *Am J Hum Genet* 67, 1154–1162.

Palmer, L.J. and Cookson, W.O. (2000). Genomic approaches to understanding asthma. *Genome Res* 10, 1280–1287.

Pillai, S.G., Chiano, M.N., White, N.J., Speer, M., Barnes, K.C., Carlsen, K., Gerritsen, J., Helms, P., Lenney, W., Silverman, M. et al. (2006). A genome-wide search for linkage to asthma phenotypes in the genetics of asthma international network families: Evidence for a major susceptibility locus on chromosome 2p. *Eur J Hum Genet* 14, 307–316.

Raby, B.A., Soto-Quiros, M.E., Avila, L., Lake, S.L., Murphy, A., Liang, C., Fournier, E., Spesny, M., Sylvia, J.S., Verner, A. et al. (2007). Sex-specific linkage to total serum immunoglobulin E in families of children with asthma in Costa Rica. *Hum Mol Genet* 16, 243–253.

Reihsaus, E., Innis, M., MacIntyre, N. and Liggett, S.B. (1993). Mutations in the gene encoding for the beta 2-adrenergic receptor in normal and asthmatic subjects. *Am J Respir Cell Mol Biol* 8, 334–339.

Rosa-Rosa, L., Zimmermann, N., Bernstein, J.A., Rothenberg, M.E. and Khurana Hershey, G.K. (1999). The R576 IL-4 receptor alpha allele correlates with asthma severity. *J Aller Clin Immunol* 104, 1008–1014.

Rosenstreich, D., Eggleston, P., Kattan, M., Baker, D., Slavin, R., Gergen, P.J., Mitchel, H., McNiff-Mortimer, K., Henry, L., Ownby, D. et al. (1997). The role of cockroach allergy and exposure to cockroach allergen in causing morbidity among inner-city children with asthma. *N Engl J Med* 336, 1356–1363.

Rosenwasser, L.J., Klemm, D.J., Dresback, J.K., Inamura, H., Mascali, J.J., Klinnert, M. and Borish, L. (1995). Promoter polymorphisms in the chromosome 5 gene cluster in asthma and atopy. *Clin Exp Allergy* 25(Suppl 2), 74–78. discussion 95–6.

Sampson, A.P., Siddiqui, S., Buchanan, D., Howarth, P.H., Holgate, S.T., Holloway, J.W. and Sayers, I. (2000). Variant LTC(4) synthase allele modifies cysteinyl leukotriene synthesis in eosinophils and predicts clinical response to zafirlukast. *Thorax* 55(Suppl 2), S28–S31.

Sandford, A., Weir, T. and Pare, P. (1996). The genetics of asthma. *Am J Respir Crit Care Med* 153, 1749–1765.

Sandford, A.J., Chagani, T., Zhu, S., Weir, T.D., Bai, T.R., Spinelli, J.J., Fitzgerald, J.M., Behbehani, N.A., Tan, W.C. and Pare, P.D. (2000). Polymorphisms in the IL4, IL4RA, and FCERIB genes and asthma severity. *J Aller Clin Immunol* 106, 135–140.

Sayers, I., Barton, S., Rorke, S., Beghe, B., Hayward, B., Van Eerdewegh, P., Keith, T., Clough, J.B., Ye, S., Holloway, J.W. et al. (2003). Allelic association and functional studies of promoter polymorphisms in the leukotriene C4 synthase gene (LTC4S) in asthma. *Thorax* 58, 417–424.

Sears, M., Burrows, B., Flannery, E., Herbison, G., Hewitt, C. and Holdaway, M. (1991). Relation between airway responsiveness and serum IgE in children with asthma and in apparently normal children. *N Engl J Med* 325, 1067–1071.

Shin, H.D., Park, B.L., Kim, L.H., Jung, J.H., Wang, J.H., Kim, Y.J., Park, H.S., Hong, S.J., Choi, B.W., Kim, D.J. et al. (2004). Association of tumor necrosis factor polymorphisms with asthma and serum total IgE. *Hum Mol Genet* 13, 397–403.

Shirakawa, T., Li, A., Dubowitz, M., Dekker, J.W., Shaw, A.E., Faux, J.A., Ra, C., Cookson, W.O. and Hopkin, J.M. (1994). Association between atopy and variants of the beta subunit of the high-affinity immunoglobulin E receptor (see comments). *Nat Genet* 7, 125–129.

Sibbald, B., Horn, M.E., Brain, E.A. and Gregg, I. (1980). Genetic factors in childhood asthma. *Thorax* 35, 671–674.

Silverman, E.S. and Drazen, J.M. (1999). The biology of 5-lipoxygenase: Function, structure, and regulatory mechanisms. *Proc Assoc Am Physicians* 111, 525–536.

Silverman, E.S., Hjoberg, J., Palmer, L.J., Tantisira, K.G., Weiss, S.T. and Drazen, J.M. (2002). Application of pharmacogenetics to the therapeutics of asthma. In *Therapeutic Targets of Airway Inflammation* (D. Huston, ed.), Marcel Dekker, New York.

Sporik, R., Holgate, S., Platts-Mills, T. and Cogswell, J. (1990). Exposure to house-dust mite allergen and the development of asthma in childhood. *N Engl J Med* 323, 502–507.

Statement of the American Thoracic Society (1987). Standardization of spirometry – 1987 update. *Am Rev Respir Dis* 136, 1285–1298.

Sullivan, S.D. and Strassels, S.A. (2002). Health economics. In *Asthma and COPD. Basic Mechanisms and Clinical Management* (N.C. Thomson, ed.), Academic Press, San Diego, pp. 657–671.

Szefler, S.J., Martin, R.J., King, T.S., Boushey, H.A., CHerniack, R.M., Chinchilli, V.M., Craig, T.J., Dolovich, M., Drazen, J.M., Fagan, J.K. et al. (2002). Significant variability in response to inhaled corticosteroids for persistent asthma. *J Aller Clin Immuno* 109, 410–418.

Tantisira, K., Small, K., Litonjua, A., Weiss, S. and LIggett, S. (2005). Molecular properties and pharmacogenetics of a polymorphism of adenylyl cyclase type 9 in asthma: Interaction between beta-agonist and corticosteroid pathways. *Hum Mol Genet* 14, 1671–1677.

Tantisira, K.G., Hwang, E., Raby, B., Silverman, E.S., Lake, S.L., Richter, B.G., Peng, S.L., Drazen, J.M., GLimcher, L.H. and Weiss, S.T. (2004). TBX21: A functional variant predicts improvement in asthma with the use of inhaled corticosteroids. *Proc Natl Acad Sci USA* 101, 18099–18104.

Tantisira, K.G., Lake, S., Silverman, E.S., Palmer, L.J., Lazarus, R., Silverman, E.K., Liggett, S.B., Gelfand, E.W., Rosenwasser, L.J., Richter, B. et al. (2004). Corticosteroid pharmacogenetics: Association of sequence variants in CRHR1 with improved lung function in asthmatics treated with inhaled corticosteroids. *Hum Mol Genet* 13, 1353–1359.

Tremblay, K., Lemire, M., Provost, V., Pastinen, T., Renaud, Y., Sandford, A.J., Laviolette, M., Hudson, T.J. and Laprise, C. (2006). Association study between the CX3CR1 gene and asthma. *Genes Immun* 7, 632–639.

Van Arsdel, P.P. and Motulsky, A.G. (1959). Frequency and heritability of asthma and allergic rhinitis in college students. *Acta Genet Med Gemellol* 9, 101–114.

van der Pouw Kraan, G.C.T.M, van Veen, A., Boeije, L.C.M, van Tuyl, S.A.P, de Groot, E.R., Stapel, S.O., Bakker, A., Verweij, C.L., Aarden, L.A. and van der Zee, J.S. (1999). An IL-13 promoter polymorphisms associated with increased risk of allergic asthma. *Genes Immunity* 1, 61–65.

Van Eerdewegh, P., Little, R.D., Dupuis, J., Del Mastro, R.G., Falls, K., Simon, J., Torrey, D., Pandit, S., McKenny, J., Braunschweiger, K. et al. (2002). Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. *Nature* 418, 426–430.

Wang, T.N., Chen, W.Y., Wang, T.H., Chen, C.J., Huang, L.Y. and Ko, Y.C. (2004). Gene–gene synergistic effect on atopic asthma: tumour necrosis factor-alpha-308 and lymphotoxin-alpha-Ncol in Taiwan's children. *Clin Exp Allergy* 34, 184–188.

Weiss, K.B., Gergen, P.J. and Wagener, D.K. (1993). Breathing better or wheezing worse? The changing epidemiology of asthma morbidity and mortality. *Annu Rev Public Health* 14, 491–513.

Wjst, M., Fischer, G., Immervoll, T., Jung, M., Saar, K., Rueschendorf, F., Reis, A., Ulbrecht, M., Gomolka, M., Weiss, E.H. et al. (1999). A genome-wide search for linkage to asthma. German Asthma Genetics Group. *Genomics* 58, 1–8.

Woo, J.G., Assa'ad, A., Heizer, A.B., Bernstein, J.A. and Hershey, G.K. (2003). The -159C T polymorphism of CD14 is associated with nonatopic asthma and food allergy. *J Aller Clin Immunol* 112, 438–444.

Xu, J., Postma, D.S., Howard, T.D., Koppelman, G.H., Zheng, S.L., Stine, O.C., Bleecker, E.R. and Meyers, D.A. (2000). Major genes regulating total serum immunoglobulin E levels in families with asthma. *Am J Hum Genet* 67, 1163–1173.

Xu, J., Meyers, D.A., Ober, C., Blumenthal, M.N., Mellen, B., Barnes, K.C., King, R.A., Lester, L.A., Howard, T.D., Solway, J. et al. (2001). Collaborative Study on the Genetics of Asthma. Genomewide screen and identification of gene-gene interactions for asthma-susceptibility loci in three US populations: Collaborative study on the genetics of asthma. *Am J Hum Genet* 68, 1437–1446.

Xu, X., Fang, Z., Wang, B., Chen, C., Guang, W., Jin, Y., Yang, J., Lewitzky, S., Aelony, A., Parker, A. et al. (2001). A genome-wide search for quantitative-trait loci underlying asthma. *Am J Hum Genet* 69, 1271–1277.

Yokouchi, Y., Nukaga, Y., Shibasaki, M., Noguchi, E., Kimura, K., Ito, S., Nishihara, M., Yamakawa-Kobayashi, K., Takeda, K., Imoto, N. et al. (2000). Significant evidence for linkage of mite-sensitive childhood asthma to chromosome 5q31-q33 near the interleukin 12B locus by a genome-wide search in Japanese families. *Genomics* 66, 152–160.

Youssef, L.A., Schuyler, M., Gilmartin, L., Pickett, G., Bard, J.D., Tarleton, C.A., Archibeque, T., Qualls, C., Wilson, B.S. and Oliver, J.M. (2007). Histamine release from the basophils of control and asthmatic subjects and a comparison of gene expressions between "releaser" and "nonreleaser" basophils. *J Immunol* 178, 4584–4594.

Zhang, J., Rowe, W.L., Clark, A.G. and Buetow, K.H. (2003). Genomewide distribution of high-frequency, completely mismatching SNP haplotype pairs observed to be common across human populations. *Am J Hum Genet* 73, 1073–1081.

Zhu, S., Chan-Yeung, M., Becker, A.B., Dimich-Ward, H., Ferguson, A.C., Manfreda, J., Watson, W.T., Pare, P.D. and Sandford, A.J. (2000). Polymorphisms of the IL-4, TNF-alpha, and Fcepsilon Rlbeta genes and the risk of allergic disorders in at-risk infants. *Am J Respir Crit Care Med* 161, 1655–1659.

Zimmermann, N., King, N.E., Laporte, J., Yang, M., Mishra, A., Pope, S.M., Muntel, E.E., Witte, D.P., Pegg, A.A., Foster, P.S. et al. (2003). Dissection of experimental asthma with DNA microarray analysis identifies arginase in asthma pathogenesis. *J Clin Invest* 111, 1863–1874.

Zou, J., Young, S., Zhu, F., Gheyas, F., SKeans, S., Wan, Y., Wang, L., Ding, W., BIllah, M., McClanahan, T., Coffman, R.L. et al. (2002). Microarray profile of differentially expressed genes in a monkey model of allergic asthma. *Genome Biol* 3, research0020.

Zou, J., Young, S., Zhu, F., Xia, L., Skeans, S., Wan, Y., Wang, L., McClanahan, T., Gheyas, F., Wei, D. et al. (2002). Identification of differentially expressed genes in a monkey model of allergic asthma by microarray technology. *Chest* 121, 26S–27S.

RECOMMENDED RESOURCE

Respiratory Genetics. (2005). Silverman, E.K., Shapiro, S.D., Lomas, D.A., Weiss, S.T. (eds.) Hodder Arnold. Comprehensive textbook of respiratory genetics.

CHAPTER

Genomics in the Evaluation and Management of Chronic Obstructive Pulmonary Disease

Peter J. Barnes

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is characterized by progressive development of airflow limitation that is not fully reversible (Barnes, 2000a). The term COPD encompasses chronic obstructive bronchiolitis with obstruction of small airways and emphysema with enlargement of airspaces and destruction of lung parenchyma, loss of lung elasticity and closure of small airways (Hogg, 2004). Chronic bronchitis, by contrast, is defined by a productive cough of more than 3 months duration for more than 2 successive years; this reflects mucous hypersecretion and is not necessarily associated with airflow limitation.

COPD is common and is increasing globally. It is now the fourth leading cause of death in the United States and the only common cause of death that is increasing. This is likely to be an underestimate as COPD is likely to be contributory to other common causes of death. It is predicted to become the third most common cause of death and the fifth most common cause of chronic disability worldwide in the next few years (Lopez et al., 2006). It currently affects more than 5% of the adult population and is underdiagnosed in the community (Chapman et al., 2006). COPD consumes an increasing proportion of health care resources, which currently exceed those devoted to asthma by more than threefold.

There is persuasive evidence that the susceptibility to develop COPD is genetically determined, although the genes have not yet been identified. Genomics and proteomics are currently in use to understand the abnormal protein expression in this disease as a way of understanding its complex pathophysiology. There are clearly different phenotypes of COPD that need to be better defined. Finally pharmacogenomics may have an impact on COPD therapy in the future. This chapter gives an overview of COPD and highlights where genomic medicine is relevant to the future understanding and management of this common disease.

PREDISPOSITION

Several environmental and endogenous factors, including genes, increase the risk of developing COPD (Table 47.1).

Environmental Factors

In industrialized countries, cigarette smoking accounts for most cases of COPD, but in developing countries other environmental pollutants, such as particulates associated with cooking in confined spaces, are important causes. It is likely that there are important interactions between environmental factors and a genetic predisposition to develop the disease. Air pollution

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

TABLE 47.1 Risk factors for COPD

Environmental factors	Endogenous (host) factors
Cigarette smoking	α_1-Antitrypsin deficiency
Active	Other genetic factors
Passive	Ethnic factors
Maternal	
Air pollution	Airway hyperresponsiveness?
Outdoor	
Indoor: biomass fuels	
Occupational exposure	Low birth weight
Dietary factors	
High salt	
Low antioxidant vitamins	
Low unsaturated fatty acids	
Infections	

(particularly sulfur dioxide and particulates), exposure to certain occupational chemicals such as cadmium and passive smoking may all be additional risk factors. The role of airway hyperresponsiveness and allergy as risk factors for COPD is still uncertain. Atopy, serum IgE and blood eosinophilia are not important risk factors. However, this is not necessarily the same type of abnormal airway responsiveness that is seen in asthma. Low birth weight is also a risk factor for COPD, probably because poor nutrition in fetal life results in small lungs, so that decline in lung function with age starts from a lower peak value.

Genetic Factors

Longitudinal monitoring of lung function in cigarette smokers reveals that only a minority (15–40% depending on definition) develop significant airflow obstruction due to an accelerated decline in lung function (2- to 5-fold higher than the normal decline of 15–30 ml FEV_1/year) compared to the normal population and the remainder of smokers who have consumed an equivalent number of cigarettes (Figure 47.1). The rate of decline in lung function in the general population (Framingham Study) has a heritability of about 50% (Gottlieb et al., 2001). This strongly suggests that genetic factors may determine which smokers are susceptible and develop airflow limitation. Further evidence that genetic factors are important is the familial clustering of patients with early onset COPD and the differences in COPD prevalence between different ethnic groups (Sandford and Silverman, 2002). Patients with α_1-antitrypsin deficiency (proteinase inhibitor (Pi) ZZ phenotype with α_1-antitrypsin levels $<$10% of normal values) develop early emphysema that is exacerbated by smoking, indicating a clear genetic predisposition to COPD (Carrell and Lomas, 2002). However, α_1-antitrypsin deficiency accounts for $<$1% of patients with COPD, and many other genetic variants of α_1-antitrypsin that are associated with lower than normal serum levels of this proteinase inhibitor have not been clearly associated with an increased risk of COPD, although analysis of several studies indicates a small risk, with an odds ratio from COPD in Pi MZ versus Pi MM individuals of approximately two (Hersh et al., 2004).

This has lead to a search for associations between COPD and single nucleotide polymorphisms (SNPs) of other candidate genes that may be involved in its pathophysiology (Cookson, 2006). Various SNPs have been associated with COPD, as defined by a reduced FEV_1, but there is emerging evidence that different aspects of COPD may relate to different genotypes (Demeo et al., 2007; Hersh et al., 2006a). For example, a reduction in gas diffusion is correlated with SNPs of the microsomal epoxide hydrolase gene, reduced exercise performance is correlated with SNPs of the latent transforming growth factor-beta binding protein-4 (LTBP4), whereas dyspnea was linked to three SNPs in transforming growth factor β1 (TGF-β1). This suggests that more careful phenotyping is required in the future to sort out susceptibility genes. A 10-fold increased risk of COPD in individuals who have a polymorphism in the promoter region of the gene for tumor necrosis factor-α (TNF-α) that is associated with increased TNF-α production has been reported in a Taiwanese population but not confirmed in Caucasian populations. So far, few significant associations have been detected between SNPs and disease, and even those reported have not usually been replicated in other studies or have no obvious functional effects. Several other genes have been implicated in COPD, but few have been replicated in different populations. One SNP that has been replicated is heme oxygenase-1 (HMOX1), which is linked to protection against oxidative stress (Hersh et al., 2005) (Table 47.2). In addition to candidate genes that have been linked to the progressive lung disease, genetic factors are also important in addictive behavior, such as the nicotine addition of smokers, which has a high degree of heritability (Goldman et al., 2005). For example, SNPs of the dopamine transporter gene *SLC6A3* and the dopamine receptor 2 gene have been associated with nicotine addiction and inability to quit (Erblich et al., 2005).

Positional cloning is more likely to identify novel genes, and genome-wide linkage studies have identified some areas of weak linkage. The most convincing evidence is for linkage on chromosome 2q in the region of the *SERPINE2* gene, which encodes an antiprotease (Demeo et al., 2006). Several SNPs of the *SERPINE2* gene have been associated with increased susceptibility to COPD. Various high density genome-wide screens are currently in progress, with the aim of identifying novel genes involved in COPD susceptibility. This approach is now possible with high throughput genotyping of thousands of SNPs in very large populations.

PATHOPHYSIOLOGY

COPD includes chronic obstructive bronchiolitis with fibrosis and obstruction of small airways, and emphysema with enlargement of airspaces and destruction of lung parenchyma, loss of lung elasticity and closure of small airways. Chronic bronchitis, by contrast, is defined by a productive cough of more than 3 months duration for more than 2 successive years; this reflects mucous hypersecretion and is not necessarily associated with airflow limitation. Most patients with COPD have all three pathological mechanisms

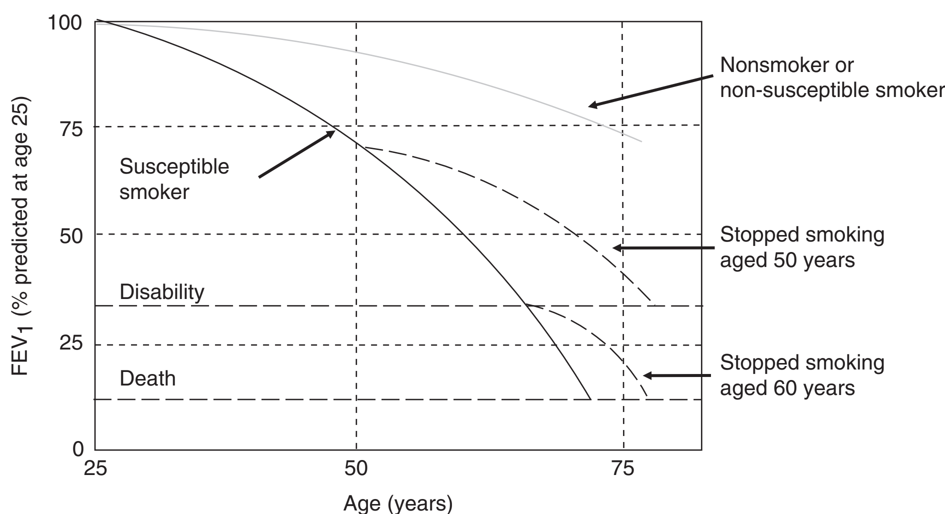


Figure 47.1 Natural history of COPD. Annual decline in airway function showing accelerated decline in susceptible smokers and effects of smoking cessation. Patients with COPD usually show an accelerated annual decline in forced expiratory volume in 1 second (FEV_1), often greater than 50 ml/year, compared to the normal decline of approximately 20 ml/year, although this is variable between patients. The classic studies of Fletcher and Peto established that 10–20% of cigarette smokers are susceptible to this rapid decline. However, with longer follow-up more smokers may develop COPD. The propensity to develop COPD amongst smokers is only weakly related to the amount of cigarettes smoked and this suggests that other factors play an important role in determining susceptibility. Most evidence points toward genetic factors, although the genes determining susceptibility have not yet been determined.

Table 47.2 Some of the genes associated with COPD susceptibility

Candidate genes	Risk
α_1-Antitrypsin	ZZ genotype high risk
	MZ, SZ genotypes small risk
α_1-Chymptrypsin	Associated in some populations
Matrix metalloproteinase-1, −2, −9, −12	Associated in some studies
Microsomal epoxide hydrolase	Increased risk
Glutathione S-transferase	Increased risk
Heme oxygenase-1	Small risk but consistent
Interleukin-13	Small risk
Vitamin D binding protein	Inconsistent
TNF-α promoter	Inconsistent
TGF-β promoter	Inconsistent

(chronic obstructive bronchitis, emphysema and mucus plugging) as all are induced by smoking, but may differ in the proportion of emphysema and obstructive bronchitis (Figure 47.2).

Small Airways

There has been debate about the predominant mechanism of progressive airflow limitation and recent pathological studies suggest that is closely related to the degree of inflammation, narrowing and fibrosis in small airways (Barnes, 2004c; Hogg et al., 2004). Emphysema may contribute to the airway narrowing in the more advanced stages of the disease, with disruption of alveolar attachments facilitating small airway closure and gas trapping. This combined effect of small airway disease and early closure on expiration results in lung hyperinflation, which results in progressive exertional dyspnea, the predominant symptom of COPD.

Emphysema

Emphysema describes loss of alveolar walls due to destruction of matrix proteins (predominantly elastin) and loss of type 1 pneumocytes as a result of apoptosis. Several patterns of emphysema are recognized: centriacinar emphysema radiates from the terminal bronchiole; panacinar emphysema involves more widespread destruction and bullae are large airspaces. Emphysema results in airway obstruction by loss of elastic recoil so that intrapulmonary airways close more readily during expiration. Emphysema with loss of gas-exchanging surface also leads to progressive hypoxia and eventually to respiratory failure.

Pulmonary Hypertension

Chronic hypoxia may lead to hypoxic vasoconstriction, with structural changes in pulmonary vessels that eventually lead to secondary pulmonary hypertension (Naeije, 2005). Inflammatory changes similar to those seen in small airways are also seen in pulmonary arterioles. Only a small proportion of COPD patients develop pulmonary hypertension, and it is likely that genetic susceptibility may play a role.

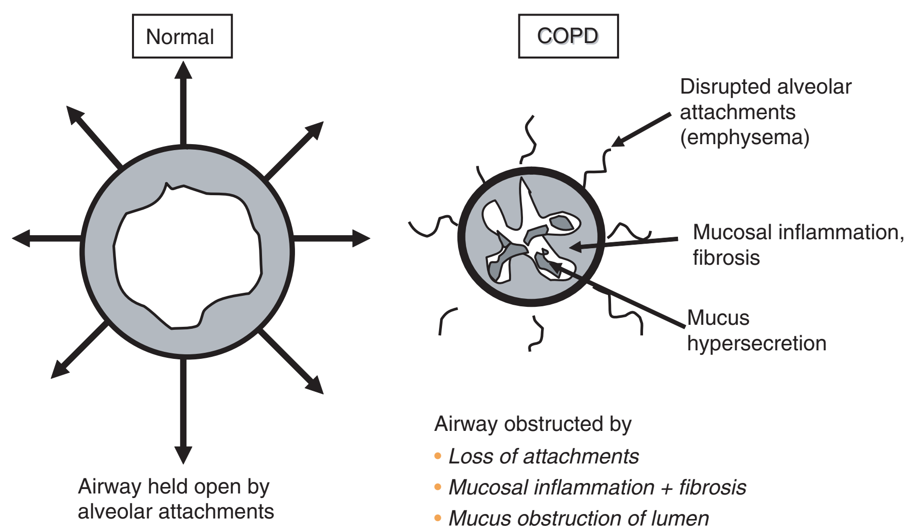


Figure 47.2 Mechanisms of airflow limitation in COPD. The airway in normal subjects is distended by alveolar attachments during expiration, allowing alveolar emptying and lung deflation. In COPD these attachments are disrupted because of emphysema thus contributing to airway closure during expiration, trapping gas in the alveoli and resulting in hyperinflation (Hogg, 2001). Peripheral airways are also obstructed and distorted by airways inflammation and fibrosis (chronic obstructive bronchiolitis) and by occlusion of the airway lumen by mucous secretions which may be trapped in the airways because of poor mucociliary clearance.

Systemic Features

Patients with severe COPD also develop systemic features, which may have an adverse effect on prognosis (Agusti et al., 2003). The most common systemic feature is weight loss due to loss of skeletal muscle bulk. This may contribute to the muscle weaknesses as a result of impaired mobility due to dyspnea. Other systemic features include osteoporosis and depression. These systemic features may be due to overspill of inflammatory mediators from the lung into the systemic circulation. There appears to be a difference in susceptibility to systemic features between patients as they do not always correlate with disease severity and this may be genetically determined. Patients with COPD also have comorbidities, particularly cardiovascular diseases. Smoking is a common risk factor for ischemic heart disease and COPD, but there may be shared genetic susceptibilities. There may be common genetic predispositions to COPD and cardiovascular diseases. COPD patients with coexisting cardiovascular disease have increased circulating levels of C-reactive protein (CRP) (Gan et al., 2004). There is evidence for genetic determinants of plasma CRP concentrations in COPD patients (Hersh et al., 2006b).

Exacerbations

An important feature of COPD are exacerbations, with worsening of dyspnea and an increase in sputum production. This may lead to hospitalization and accounts for a high proportion of the costs of COPD. Exacerbations are usually due to infections, either due to bacteria (especially *Haemophilus influenzae* or *Steptococcus pneumoniae*) or to upper respiratory tract virus infections (especially rhinovirus or respiratory syncytial virus) (Papi et al., 2006). Exacerbations of COPD tend to increase as the disease progresses, but some patients appear to have more frequent exacerbations than others, which may suggest genetic predisposing factors.

CELLULAR AND MOLECULAR MECHANISMS

There is chronic inflammation predominantly in small airways and the lung parenchyma, with an increase in numbers of macrophages and neutrophils in early stages of the disease indicating an enhanced innate immune response, but in more advanced stages of the disease there is an increase in lymphocytes (particularly cytotoxic $CD8^+$ T cells), including lymphoid follicles that contain B- and T-lymphocytes, indicating acquired immunity (Hogg et al., 2004) (Figure 47.3).

Macrophages

Alveolar macrophages play a critical role in the orchestration of this pulmonary inflammation, since they are activated by inhaled irritants such as cigarette smoke and release chemokines that attract inflammatory cells, such as monocytes, neutrophils and T cells, into the lungs (Barnes, 2004a; Barnes et al., 2003). Monocytes are attracted into the lung to differentiate into tissue and alveolar macrophages, the number of which are increased more than 20-fold in COPD compared to cigarette smokers without COPD (Retamales et al., 2001). These cells then

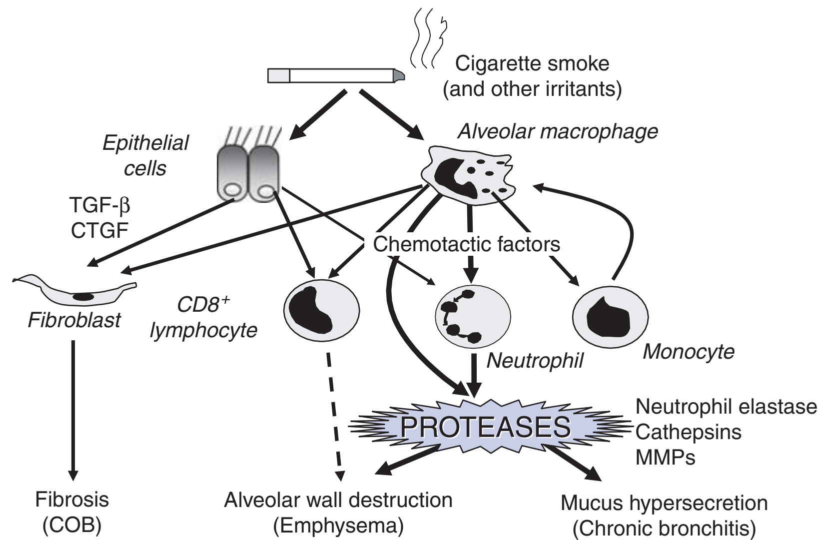


Figure 47.3 Inflammatory mechanisms in COPD. Cigarette smoke (and other irritants) activate macrophages in the respiratory tract that release neutrophil chemotactic factors, including interleukin-8 (IL-8) and leukotriene B4 (LTB4). These cells then release proteases that break down connective tissue in the lung parenchyma, resulting in emphysema, and also stimulate mucus hypersecretion. These enzymes are normally counteracted by protease inhibitors, including α1-antitrypsin, secretory leukoprotease inhibitor (SLPI) and tissue inhibitor of matrix metalloproteinases (TIMP). Cytotoxic T cells ($CD8^+$) may also be recruited and may be involved in alveolar wall destruction.

all release a variety of inflammatory mediators and proteases, which collectively result in the typical pathology of COPD (Barnes, 2004b).

Mediators

Prominent mediators are those that amplify inflammation, such as TNF-α, interleukin(IL)-1β (IL-1β) and IL-6, and chemokines which attract inflammatory cells such as CXCL8 (IL-8), CXCL1 (GROα), CXCL10 (IP-10), CCL1 (MCP) and CCL5 (RANTES) (Donnelly and Barnes, 2006). Elastolytic enzymes account for the tissue destruction of emphysema and include neutrophil elastase and matrix metalloproteinase-9 (MMP-9). There is an imbalance between increased production of elastases and a deficiency of endogenous antiproteases, such as α_1-antritrypin, secretory leukoprotease inhibitor and tissue inhibitors of MMPs. MMP-9 may be the predominantly elastolytic enzyme causing emphysema and also activates TGF-β, a cytokine that is expressed particularly in small airways that may result in the characteristic peribronchiolar fibrosis.

Oxidative stress is a prominent feature of COPD and is due to exogenous oxidants in cigarette smoke and endogenous oxidants release from activated inflammatory cells, such as neutrophils and macrophages (Bowler et al., 2004). Endogenous antioxidants may also be defective. Oxidative stress enhances inflammation and may lead to corticosteroid resistance. There is also increase production of nitric oxide (NO) in peripheral lung of COPD patients, with increased formation of peroxynitrite (Brindicci et al., 2005; Ricciardolo et al., 2005).

Differences from Asthma

Although both COPD and asthma involve chronic inflammation of the respiratory tract, there are marked differences in the inflammatory process between these diseases, which are summarized in Table 47.3 (Barnes, 2000b). While there are striking differences between the inflammation in mild asthma and COPD, patients with severe asthma become much more similar to patients with COPD, with involvement of neutrophils, macrophages, TNF-α, CXCL8, oxidative stress and a poor response to corticosteroids (Barnes, 2006a). This has suggested that there may be similarities in genetic predisposition to COPD in smokers and severe asthma. Several novel susceptibility genes, including DPP10, GPRA, PHF11 and ADAM33, that have been identified in severe asthma have now also been identified in COPD patients (van Diemen et al., 2005).

Gene Regulation

Proinflammatory mediators, such as TNF-α and IL-1β, activate the transcription factor nuclear factor-κB (NF-κB), which is activated in the airways and lung parenchyma of COPD patients, particularly in epithelial cells and macrophages (Di Stefano et al., 2002). There is further activation of NF-κB during exacerbations (Caramori et al., 2003). NF-κB switches on many of the inflammatory genes that are activated in COPD lungs, including chemokines, adhesion molecules such as ICAM-1 and E-selectin, inflammatory enzymes such as cyclo-oxygenase-2 and inducible nitric oxide synthase, elastolytic enzymes such as MMP-9 and

TABLE 47.3 Differences between inflammation in COPD and asthma

Inflammation	COPD	Asthma
Inflammatory cells	Neutrophils	Eosinophils
	CD8+ T cells+++	Mast cells
	CD4+ T cells+	CD4+ T cells
	Macrophages+++	Macrophages+
Inflammatory mediators	LTB_4	LTD_4, histamine
	TNF-α	IL-4, IL-5, IL-13
	CXCL1, CXCL8	CCL11
	Oxidative stress+++	Oxidative stress+
Inflammatory effects	Epithelial metaplasia	Epithelial shedding
	Fibrosis++	Fibrosis+
	Mucus secretion+++	Mucus secretion+
	AHR±	AHR+++
Location	Peripheral airways	All airways
	Predominantly	No parenchymal effects
	Parenchymal destruction	
Response to corticosteroids	±	+++

LT, leukotriene; TNF, tumor necrosis factor; IL, interleukin; GRO, growth-related oncogene; AHR, airway hyperresponsiveness.

proinflammatory mediators such as TNF-α and IL-1β which themselves activate NF-κB. NF-κB-activated genes result in acetylation of core histones (particularly histone-4), which is necessary for activation of inflammatory genes and this is reversed by histone deacetylase-2 (HDAC2) (Barnes et al., 2005). There is a marked reduction in HDAC2 activity and expression in lung parenchyma, airways and alveolar macrophages of COPD patients (Ito et al., 2005). This is a mechanism that can account for the amplified pulmonary inflammation seen in COPD compared to smokers with normal lung function and also explains why COPD patients are not responsive to corticosteroids, since HDAC2 is the mechanism whereby corticosteroids switch off activated inflammatory genes (Barnes, 2006b) (Figure 47.4). The reduction in HDAC2 expression in COPD appears to be secondary to increased oxidative and nitrative stress and the formation of peroxynitrite, which nitrates tyrosine residues on HDAC2.

Genomics

Better knowledge of the complex underlying pathophysiology has identified many candidate genes that appear to be involved in the molecular mechanisms of COPD, and polymorphisms of several of these have been investigated for association with COPD. With gene microarrays it is now possible to study the global expression of all human genes in cells and tissues (transcriptomics). Few studies of global gene expression have been conducted in COPD, but several laboratories are now looking at the differences in gene expression between COPD patients and smokers exposed to similar numbers of cigarettes who have normal lung function. These studies are using alveolar macrophages, circulating monocytes, airway epithelial cells and peripheral lung. Serial analysis of gene expression and microarray analysis in peripheral lung from mild to moderate COPD patients versus normal smokers has demonstrated that a total of 261 genes showed differential expression (Ning et al., 2004). Many of these genes encode inflammatory transcription factors and growth factors. The transcription factor Egr-1 was notably upregulated and shown to increase the activity of mitogen-activated protein kinase pathways (MAPK), which regulate multiple inflammatory genes. A microarray study of human bronchial epithelial cells exposed to nicotine has identified upregulation of several genes in the MAPK pathway (Tsai et al., 2006). Gene expression profiling of alveolar macrophages has demonstrated 40 genes upregulated and 35 downregulated in smokers compared to nonsmokers. Most of these genes belong to the functional categories of immune/inflammatory response, cell adhesion and extracellular matrix, proteases/antiproteases, antioxidants, signal transduction and regulation of transcription (Heguy et al., 2006). The amplification of inflammation in COPD patients compared to equivalent smokers with normal lung function may be genetically determined, through amplification of inflammatory genes or defective expression of anti-inflammatory or protective mechanisms. Studies looking at susceptibility to the effects of cigarette smoke in different strains of mice may also be informative. For example, the susceptible strain AKR/J shows increased expression of cytokines linked to T helper 1 cells after exposure to tobacco smoke compared to non-susceptible strains (C57BL/6/J and NZWlac/J) (Guerassimov et al., 2004). Molecular genomics may identify markers of risk, but may also reveal novel molecular targets for the development of treatments of the future.

A whole genome-wide linkage scan in relatives of patients with familial early-onset COPD found evidence for linkage of airway obstruction with chromosome 2 (LOD 2.60 at 216 cM). In a smokers-only analysis, evidence for linkage was also observed with chromosome 12 (LOD 5.03 at 35 cM) and chromosome 2 (LOD 4.13 at 229 cM), but the specific genes involved have not been identified (Demeo et al., 2004).

Proteomics

While gene microarray analysis gives considerable information about proteins that are transcriptionally regulated it is

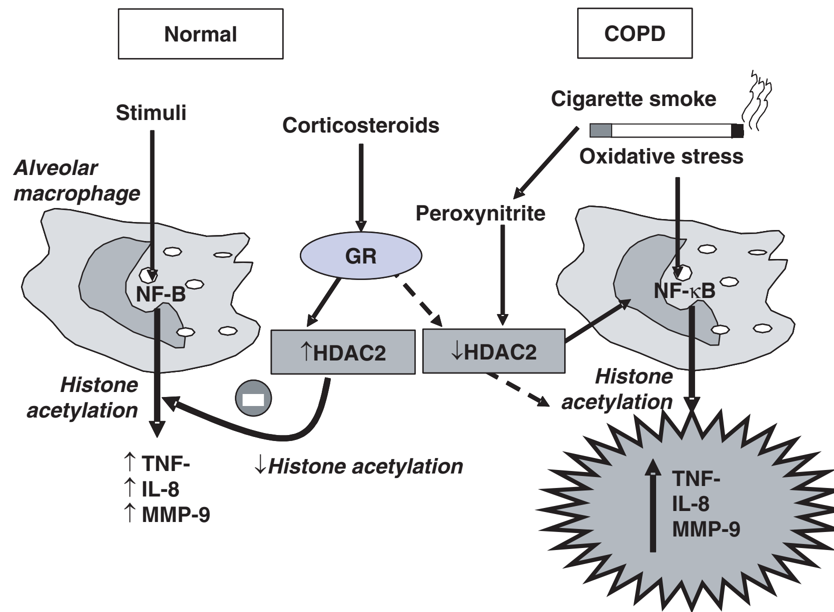


Figure 47.4 Proposed mechanism of corticosteroid resistance in COPD patients. Stimulation of normal alveolar macrophages activates nuclear factor-κB (NF-κB) and other transcription factors to switch on histone acetyltransferase leading to histone acetylation and subsequently to transcription of genes encoding inflammatory proteins, such as TNF-α and IL-8. Corticosteroids reverse this by binding to glucocorticoid receptors (GR) and recruiting histone deacetylase-2 (HDAC2). This reverses the histone acetylation induced by NF-κB and switches off the activated inflammatory genes. In COPD patients cigarette smoke activates macrophages, as in normal subjects, but oxidative stress (acting through the formation of peroxynitrite) impairs the activity of HDAC2. This amplifies the inflammatory response to NF-κB activation, but also reduces the anti-inflammatory effect of corticosteroids as HDAC2 is now unable to reverse histone acetylation.

now recognized that many genes involved in inflammation are regulated post-transcriptionally and that it is necessary to measure protein expression using proteomic approaches (Bowler et al., 2006). A recent study used a proteomic approach to investigate the plasma concentrations of multiple inflammatory and immune proteins during COPD exacerbations compared to baseline measurements (Hurst et al., 2006). CRP was found to be the most discriminant marker and was linked to various proteins associated with monocytic and lymphocytic inflammatory pathways. In another study 143 plasma proteins were assayed in COPD patients compared to control subjects and 25 found to show some correlation with clinical parameters in the COPD patients (Pinto-Plata et al., 2006). Proteomic analysis may also be applied to cells, lung tissue or sputum (Barnes et al., 2006).

DIAGNOSIS AND SCREENING

Diagnosis

Diagnosis is commonly made from the history of progressive dyspnea in a chronic smoker and is confirmed by spirometry, which shows an FEV_1/VC ratio of <70% and FEV_1<80% predicted. Staging of severity is made on the basis of FEV_1, but exercise capacity and the presence of systemic features may be more important determinants of clinical outcome (Celli et al., 2004). Measurement of lung volumes by body plethysmography shows an increase in total lung capacity, residual volume and functional residual capacity, with consequent reduction in inspiratory capacity, representing hyperinflation as a result of small airway closure. This results in dyspnea which may be measured by dyspnea scales and reduced exercise tolerance, which may be measured by a 6 min or shuttle walking test (Figure 47.5). Carbon monoxide diffusion is reduced in proportion to the extent of emphysema.

A chest X-ray is rarely useful but may show hyperinflation of the lungs and the presence of bullae. High resolution computerized tomography demonstrates emphysema, but is not used as a routing diagnostic test. Blood tests are rarely useful; a normocytic normochromic anemia is more commonly seen in patients with severe disease than polycythemia due to chronic hypoxia. Arterial blood gases demonstrate hypoxia and in some patients hypercapnia.

Screening

COPD is grossly underdiagnosed, as symptoms present late in the progression of disease (Chapman et al., 2006). Population screening with measurement of FEV_1 and FEV_1/VC ratio would pick up early COPD in smokers over 40 years. However, this would not identify COPD due to nonsmoking causes, which constitute 10–20% of the total. This would require

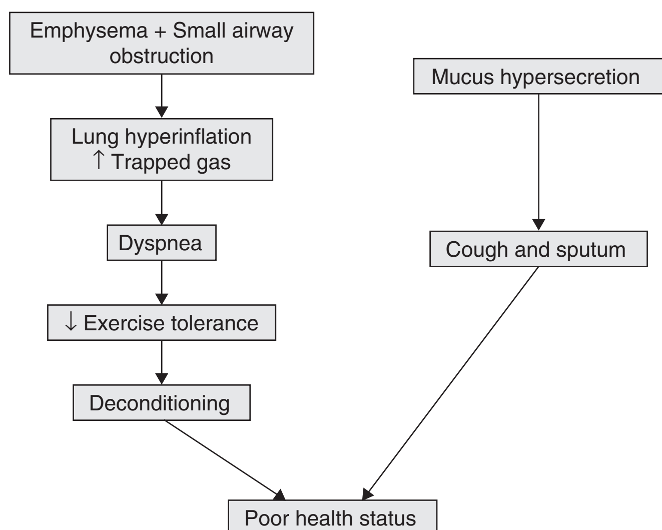


Figure 47.5 Symptoms of COPD. The most prominent symptom of COPD is dyspnea, which is largely due to hyperinflation of the lungs as a result of small airways collapse due to emphysema and narrowing due to fibrosis, so that the alveoli are not able to empty. Hyperinflation induces an uncomfortable sensation and reduces exercise tolerance. This leads to immobility and deconditioning and results in poor health status. Other common symptoms of COPD are cough and sputum production as a result of mucus hypersecretion, but not all patients have these symptoms and many smokers with these symptoms do not have airflow obstruction (simple chronic bronchitis).

screening spirometry of all individuals over 40 years in addition to the 20–40% of the population who smoke. It is hoped that gene expression studies might result in a biomarker that is correlated with COPD susceptibility, but this is unlikely as so many genes are likely to be involved (Barnes et al., 2006).

PROGNOSIS

COPD is slowly progressive with an accelerated decline in FEV_1, leading to slowly increasing symptoms, fall in lung function and eventually to respiratory failure (Figure 47.1). The only strategy to reduce disease progression is smoking cessation, although this is relatively ineffective once FEV_1 has fallen below 50% predicted and the patient is symptomatic. Patients with more severe exacerbations develop acute exacerbations, which have a prolonged effect on quality of life for many months. There is still debate about the role of acute exacerbations on disease progression, but the decline in lung function may be accelerated further following an acute exacerbation (Donaldson et al., 2002). Factors other than FEV_1 are important in the prognosis of the disease, including exercise performance and muscle weakness, which signal a greater mortality (Celli et al., 2004). Patients who develop right heart failure (*cor pulmonale*) also have poor survival, although this may be improved by long-term oxygen therapy. It is hoped in the future that genetic approaches will help to improve the prediction of prognosis and identify what the most effective management strategies may be.

MANAGEMENT

COPD is managed according to the severity of the disease, with a progressive escalation of therapy as the disease progresses (Barnes and Stockley, 2005; Rennard, 2004).

Anti-Smoking Measures

Smoking cessation is the only measure so far shown to slow the progression of COPD, but in advanced disease stopping smoking has little effect and the chronic inflammation persists. Nicotine replacement therapy (gum, transdermal patch, inhaler) helps in quitting smoking, but bupropion, a noradrenergic antidepressant, is more effective. More effective anti-smoking therapies, including partial nicotinic agonists and cannabinoid receptor antagonists are currently being evaluated. As discussed above, there may be genetic determinants of smoking cessation that make it very difficult for some patients to quit.

Bronchodilators

Bronchodilators are the mainstay of current drug therapy for COPD. The bronchodilator response measured by an increase in FEV_1 is limited in COPD, but bronchodilators may improve symptoms by reducing hyperinflation and therefore dyspnea, and may improve exercise tolerance, despite the fact that there is little improvement in spirometric measurements. Previously short-acting bronchodilators, including β_2-agonists and anticholinergics, were most widely used, but more recently long-acting bronchodilators have been introduced. These include the inhaled long-acting β_2-agonists (LABA) salmeterol and formoterol and the once daily inhaled anticholinergic tiotropium bromide. In patients with more severe disease, these therapies appear to be additive. Theophylline is also used as an add-on bronchodilator in patients with very severe disease, but systemic side effects may limit its value.

Antibiotics

Acute exacerbations of COPD are commonly assumed to be due to bacterial infections, since they may be associated with increased volume and purulence of the sputum. However, it is increasingly recognized that exacerbations may be due to upper respiratory tract viral infections or may be non-infective, questioning the place of antibiotic treatment in many patients. Controlled trials of antibiotics in COPD show a relatively minor benefit of antibiotics in terms of clinical outcomes and lung function. Although antibiotics are still widely used in exacerbations of COPD, methods that can reliably diagnose bacterial infection in the respiratory tract are needed so that antibiotics are not used inappropriately. There is no evidence that prophylactic antibiotics prevent acute exacerbations.

Oxygen

Home oxygen accounts for a large proportion (over 30% in the United States) of health care spending on COPD. Long-term oxygen therapy was justified by two large trials showing reduced mortality and improvement in quality of life in patients with severe COPD and chronic hypoxemia ($PaO_2 < 55$ mmHg). More recent studies have demonstrated that patients with less severe hypoxemia do not appear to benefit in terms of increased survival, so that selection of patients is important in prescribing this expensive therapy. Similarly, nocturnal treatment with oxygen does not appear to be beneficial in terms of survival or delaying the prescription of long-term oxygen therapy in patients with COPD who have nocturnal hypoxemia.

Corticosteroids

Inhaled corticosteroids are now the mainstay of chronic asthma therapy and the recognition that chronic inflammation is also present in COPD provided a rationale for their use in COPD. Indeed, inhaled corticosteroids are now widely prescribed in the treatment of COPD. However, the inflammation in COPD shows little or no suppression even by high doses of inhaled or oral corticosteroids. This may reflect the fact that neutrophilic inflammation is not suppressible by corticosteroids as neutrophil survival is prolonged by steroids. There is also evidence for an active cellular resistance to corticosteroids, with no evidence that even high doses of corticosteroids suppress the synthesis of inflammatory mediators or enzymes. This is related to a decreased activity and expression of HDAC2 (Barnes, 2006b). Approximately 10% of patients with stable COPD show some symptomatic and objective improvement with oral corticosteroids, and it is likely that these patients have concomitant asthma, as both diseases are very common. Furthermore, these patients have elevated sputum eosinophils and exhaled NO, which are features of asthmatic inflammation. Long-term treatment with high doses of inhaled corticosteroids fails to reduce disease progression, even at the early stages of the disease. However, there is a small protective effect against acute exacerbations (approximately 20% reduction) in patients with severe disease. In view of the risk of systemic side effects in this susceptible population, inhaled corticosteroid are only recommended in patients with $FEV_1 < 50\%$ predicted who have two or more severe exacerbations a year.

There is a small beneficial effect of systemic corticosteroids in treating acute exacerbations of COPD, with improved clinical outcome and reduced length of hospital admission. The reasons for this discrepancy between steroid responses in acute versus chronic COPD may relate to differences in the inflammatory response (increased eosinophils) or airway edema in exacerbations.

Other Drug Therapies

Systematic reviews show that mucolytic therapies reduce exacerbation by about 20%, but most of the benefit appears to derive from *N*-acetylcysteine which is also an antioxidant. A controlled trial, however, did not show any overall benefit in reducing exacerbations or disease progression (Decramer et al., 2005).

Pulmonary Rehabilitation

Pulmonary rehabilitation consists of a structured program of education, exercises and physiotherapy and has been shown in controlled trials to improve the exercise capacity and quality of life of patients with severe COPD, with a reduction in health care utilization. Pulmonary rehabilitation is now an important part of the management plan in patients with severe COPD. There is debate about the duration and frequency of pulmonary rehabilitation (Troosters et al., 2005). Most of the benefits appear to relate to exercise so that modified simplified programs are now often used.

Lung Volume Reduction

Surgical removal of emphysematous lung improves ventilatory function in carefully selected patients (Wouters, 2004). The reduction in hyperinflation improves the mechanical efficiency of the inspiratory muscles. Careful patient selection after a period of pulmonary rehabilitation is essential. Patients with localized upper lobe emphysema with poor exercise capacity do best, but there is a relatively high operative mortality, particularly with patients who have a low diffusing capac ity. Significant functional improvements include increased FEV_1, reduced total lung capacity and functional residual capacity, improved function of respiratory muscles, improved exercise capacity and improved quality of life. Benefits persist for at least a year in most patients, but careful long-term follow-up is needed in order to evaluate the long-term benefits of this therapy. More recently nonsurgical bronchoscopic lung volume reduction has been achieved by insertion of one-way valves by fiberoptic bronchoscopy. This gives significant improvement in some patients and appears to be safe, but collateral ventilation reduces the efficacy of this treatment so that significant deflation of affected lung may not be achieved.

Management of Acute Exacerbations

Acute exacerbations of COPD should be managed by supplementary oxygen therapy, initially 24% oxygen and checking that there is no depression of ventilation. Antibiotics should be given if the sputum is purulent or there are other signs of bacterial infection, High doses of corticosteroids reduce hospital stay and are routinely given. Chest physiotherapy is usually given, but there is little objective evidence for benefit. Noninvasive ventilation is indicated for incipient respiratory failure and reduces the need for intubation.

Pharmacogenomics

There is no information about the impact of pharmacogenomics on COPD therapy. Polymorphism of the β_2-receptor affect the bronchodilator response to short- and long-acting β_2-agonists in asthma, with reduce responses seen in patients with the Arg–Arg[16] polymorphism (Liggett, 2002). However this is not a large effect and is unlikely to have a major clinical impact. It is possible that polymorphism in corticosteroids signaling pathways may contribute to the corticosteroids resistance in COPD

but this has not yet been investigated. Using genomics to better define the clinical phenotypes may lead to a more rational use of specific therapies in the future.

Impacts of Genomics on Therapy

As indicated above, molecular genetics and genomics may identify novel targets for the development of new therapies. Pharmacogenomics might have an impact on choice of therapy in the future. It is unlikely that gene-based therapies will be useful, but silencing of inflammatory genes by inhaled interfering RNAs and antisense oligonucleotides might be a feasible approach in the future (Ulanova et al., 2006). Micro-RNAs (miRNA) are small regulatory RNAs that inhibit the translation of several activated genes and may have application to the suppression of multiple inflammatory genes in the future (Ying et al., 2006).

2009 UPDATE

Gene Polymorphism Studies

Several new associations between COPD and SNPs in candidate genes have recently been described. Stratified linkage analysis in early onset COPD patients with emphysema identified an association with an intronic SNP of transforming growth factor receptor-3 (TGFR3, betaglycan), and this was confirmed in an independent COPD population (Hersh et al., 2009). Studies in mice and microarray studies in COPD have also identified this target, consistent with the previously demonstrated SNP of TGFβ1, indicating an abnormality in the TGF signaling pathway that may be relevant to the small airway fibrosis that is a key component of COPD. The association between SNPs of TGFβ in emphysema patients has also been confirmed in a Japanese population (Ito et al., 2008).

One of the problems in the genetics of COPD is the clinical heterogeneity of the disease. There is now an effort to link different aspects of the disease to specific genetic associations in order to classify disease subtypes that might be useful in selecting patients for different types of therapy in the future. Several SNPs were reported to be associated with severe COPD in almost 400 patients who participated in the National Emphysema Treatment Trial (NETT). Two of these genes – microsomal epoxide hydrolase (EPHX1) and SERPINE2 – were found to be significantly associated with hypoxema. In a separate population of early onset COPD patients, the same SNPs were associated with the requirement of domiciliary oxygen (Castaldi et al., 2008). Another SNP in the surfactant protein B (SFTPB) gene was associated with pulmonary hypertension. The same research group has also linked bronchodilator responsiveness to the inhaled β_2-agonist albuterol with SNPs in the SERPINE2 and EPHX1 genes, as well as in the β_2-adrenergic receptor (ADRB2) gene (Kim et al., 2008). Extracellular superoxide dismutase (SOD3) is an important antioxidant enzyme in the lungs, and two novel SNPs of SOD3 have been linked to reduced FEV_1 values in a Danish population study and these polymorphisms were associated with hospitalization and mortality from COPD (Dahl et al., 2008).

Matrix metalloproteinases appear to play an important role in the remodeling of COPD lungs and MMP-9 is the predominant elastolytic enzyme in the lung periphery. There is a significant association between a genetic variant of MMP-9, Gln279Arg, which is associated with the increased substrate binding that is increased in COPD patients (Tesfaigzi et al., 2006).

Genome-Wide Association Studies

Although genome-wide association studies (GWAS) are now widely applied to complex diseases, this approach has not yet been applied successfully in COPD patients. The GWAS approach has been applied to over 70,000 samples from the Framingham Heart Study, and reduced lung function (lower FEV_1 and FEV_1/FVC ratio) was significantly associated with SNPs in the IL-6 receptor and glutathione *S*-transferase omega-2 genes, as well as SOD3 (Wilk et al., 2007). Several GWAS in COPD are currently underway.

Biomarker Studies

There is continued search for novel biomarkers in COPD that might predict disease progression and response to therapy. Desmosine and isodesmosine are amino acids that are unique to elastin fibers and may therefore have potential as markers of elastin degradation in COPD lungs. However, they have previously proved to be of little value because of the high background levels. Novel separation methods and more sensitive detection methods are now being applied and indicate that these biomarkers are increased in urine, plasma and sputum of COPD patients (Luisetti et al., 2008).

Gene Expression Studies

There is increasing use of microarrays to study abnormal patterns of gene expression in cells from COPD patients. Airway epithelial cells from COPD patients show downregulation of the Notch signaling pathway, which is involved in epithelial regeneration (Tilley et al., 2009). Microarray analysis of peripheral lung from COPD patients also shows abnormal expression of genes involved in epithelial repair and matrix synthesis and degradation (Wang et al., 2008). However, there appears to be enormous variability in the expression of cytokine

and chemokine genes in small airway epithelial cells and macrophages in response to cigarette smokers, which is greater in COPD patients than in normal smokers and nonsmokers (Ammous et al., 2008; Howrylak and Choi, 2008). This variability means that large numbers of samples are likely to be needed to characterize the COPD transcriptome (Zeskind et al., 2008).

Disease Mechanisms

Increased oxidative stress is important in the pathophysiology of COPD and may be an important driver of chronic inflammation and corticosteroid resistance. As well as an increase in oxidants from cigarette smoke and other inhaled particles and from activated inflammatory cells, there may also be a defect in antioxidant defenses in the lung, which may be genetically determined. The transcription factor nuclear erythroid- 2 related factor 2 (Nrf2) is important in regulating the transcription of many antioxidant and detoxifying genes. A recent study has demonstrated that Nrf2 activity is reduced in the lungs and epithelial cells of COPD patients and that this is associated with reduced expression of several antioxidant genes (Malhotra et al., 2008). Nrf2 activity has also been found to be reduced in alveolar macrophages from COPD patients (Suzuki et al., 2008) This defect is related to reduced nuclear translocation of Nrf2 in response to oxidative stress as a result of defect in the protein DJ1, which normally stabilizes Nrf2 in the cytoplasm. There is no obvious association between SNPs of Nrf2 and COPD, but other genes in the pathway such as DJ1 have not yet been studied. This defect in Nrf2 may be reversed using drugs, such as sulforaphane, which occurs naturally in broccoli (Barnes, 2008). A more potent and selective Nrf2 activator triterpenoid CDDO-imidazolide prevents cigarette smoke-induced emphysema in mice (Sussan et al., 2009).

2009 UPDATE REFERENCES

Ammous, Z., Hackett, N.R., Butler, M.W., Raman, T., Dolgalev, I., O'Connor, T.P., Harvey, B.G. and Crystal, R.G. (2008). Variability in small airway epithelial gene expression among normal smokers. *Chest* 133, 1344–1353.

Barnes, P.J. (2008). Defective antioxidant gene regulation in COPD: A case for broccoli. *Am J Respir Crit Care Med* 178, 552–554.

Castaldi, P.J., Hersh, C.P., Reilly, J.J. and Silverman, E.K. (2008). Genetic associations with hypoxemia and pulmonary arterial pressure in COPD. *Chest* 135(3), 737–744.

Dahl, M., Bowler, R.P., Juul, K., Crapo, J.D., Levy, S. and Nordestgaard, B.G. (2008). Superoxide dismutase 3 polymorphism associated with reduced lung function in two large populations. *Am J Respir Crit Care Med* 178, 906–912.

Hersh, C.P., Hansel, N.N., Barnes, K.C., Lomas, D.A., Pillai, S.G., Coxson, H.O., Mathias, R.A., Rafaels, N.M., Wise, R.A., Connett, J.E. et al. (2009). Transforming growth factor beta receptor-3 is associated with pulmonary emphysema. *Am J Respir Cell Mol Biol* [Epub ahead of print].

Howrylak, J. and Choi, A.M. (2008). Gene expression profiling in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 177, 364–365.

Ito, M., Hanaoka, M., Droma, Y., Hatayama, O., Sato, E., Katsuyama, Y., Fujimoto, K. and Ota, M. (2008). The association of transforming growth factor beta 1 gene polymorphisms with the emphysema phenotype of COPD in Japanese. *Intern Med* 47, 1387–1394.

Kim, W.J., Hersh, C.P., Demeo, D.L., Reilly, J.J. and Silverman, E.K. (2008). Genetic association analysis of COPD candidate genes with bronchodilator responsiveness. *Respir Med* 103(4), 552–557.

Luisetti, M., Ma, S., Iadarola, P., Stone, P.J., Viglio, S., Casado, B., Lin, Y.Y., Snider, G.L. and Turino, G.M. (2008). Desmosine as a biomarker of elastin degradation in COPD: Current status and future directions. *Eur Respir J* 32, 1146–1157.

Malhotra, D., Thimmulappa, R., Navas-Acien, A., Sandford, A., Elliott, M., Singh, A., Chen, L., Zhuang, X., Hogg, J., Pare, P. et al. (2008). Decline in NRF2 regulated antioxidants in COPD lungs due to loss of its positive regulator DJ-1. *Am J Respir Crit Care Med* 178, 592–604.

Sussan, T.E., Rangasamy, T., Blake, D.J., Malhotra, D., El Haddad, H., Bedja, D., Yates, M.S., Kombairaju, P., Yamamoto, M., Liby, K.T. et al. (2009). Targeting Nrf2 with the triterpenoid CDDO-imidazolide attenuates cigarette smoke-induced emphysema and cardiac dysfunction in mice. *Proc Natl Acad Sci U S A* 106, 250–255.

Suzuki, M., Betsuyaku, T., Ito, Y., Nagai, K., Nasuhara, Y., Kaga, K., Kondo, S. and Nishimura, M. (2008). Down-regulated NF-E2-related factor 2 in pulmonary macrophages of aged smokers and patients with chronic obstructive pulmonary disease. *Am J Respir Cell Mol Biol* 39, 673–682.

Tesfaigzi, Y., Myers, O.B., Stidley, C.A., Schwalm, K., Picchi, M., Crowell, R.E., Gilliland, F.D. and Belinsky, S.A. (2006). Genotypes in matrix metalloproteinase 9 are a risk factor for COPD, *Int J Chron Obstruct Pulmon Dis* 1, 267–278.

Tilley, A.E., Harvey, B.G., Heguy, A., Hackett, N.R., Wang, R., O'Connor, T.P. and Crystal, R.G. (2009). Downregulation of the notch pathway in human airway epithelium in association with smoking and COPD. *Am J Respir Crit Care Med* 179(6), 457–466.

Wang, I.M., Stepaniants, S., Boie, Y., Mortimer, J.R., Kennedy, B., Elliott, M., Hayashi, S., Loy, L., Coulter, S., Cervino, S. et al. (2008). Gene expression profiling in patients with chronic obstructive pulmonary disease and lung cancer. *Am J Respir Crit Care Med* 177, 402–411.

Wilk, J.B., Walter, R.E., Laramie, J.M., Gottlieb, D.J. and O'Connor, G.T. (2007). Framingham Heart Study genome-wide association: Results for pulmonary function measures. *BMC Med Genet* 8(Suppl 1) S8.

Zeskind, J.E., Lenburg, M.E. and Spira, A. (2008). Translating the COPD transcriptome: Insights into pathogenesis and tools for clinical management. *Proc Am Thorac Soc* 5, 834–841.

REFERENCES

Agusti, A.G., Noguera, A., Sauleda, J., Sala, E., Pons, J. and Busquets, X. (2003). Systemic effects of chronic obstructive pulmonary disease. *Eur Respir J* 21, 347–360.

Barnes, P.J. (2000a). Chronic obstructive pulmonary disease. *N Engl J Med* 343, 269–280.

Barnes, P.J. (2000b). Mechanisms in COPD: Differences from asthma. *Chest* 117, 10S–14S.

Barnes, P.J. (2004a). Macrophages as orchestrators of COPD. *J COPD* 1, 59–70.

Barnes, P.J. (2004b). Mediators of chronic obstructive pulmonary disease. *Pharm Rev* 56, 515–548.

Barnes, P.J. (2004c). Small airways in COPD. *N Engl J Med* 350, 2635–2637.

Barnes, P.J. (2006a). Against the Dutch hypothesis: Asthma and chronic obstructive pulmonary disease are distinct diseases. *Am J Respir Crit Care Med* 174, 240–243.

Barnes, P.J. (2006b). Reduced histone deacetylase in COPD: Clinical implications. *Chest* 129, 151–155.

Barnes, P.J. and Stockley, R.A. (2005). COPD: Current therapeutic interventions and future approaches. *Eur Respir J* 25, 1084–1106.

Barnes, P.J., Shapiro, S.D. and Pauwels, R.A. (2003). Chronic obstructive pulmonary disease: Molecular and cellular mechanisms. *Eur Respir J* 22, 672–688.

Barnes, P.J., Adcock, I.M. and Ito, K. (2005). Histone acetylation and deacetylation: Importance in inflammatory lung diseases. *Eur Respir J* 25, 552–563.

Barnes, P.J., Chowdhury, B., Kharitonov, S.A., Magnussen, H., Page, C.P., Postma, D. and Saetta, M. (2006). Pulmonary biomarkers in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 174, 6–14.

Bowler, R.P., Barnes, P.J. and Crapo, J.D. (2004). The role of oxidative stress in chronic obstructive pulmonary disease. *J COPD* 2, 255–277.

Bowler, R.P., Ellison, M.C. and Reisdorph, N. (2006). Proteomics in pulmonary medicine. *Chest* 130, 567–574.

Brindicci, C., Ito, K., Resta, O., Pride, N.B., Barnes, P.J. and Kharitonov, S.A. (2005). Exhaled nitric oxide from lung periphery is increased in COPD. *Eur Respir J* 26, 52–59.

Caramori, G., Romagnoli, M., Casolari, P., Bellettato, C., Casoni, G., Boschetto, P., Fan, C.K., Barnes, P.J., Adcock, I.M., Ciaccia, A. et al. (2003). Nuclear localisation of p65 in sputum macrophages but not in sputum neutrophils during COPD exacerbations. *Thorax* 58, 348–351.

Carrell, R.W. and Lomas, D.A. (2002). Alpha1-antitrypsin deficiency – a model for conformational diseases. *N Engl J Med* 346, 45–53.

Celli, B.R., Cote, C.G., Marin, J.M., Casanova, C., Montes, d.O., Mendez, R.A., Pinto, P.V. and Cabral, H.J. (2004). The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. *N Engl J Med* 350, 1005–1012.

Chapman, K.R., Mannino, D.M., Soriano, J.B., Vermeire, P.A., Buist, A.S., Thun, M.J., Connell, C., Jemal, A., Lee, T.A. et al. (2006). Epidemiology and costs of chronic obstructive pulmonary disease. *Eur Respir J* 27, 188–207.

Cookson, W.O. (2006). State of the art. Genetics and genomics of chronic obstructive pulmonary disease. *Proc Am Thorac Soc* 3, 473–475.

Decramer, M., Rutten-van Molken, M., Dekhuijzen, P.N., Troosters, T., van Herwaarden, C., Pellegrino, R., Van Schayck, C.P., Olivieri, D., Del Donno, M., De Backer, W. et al. (2005). Effects of N-acetylcysteine on outcomes in chronic obstructive pulmonary disease (Bronchitis Randomized on NAC Cost-Utility Study, BRONCUS): A randomised placebo-controlled trial. *Lancet* 365, 1552–1560.

Demeo, D.L., Celedon, J.C., Lange, C., Reilly, J.J., Chapman, H.A., Sylvia, J.S., Speizer, F.E., Weiss, S.T. and Silverman, E.K. (2004). Genome-wide linkage of forced mid-expiratory flow in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 170, 1294–1301.

Demeo, D.L., Mariani, T.J., Lange, C., Srisuma, S., Litonjua, A.A., Celedon, J.C., Lake, S.L., Reilly, J.J., Chapman, H.A., Mecham, B.H. et al. (2006). The SERPINE2 gene is associated with chronic obstructive pulmonary disease. *Am J Hum Genet* 78, 253–264.

Demeo, D.L., Hersh, C.P., Hoffman, E.A., Litonjua, A.A., Lazarus, R., Sparrow, D., Benditt, J.O., Criner, G., Make, B., Martinez, F.J. et al. (2007). Genetic determinants of emphysema distribution in the National Emphysema Treatment Trial. *Am J Respir Crit Care Med* 176, 42–48.

Di Stefano, A., Caramori, G., Capelli, A., Lusuardi, M., Gnemmi, I., Ioli, F., Chung, K.F., Donner, C.F., Barnes, P.J. and Adcock, I.M. (2002). Increased expression of NF-κB in bronchial biopsies from smokers and patients with COPD. *Eur Respir J* 20, 556–563.

Donaldson, G.C., Seemungal, T.A., Bhowmik, A. and Wedzicha, J.A. (2002). Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. *Thorax* 57, 847–852.

Donnelly, L.E. and Barnes, P.J. (2006). Chemokine receptors as therapeutic targets in chronic obstructive pulmonary disease. *Trend Pharmacol Sci* 27, 546–553.

Erblich, J., Lerman, C., Self, D.W., Diaz, G.A. and Bovbjerg, D.H. (2005). Effects of dopamine D2 receptor (DRD2) and transporter (SLC6A3) polymorphisms on smoking cue-induced cigarette craving among African-American smokers. *Mol Psychiatr* 10, 407–414.

Gan, W.Q., Man, S.F., Senthilselvan, A. and Sin, D.D. (2004). Association between chronic obstructive pulmonary disease and systemic inflammation: A systematic review and a meta-analysis. *Thorax* 59, 574–580.

Goldman, D., Oroszi, G. and Ducci, F. (2005). The genetics of addictions: Uncovering the genes. *Nat Rev Genet* 6, 521–532.

Gottlieb, D.J., Wilk, J.B., Harmon, M., Evans, J.C., Joost, O., Levy, D., O'Connor, G.T. and Myers, R.H. (2001). Heritability of longitudinal change in lung function. The Framingham study. *Am J Respir Crit Care Med* 164, 1655–1659.

Guerassimov, A., Hoshino, Y., Takubo, Y., Turcotte, A., Yamamoto, M., Ghezzo, H., Triantafillopoulos, A., Whittaker, K., Hoidal, J.R. and Cosio, M.G. (2004). The development of emphysema in cigarette smoke-exposed mice is strain dependent. *Am J Respir Crit Care Med* 170, 974–980.

Heguy, A., O'Connor, T.P., Luettich, K., Worgall, S., Cieciuch, A., Harvey, B.G., Hackett, N.R. and Crystal, R.G. (2006). Gene expression profiling of human alveolar macrophages of phenotypically normal smokers and nonsmokers reveals a previously unrecognized subset of genes modulated by cigarette smoking. *J Mol Med* 84, 318–328.

Hersh, C.P., Dahl, M., Ly, N.P., Berkey, C.S., Nordestgaard, B.G. and Silverman, E.K. (2004). Chronic obstructive pulmonary disease in alpha1-antitrypsin PI MZ heterozygotes: A meta-analysis. *Thorax* 59, 843–849.

Hersh, C.P., Demeo, D.L., Lange, C., Litonjua, A.A., Reilly, J.J., Kwiatkowski, D., Laird, N., Sylvia, J.S., Sparrow, D., Speizer, F.E.

et al. (2005). Attempted replication of reported chronic obstructive pulmonary disease candidate gene associations. *Am J Respir Cell Mol Biol* 33, 71–78.

Hersh, C.P., Demeo, D.L., Lazarus, R., Celedon, J.C., Raby, B.A., Benditt, J.O., Criner, G., Make, B., Martinez, F.J., Scanlon, P.D. et al. (2006a). Genetic association analysis of functional impairment in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 173, 977–984.

Hersh, C.P., Miller, D.T., Kwiatkowski, D.J. and Silverman, E.K. (2006b). Genetic determinants of C-reactive protein in chronic obstructive pulmonary disease. *Eur Respir J.*

Hogg, J.C. (2001). Chronic obstructive pulmonary disease: An overview of pathology and pathogenesis. *Novartis Found Symp* 234, 4–19.

Hogg, J.C. (2004). Pathophysiology of airflow limitation in chronic obstructive pulmonary disease. *Lancet* 364, 709–721.

Hogg, J.C., Chu, F., Utokaparch, S., Woods, R., Elliott, W.M., Buzatu, L., Cherniack, R.M., Rogers, R.M., Sciurba, F.C., Coxson, H.O. et al. (2004). The nature of small-airway obstruction in chronic obstructive pulmonary disease. *N Engl J Med* 350, 2645–2653.

Hurst, J.R., Donaldson, G.C., Perea, W.R., Wilkinson, T.M., Bilello, J.A., Hagan, G.W., Vessey, R.S. and Wedzicha, J.A. (2006). Utility of plasma biomarkers at exacerbation of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*.

Ito, K., Ito, M., Elliott, W.M., Cosio, B., Caramori, G., Kon, O.M., Barczyk, A., Hayashi, M., Adcock, I.M., Hogg, J.C. et al. (2005). Decreased histone deacetylase activity in chronic obstructive pulmonary disease. *N Engl J Med* 352, 1967–1976.

Liggett, S.B. (2002). Polymorphisms of the beta2-adrenergic receptor. *N.Engl J Med* 346, 536–538.

Lopez, A.D., Shibuya, K., Rao, C., Mathers, C.D., Hansell, A.L., Held, L.S., Schmid, V. and Buist, S. (2006). Chronic obstructive pulmonary disease: Current burden and future projections. *Eur Respir J* 27, 397–412.

Naeije, R. (2005). Pulmonary hypertension and right heart failure in chronic obstructive pulmonary disease. *Proc Am Thorac Soc* 2, 20–22.

Ning, W., Li, C.J., Kaminski, N., Feghali-Bostwick, C.A., Alber, S.M., Di, Y.P., Otterbein, S.L., Song, R., Hayashi, S., Zhou, Z. et al. (2004). Comprehensive gene expression profiles reveal pathways related to the pathogenesis of chronic obstructive pulmonary disease. *Proc Natl Acad Sci USA* 101, 14895–14900.

Papi, A., Luppi, F., Franco, F. and Fabbri, L.M. (2006). Pathophysiology of exacerbations of chronic obstructive pulmonary disease. *Proc Am Thorac Soc* 3, 245–251.

Pinto-Plata, V., Toso, J., Lee, K., Bilello, J., Mullerova, H., De Souza, M., Vessey, R. and Celli, B. (2006). Use of proteomic patterns of serum biomarkers in patients with chronic obstructive pulmonary disease: Correlation with clinical parameters. *Proc Am Thorac Soc* 3, 465–466.

Rennard, S.I. (2004). Treatment of stable chronic obstructive pulmonary disease. *Lancet* 364, 791–802.

Retamales, I., Elliott, W.M., Meshi, B., Coxson, H.O., Pare, P.D., Sciurba, F.C., Rogers, R.M., Hayashi, S. and Hogg, J.C. (2001). Amplification of inflammation in emphysema and its association with latent adenoviral infection. *Am J Respir Crit Care Med* 164, 469–473.

Ricciardolo, F.L., Caramori, G., Ito, K., Capelli, A., Brun, P., Abatangelo, G., Papi, A., Chung, K.F., Adcock, I., Barnes, P.J. et al. (2005). Nitrosative stress in the bronchial mucosa of severe chronic obstructive pulmonary disease. *J Allergy Clin Immunol* 116, 1028–1035.

Sandford, A.J. and Silverman, E.K. (2002). Chronic obstructive pulmonary disease. 1: Susceptibility factors for COPD the genotype–environment interaction. *Thorax* 57, 736–741.

Troosters, T., Casaburi, R., Gosselink, R. and Decramer, M. (2005). Pulmonary rehabilitation in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 172, 19–38.

Tsai, J.R., Chong, I.W., Chen, C.C., Lin, S.R., Sheu, C.C. and Hwang, J.J. (2006). Mitogen-activated protein kinase pathway was significantly activated in human bronchial epithelial cells by nicotine. *DNA Cell Biol* 25, 312–322.

Ulanova, M., Schreiber, A.D. and Befus, A.D. (2006). The future of antisense oligonucleotides in the treatment of respiratory diseases. *BioDrugs* 20, 1–11.

van Diemen, C.C., Postma, D.S., Vonk, J.M., Bruinenberg, M., Schouten, J.P. and Boezen, H.M. (2005). A disintegrin and metalloprotease 33 polymorphisms and lung function decline in the general population. *Am J Respir Crit Care Med* 172, 329–333.

Wouters, E.F. (2004). Management of severe COPD. *Lancet* 364, 883–895.

Ying, S.Y., Chang, D.C., Miller, J.D. and Lin, S.L. (2006). The microRNA: Overview of the RNA gene that modulates gene functions. *Methods Mol Biol* 342, 1–18. .

CHAPTER

Genetics and Genomics of Interstitial Lung Disease

Paul W. Noble and Mark P. Steele

INTRODUCTION

Interstitial lung diseases (ILDs), also referred to as diffuse parenchymal lung diseases (DPLDs), are a diverse group of lung diseases that can be classified according to clinical, radiologic, physiologic, or pathologic criteria that result in fibrosis of the alveolar interstitium and impairment of gas exchange. The term DPLD more accurately describes these entities, since among the ILDs there is substantial variation in the degree of involvement of lung structures other than the alveolar interstitium, including capillaries, terminal and respiratory bronchioles, and lymphatics along the bronchovascular bundle and interlobular septae. A general classification scheme of ILD/DPLD categorizes DPLD related to connective tissue diseases, drug-induced diseases, occupational and environmental exposures, granulomatous diseases, inherited conditions such as familial interstitial pneumonia (FIP) and Hermansky-Pudlak syndrome, unique conditions such as eosinophilic granuloma and amyloidosis, and the idiopathic interstitial pneumonias (IIPs). While the pathogenic mechanisms are known or inferred in some of the DPLDs such as those related to environmental exposure, drug exposure, or autoimmune mechanisms, the pathogenesis of most of these entities is poorly understood. Furthermore, it is well recognized that there is considerable individual variation in the natural history of DPLD, particularly regarding the susceptibility to agents known to cause pulmonary fibrosis, such as radiation or asbestos. Since the lung responds to these injuries in a limited fashion, when using standard radiologic and pathologic evaluation, there is significant overlap and diagnostic uncertainty with DPLD. Consequently, there is considerable interest in the study of the genetics, genomics, and proteomics of ILD/DPLD to better understand the pathogenesis of these diseases and individual disease susceptibility, to discover biomarkers for disease diagnosis and prognosis, as well as to develop effective treatment interventions.

Four lines of evidence suggest that the development of pulmonary fibrosis is, at least in part, determined by genetic factors. First, clustering of pulmonary fibrosis, an uncommon disease, has been reported in monozygotic twins raised in different environments (Bonanni et al., 1965; Javaheri et al., 1980; Solliday et al., 1973), in genetically related members of several families (Bitterman et al., 1986; Bonanni et al., 1965; Hughes, 1964; Swaye et al., 1969), in consecutive generations of the same families (Bonanni et al., 1965; Hodgson et al., 2002; Lee et al., 2005), and in family members separated at an early age (Swaye et al., 1969). While a single report suggests that FIP is inherited as an autosomal recessive trait (Tsukahara and Kajii, 1983), other pedigrees demonstrate an autosomal dominant pattern of inheritance (Adelman et al., 1966; McKusick and Fisher, 1958; Swaye et al., 1969), perhaps with reduced penetrance (Adelman et al., 1966; Bitterman et al., 1986; Hughes, 1964; Javaheri et al., 1980; Marshall et al., 2000; Musk et al., 1986a, b;

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

Solliday et al., 1973; Swaye et al., 1969). Second, pulmonary fibrosis is observed in genetic disorders with pleiotropic presentation, including Hermansky-Pudlak syndrome (Depinho and Kaplan, 1985), neurofibromatosis (Riccardi, 1981), tuberous sclerosis (Harris et al., 1969; Makle et al., 1970), Neimann-Pick disease (Terry et al., 1954), Gaucher disease (Schneider et al., 1977), familial hypocalciuric hypercalcemia (Auwerx et al., 1985), and familial surfactant protein C mutation (Thomas et al., 2002). Third, considerable variability exists in the development of pulmonary fibrosis among workers exposed to similar concentrations of fibrogenic dusts or organic antigens. For instance, following exposure to asbestos, similarly exposed individuals may experience very different outcomes (Polakoff et al., 1979; Selikoff et al., 1979). Fourth, inbred strains of mice differ in their susceptibility to fibrogenic agents. In comparison to BALB/c or 129 mice, C57BL/6 mice develop more lung fibrosis when challenged with either bleomycin (Ortiz et al., 1998; Rossi et al., 1987) or asbestos (Corsini et al., 1994; Warshamana et al., 2002).

This review will focus on the genetic, genomic, and proteomic approaches to identify disease susceptibility genes that predispose to DPLD or to aid in the diagnostic classification of DPLD. To better understand the pathogenesis of pulmonary sarcoidosis, both candidate gene and genome-wide linkage strategies have been utilized, as well as proteomic approaches to identify pathogenic antigens causing granulomatous inflammation. Surfactant protein C deficiency has been linked to familial cases of pediatric and adult interstitial pneumonia and represents the most successful candidate gene approach to studying genetic susceptibility to DPLD. Microarray-based expression profiling is being successfully utilized to aid in the classification of DPLDs and to identify novel disease-susceptibility genes. Recombinant inbred and congenic mouse strains have been successfully used to map susceptibility loci related to radiation-, bleomycin-, and asbestos-induced pulmonary fibrosis. Finally, genome-wide scans are being utilized to identify susceptibility loci that predispose to the development of IIPs.

GENETIC DETERMINANTS OF SARCOIDOSIS

Evidence for Genetic Basis

Sarcoidosis represents a complex disease with racial and ethnic differences in disease prevalence, an association with environmental exposures, and a genetic predisposition. A genetic basis for pulmonary sarcoidosis is suggested by familial clustering of sarcoidosis (Buck and Mc, 1961; Harrington et al., 1994; Headings et al., 1976; Moura et al., 1990; Wiman, 1972) and racial differences in disease prevalence (Rybicki et al., 2001a, b). A large multicenter case control study investigating the familial aggregation of sarcoidosis in the United States demonstrated a 5.8-fold increase in the relative risk of developing sarcoidosis among first degree relatives (Rybicki et al., 2001a, 2005a).

Candidate Gene Studies

Sarcoidosis demonstrates characteristic granulomatous inflammation with morphologic features of the granuloma having a concentration of CD4 positive T cells at the central core, CD8 positive T cells at the periphery, and a host of associated immunologic abnormalities (Statement on sarcoidosis, 1999). Immunologic abnormalities observed in patients with sarcoidosis include expansion of T cells bearing restricted T cell receptor suggesting oligoclonality; increased expression of members of the TNF-ligand and TNF-receptor superfamilies by T cells; B cell hyperactivity and spontaneous *in situ* production of immunoglobulin; and accumulation of monocytes/macrophages with antigen presenting capacity associated with increased release of macrophage-derived cytokines (IL-1, IL-6, IL-8, IL-15, TNF-α, IFN-γ, GM-CSF), chemokines (RANTES, MIP-1α, IL-16), and fibrogenic cytokines (TGF-β, PDGF). Genes involved in these pathways are plausible biologic candidate genes.

Given the known racial and ethnic differences in disease prevalence, the characteristic immunologic features of the disease, attention has been directed to HLA region genes that predispose to the development of sarcoidosis. HLA class II alleles have been most frequently reported to be associated with the risk of developing sarcoidosis. Studies performed in different populations have produced conflicting results demonstrating either increased risk or protection from various HLA alleles. There is a consistent association with HLA-DR3 haplotypes and a more favorable prognosis in Czech, German, Italian, Japanese, Polish, and Scandinavian populations (Berlin et al., 1997; Bogunia-Kubik et al., 2001; Foley et al., 2001; Gardner et al., 1984; Ina et al., 1989; Ishihara et al., 1994; Martinetti et al., 1995; Swider et al., 1999). The HLA-DRB1 and HLA-DQB1 alleles have been associated with milder forms of the disease (erythema nodosum, Löfgren's syndrome, stage 0/1 CXR findings) in patients from Scandinavia, UK, and the Netherlands (Berlin et al., 1997; Sato et al., 2002). Of note, HLA-DQB1 is in linkage disequilibrium (LD) with HLA-DRB1, which is in close proximity to non-HLA-related genes such as TNF that may be similarly influencing outcomes in sarcoidosis (Mrazek et al., 2005). The favorable outcome associated with increased TNF-α production based on the A2 promoter allele may be related to a common haplotype shared by HLA-DR3 (Seitzer et al., 2002). Other HLA loci that confer disease susceptibility include HLA-A1, -B8, -B22, -B13, -DR15, and -DR16, whereas protection from disease or milder forms of the disease have been associated with HLA-DR17 and -DRw52 (Berlin et al., 1997; Grunewald et al., 2004; Maliarik et al., 1998a; Martinetti et al., 1995; Rossman et al., 2003).

A number of non-HLA genes have been investigated for an association with sarcoidosis. The most heavily investigated is an insertion/deletion located in intron 16 of the gene for angiotensin converting enzyme (ACE), which is known to affect serum ACE levels. While this mutation affects serum ACE levels, there appears to be no relationship to disease susceptibility (Alia et al., 2005; Maliarik et al., 1998b; McGrath et al., 2001). Other non-HLA candidate genes studied in sarcoidosis and their associations are summarized in Table 48.1. In interpreting these data,

it is important to look closely at the method of choosing control populations to avoid spurious associations due to population stratification, as well as statistical genetic methods such as verification that genetic markers are in Hardy–Weinberg equilibrium, and correction for multiple comparisons. It is also important to look for evidence of linkage disequilibrium (LD) flanking putative disease susceptibility single nucleotide polymorphisms (SNPs) since unknown mutations in LD with the putative disease susceptibility SNP may be responsible for the apparent association. In general, replication studies in multiple populations are necessary before a candidate gene SNP is unequivocally linked as a sarcoidosis susceptibility gene. Currently, utilizing mutation screening in candidate genes, only HLA alleles have been validated unequivocally as susceptibility genes for sarcoidosis.

Genetic Epidemiology

Evidence that sarcoidosis aggregates in some families, and the ability to accurately identify and phenotype these families, is critical to the successful completion of family-based linkage studies. The most comprehensive study of the familial aggregation of sarcoidosis comes from A Case Control Etiologic Study of Sarcoidosis (ACCESS) (Rybicki et al., 2001a). The study population was drawn from 10,862 first-degree and 17,047 second-degree relatives identified by 706 sarcoidosis case-control pairs. Controls were matched to cases on race, sex, age, and three-digit phone numbers. The familial relative risk of sarcoidosis was 5.8 (OR = 5.8, [2.1, 15.9]) for sibs, and 3.8 (OR = 3.8, [1.2, 11.3]) for parents.

Linkage Analysis

The first published genome-wide linkage study in sarcoidosis was from 63 German families consisting mostly of affected sibling pairs having 138 affected siblings and 95 first-degree relatives (Schurmann et al., 2001). This genome-wide scan utilized 225 microsatellite markers, and was analyzed using the NPL score (Genehunter 2.0), a nonparametric, model-free approach where modes of inheritance and penetrance values are not required. The highest NPL score was 2.99, $p=0.001$, obtained with marker D6S1666, which resides in the MHC class II gene. The genome-wide significant NPL score is approximately 3.6, $p<0.001$. Therefore, the NPL score of 2.99 is very suggestive but not conclusive evidence for linkage to the MHC class II gene. Other chromosome regions showing minor peaks ($p<0.05$) include NPL scores of 2.39 on chromosome 3p21, 1.87 on chromosome 1p22, 1.82 for chromosome 9q33, 1.64 on chromosome X, and 1.92 on chromosomes 7q22 and 7q36. The second linkage study in sarcoidosis is from the Sarcoidosis Genetic Analysis Consortium (SAGA) reporting linkage analysis of sibling pairs from 229 African American families utilizing 380 microsatellite markers. These investigators reported *p*-values based on Haseman-Elston regression. The smallest *p*-value was 0.0005 on chromosome 5 at marker D5S2500. Interestingly, despite using a higher density of markers in the MHC class II region, the SAGA investigators did not find evidence for linkage in the MHC class II region. The different linkage results in the German and African American populations would be consistent with ethnicity-related locus heterogeneity. Also, the sample sizes in these two studies are relatively small for a complex disease, and may account for the different results. Nevertheless, these studies indicate that genome-wide linkage studies can be utilized to identify disease susceptibility genes in sarcoidosis.

Based on the initial linkage analysis of the 63 German families demonstrating linkage to chromosome 6p21, SNP-based fine mapping of the region was performed using extended families and trios to conduct transmission disequilibrium test (TDT), and case-control association analysis (Valentonyte et al., 2005). The results demonstrated an association with SNP rs2076530 ($P_{\mathrm{TDT}} = 3 \times 10^{-6}$, $P_{\mathrm{case-control}} = 1.1 \times 10^{-8}$) located in the butyrophilin-like2 gene (BTNL2) located adjacent to HLA-DR1. Additional analysis demonstrated G to A transition in BTNL2 that leads to the use of a cryptic slice site resulting in a 4bp

TABLE 48.1 Non-HLA candidate gene polymorphisms in sarcoidosis

Candidate gene	Polymorphism	Result	Reference
ACE	Intron 16 in/del	Population specific	Berlin et al. (1997); McGrath et al. (2001)
Vitamin D receptor	BsmI RFLP	Increased risk	Niimi et al. (2000b)
IL-1 cluster	IL-α-889	Increase risk 2X	Niimi et al. (2000a)
TGF β3	4785A	Fibrotic sarcoid	Kruit et al. (2006)
HSP-70 hom	2763, 2437	Löfgren's syndrome	Bogunia-Kubik et al. (2006)
BTNL2	10 intron/exon 5 SNP	Increased risk	Rybicki et al. (2005b)
TLR4	A299G, T399I	Increased in chronic sarcoid	Pabst et al. (2006)
Nod2/Card15	R702W, G908R, 1007FsinC	No effect	Kanazawa et al. (2005); Martin et al. (2003)
TNFα	A2 promoter allele	Favorable prognosis	Sabounchi-Schutt et al. (2003)
CCR5	HHC haplotype	Persistent lung disease	Spagnolo et al. (2005)
BTNL2	3 locus haplotype	Increased risk	Akahoshi et al. (2004)
HSP70-hom	C2437T	Lofgren's syndrome	Zorzetto et al. (2002)
TGFβ	4875A	Fibrotic sarcoid	Takada et al. (2001)

deletion that generates a premature stop codon, and the resulting protein lacks the C-terminal immunoglobulin-like constant domain and transmembrane domain. The authors demonstrate that the disease allele encodes for a protein that is located in the cytoplasm rather than on the plasma membrane. The odds ratio (OR) for developing sarcoidosis when heterozygous for the susceptibility allele is 1.6, and 2.75 in homozygotes. The authors did not comment on the overall frequency of the susceptibility allele in their population. Given the complexity of sarcoidosis, and the modest OR in the range of 1.6–2.75, one would expect that BTNL2 is one of the several sarcoid susceptibility genes, and further investigations of BTNL2 in other non-German populations will be important. Additional studies of the role of the BTNL2 gene in sarcoidosis may identify novel mechanisms predisposing individuals to sarcoidosis, or other granulomatous inflammation.

Genomic Medicine and Sarcoidosis

The identification of BTNL2 mutations, or mutations in other genes that lead to the susceptibility of sarcoidosis, will likely aid several aspects of the clinical approach to patients with sarcoidosis. A very clinically relevant application of sarcoid susceptibility alleles would be the development of a diagnostic test that distinguishes chronic progressive forms of sarcoidosis from spontaneous remitting disease. Another useful risk allele would be one that identifies risk for extra-pulmonary sarcoidosis such as cardiac or central nervous system involvement since involvement of these organs results in substantial morbidity and mortality. The identification of risk alleles that distinguish progressive forms of the disease, or forms of the disease associated with higher morbidity would identify a group of patients needing aggressive anti-inflammatory therapy while sparing those with milder, self-limited disease the toxicity of these therapies. In addition, the identification of sarcoid susceptibility genes might provide insight into development of new pharmacologic therapies.

Proteomics in Sarcoidosis

Sarcoidosis is characterized by granulomatous inflammation with expansion of T cells bearing restricted T cell receptor suggesting oligoclonality, and polyclonal antibody production. The intradermal injection of the Kveim reagent, homogenates of diseased sarcoid tissue obtained from spleen, or lymph nodes, induces characteristic granulomatous inflammation known as the Kveim reaction (Siltzbach, 1961). The biochemical features of the Kveim reagent include neutral detergent insolubility; heat, acid, and protease resistance; and sensitivity to potent denaturants, suggesting a poorly soluble protein or protein aggregate (Chase and Siltzbach, 1967; Lyons et al., 1992). Proteomic approaches have been applied to better characterize the inflammatory response of sarcoidosis, and to identify potential antigenic proteins present in the Kveim reagent.

Several studies have analyzed sarcoid bronchoalveolar lavage fluid (BALF) protein by 2-D gel electrophoresis and mass spectroscopy. These studies demonstrate increased concentrations of many plasma proteins in sarcoid BALF (Magi et al., 2002; Rottoli et al., 2005; Sabounchi-Schutt et al., 2003). The largest study to date compared alveolar proteins in patients with acute forms of sarcoidosis with chest radiographs having only bilateral hilar adenopathy to those with more chronic forms of sarcoidosis (Kriegova et al., 2006). Forty differentially expressed proteins were seen and some were identified as albumin, alpha-1-antitrypsin, and protocadherin-2 precursor. These data demonstrate that BALF analyzed by SELDI-TOF or MALDI-TOF mass spectroscopy can identify unique disease-related protein profiles, and that these protein profiles may change with the stage of the disease or disease progression.

Epidemiologic studies suggest an association between infective agents and sarcoidosis based on seasonal variations and case clustering (Baughman et al., 2003). Several lines of evidence suggest mycobacteria as a candidate organism. First, while mycobacteria in general cannot be isolated from sarcoid tissue using conventional microbiologic techniques, cell wall deficient L forms persisting as an intracellular organism can be identified. Second, several PCR-based studies have identified mycobacterium DNA in sarcoid tissue homogenates (Song et al., 2005). Utilizing biochemical techniques similar to that used to isolate the Kveim reagent and immunoblotting using immunoglobulin obtained from the serum of sarcoid patients, investigators identified immunoreactive proteins unique to sarcoid tissue homogenates. These antigenic bands were excised from gels, subjected to trypsin digestion, and analyzed by MALDI-TOF mass spectroscopy. The peptide fingerprints were identified as mycobacterium tuberculosis catalase–peroxidase. These authors hypothesize that insoluble aggregates of mycobacteria-derived catalase–peroxidase drive the sarcoid immune response in genetically predisposed individuals.

SURFACTANT PROTEINS AND DPLD

Pulmonary surfactant

Pulmonary surfactant is a complex mixture of phospholipids and proteins that functions to reduce surface tension at the alveolar air interface preventing atelectasis. Deficiency of pulmonary surfactant is the principal cause of respiratory distress syndrome in premature infants (Whitsett and Weaver, 2002). Four surfactant-associated proteins, surfactant proteins A, B, C, and D, have been described, and two have been associated with DPLD. Surfactant protein C (SP-C) is a highly hydrophobic protein that enhances the surface tension lowering properties of pulmonary surfactant. Familial cases of neonatal respiratory distress have been associated with surfactant protein B deficiency, but respiratory distress of neonates is not considered a form of DPLD/ILD (Nogee et al., 2000). Genetic variants of SP-A have been associated with increased risk of idiopathic pulmonary fibrosis (IPF; Lawson et al., 2004).

Family Studies with SP-C

Nogee et al. (2001) reported a full-term baby girl born to a woman who had desquamative interstitial pneumonia, a type of IIP, at 1-year of age. The infant's maternal grandfather died of an

unknown lung disease. The infant developed respiratory distress at the age of 6 weeks, and surgical lung biopsy demonstrated non-specific interstitial pneumonia (NSIP). Both the infant and the mother had minimal SP-C by either immunohistochemical staining, or immunoblotting of lung tissue. DNA sequence analysis of the SP-C gene demonstrated a heterozygous substitution of A to G at the first base of intron 4 that abolished the normal donor splice site resulting in a truncated mRNA. Subsequently, there have been several other similar families with SP-C mutations and interstitial pneumonia (Chibbar et al., 2004; Thomas et al., 2002). In the largest kindred, a heterozygous T to A substitution was identified in exon 5 causing glutamine for threonine substitution. In this pedigree, there was both adult-onset interstitial pneumonia of the usual interstitial pneumonia (UIP) histology and children with the cellular NSIP (Thomas et al., 2002). Immunohistochemical analysis of these patients demonstrated intracellular aggregates of SP-C and *in vitro* expression studies demonstrated abnormal intracellular processing of SP-C in alveolar type II cells.

SP-B and SP-C Genetic Variants Associated with IPF

In a single study by Selman and colleagues, the SP-A1 $6A^4$ haplotype is associated with a substitution of 3 amino acids at positions 19, 50, and 219, with the 219 variant being a tryptophan for arginine substitution (Lawson et al., 2004; Selman et al., 2003). The amino acid 219 variant is associated with IPF in smokers and nonsmokers (OR = 3.67 [1.34, 10.07], $p = 0.01$).

GENETIC DETERMINANTS OF PULMONARY FIBROSIS IDENTIFIED IN RARE INHERITED DISORDERS

Pulmonary fibrosis is observed in genetic disorders with pleiotropic presentation, including Hermansky-Pudlak syndrome, neurofibromatosis, tuberous sclerosis, Neimann-Pick disease, Gaucher disease, familial hypocalciuric hypercalcemia, familial SP-C mutation, and most recently in dyskeratosis congenital. Recently, mutations associated with the dyskeratosis congenita syndrome have been in a small percentage of FIP (Armanios et al., 2007). Specifically, 8% of 73 families with more than one case of IIP were found to have heterozygous mutations in telomerase reverse transcriptase, resulting in shortening of telomeres. These authors suggest that telomere shortening may cause apoptosis of the alveolar epithelium.

Emerging Concepts from Genomic Studies of Associated DPLD

An interesting feature of these studies is the variable histopathologic features among family members sharing the identical SP-C mutation, suggesting modification of disease phenotype by other unknown factors. Mutations in SP-C and telomerase reverse transcriptase also suggest that abnormalities of the alveolar epithelium, particularly alveolar type II epithelial cells, the major source of pulmonary surfactant and also the progenitor cell for alveolar type I epithelial cell, are critical for the development of pulmonary fibrosis and interstitial pneumonia. In the instance of SP-C mutations, intracellular aggregates of abnormally processed SP-C may induce functional abnormalities of alveolar epithelial cells. These studies have resulted in a paradigm shift in the mechanism and the pathogenesis of IIP, away from an inflammatory hypothesis, toward one of abnormal injury and repair of the alveolar epithelium and the fibroproliferative response. These data support the lack of efficacy of glucocorticoid and cytotoxic treatments of IPF.

GENETIC DETERMINANTS OF FIP

Clinical Features

The interstitial pneumonias can be classified as sporadic or idiopathic (IIP), in which no positive family history can be identified, or FIP. There are no data on the relative proportion of interstitial pneumonias that are familial, but estimates are in the range of 5%. Familial aggregation has been reported in a variety of studies in twins, siblings raised apart, and multigenerational families. Steele et al. (2005) reported the largest collection of families, identifying 111 families from the United States, and 20 multigenerational pedigrees were consistent with autosomal dominant inheritance. Forty-five percent of the families demonstrated phenotypic heterogeneity with some families having bronchiolitis obliterans or NSIP and UIP within the same pedigree. Cigarette smoking was associated with affection status among siblings (OR = 3.6, [1.3, 9.8], $p = 0.01$). The histopathologic heterogeneity in these families has been confirmed in a subsequent study with multiple independent pathologists. UIP is found in 40% of these families, but the predominant histopathologic pattern (60%) for the FIP cohort was difficult to classify pattern characterized by advanced fibrosis with a high incidence of microscopic honeycombing, smooth muscle proliferation and fibrosis, and variable diffuse alveolar septal and airway-centered scarring. Similar to the findings with SP-C mutations, these data suggest that histologically distinct forms of pulmonary fibrosis may have common pathogenic mechanisms, and that cigarette smoking may contribute to the development of pulmonary fibrosis in individuals who are genetically prone to this disease.

Candidate Genes Identified by Microarray Analysis in Familial Pulmonary Fibrosis

Yang and colleagues performed microarray analysis of 16 cases of sporadic IIP (14 UIP, 2 NSIP), and 10 cases of familial IIP (6 UIP, 4 NSIP), compared to 9 normal lung controls. RNA was extracted from diseased lung tissue (specimens from surgical lung biopsy, autopsy, or explanted lung at the time of lung transplant), or normal controls (9 lung samples taken from donor lung at the time of lung transplant). cRNA was synthesized and hybridized

onto whole human genome arrays modified with an additional 657 probes for genes/ESTs that would be potentially informative based on preliminary linkage data (Yang et al., 2005). Expression profiling was performed using standard protocols (co-hybridized with human universal cell line reference RNA; replicates with the two dyes swapped; Lowess-normalized intensities; analysis using the TGR MIDAS and MeV software), and differentially expressed genes were identified using significance analysis of microarrays (SAM) with 100 permutations. 558 differentially expressed transcripts between cases compared to controls were identified, with 135 genes being up- or down-regulated greater than 1.8-fold. When hierarchical clustering was applied to the set of 135 genes, all but two samples clustered according to disease versus no disease, and familial disease segregated from sporadic disease (Figure 48.1). Sixty-nine differentially expressed genes were identified that distinguish sporadic and familial IP, and these

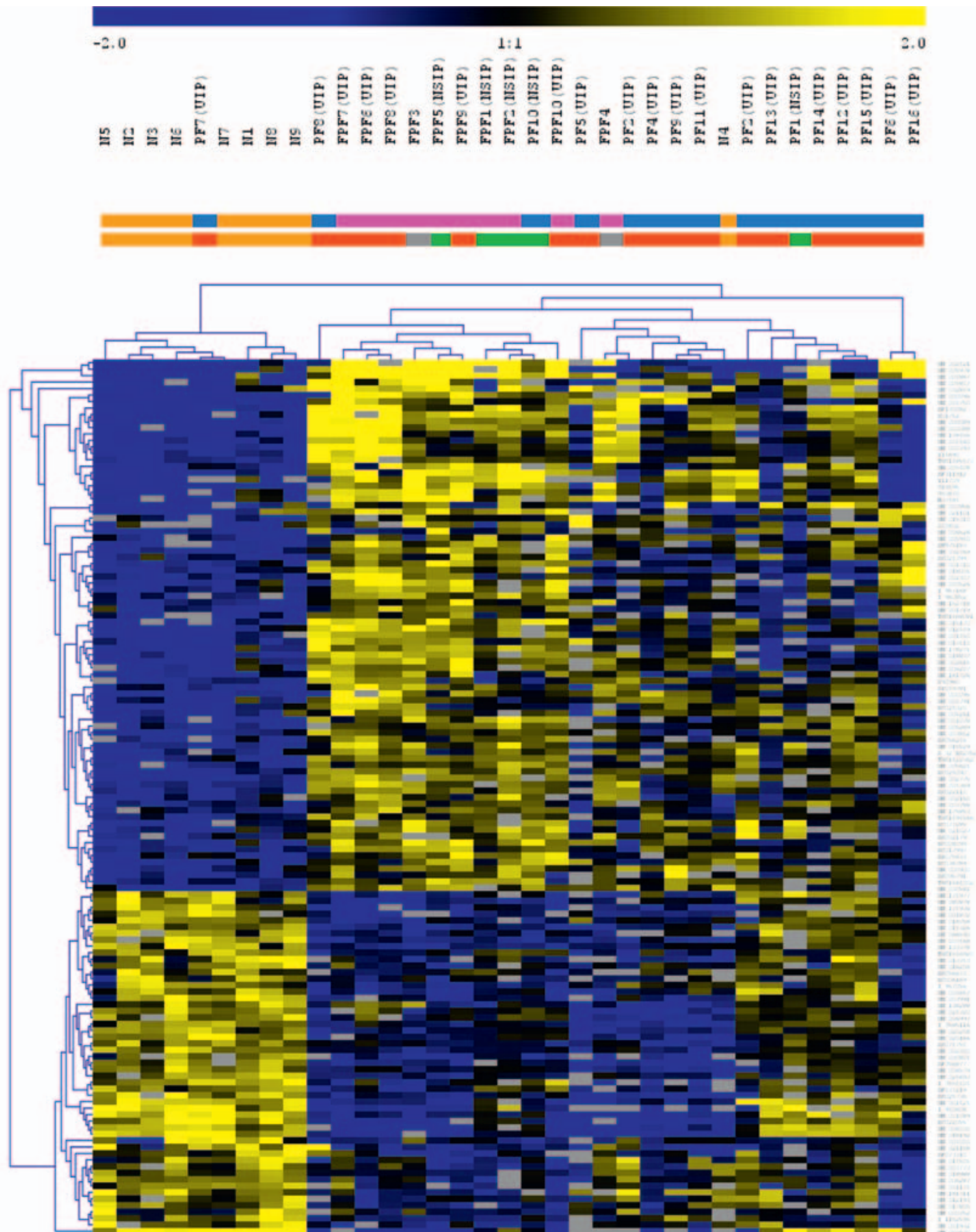


Figure 48.1 Hierarchical clustering of 35 cases and 16 controls. Hierarchical clustering of 35 samples based on 135 transcripts that best differentiate patients with pulmonary fibrosis from normal controls. Samples are color-coded according to the disease inheritance status (orange – normal, blue – sporadic IIP, magenta – familial IIP) or histological features (orange – normal, green – NSIP, red – UIP, gray – no histological evaluation). Average linkage clustering with Euclidian distance metric was used.

are broadly grouped into functional classes (Figure 48.2), with a wide variety of chemokines, extracellular matrix, and growth-related genes that are differentially expressed. These data indicate that familial and sporadic IIP are transcriptionally distinct and also suggest many similarities between the histologic subtypes of UIP and NSIP. This study indicates that additional candidate genes that may be important in the pathogenesis of IIP, or that might be useful as biomarkers of disease activity, can be identified using whole-genome microarray analysis.

Relevance of Genomics and Microarray Studies to DPLD/ILD

A number of clinical challenges exist that impede better understanding of disease mechanisms in DPLD and are potentially addressed using expression profiling, proteomic, and genomic approaches. First, even in the hands of expert clinicians, radiologists, and pathologists specializing in DPLD, a definitive consensus diagnosis is achievable in about 85% of patients due to the clinical, radiologic, and pathologic overlap of these

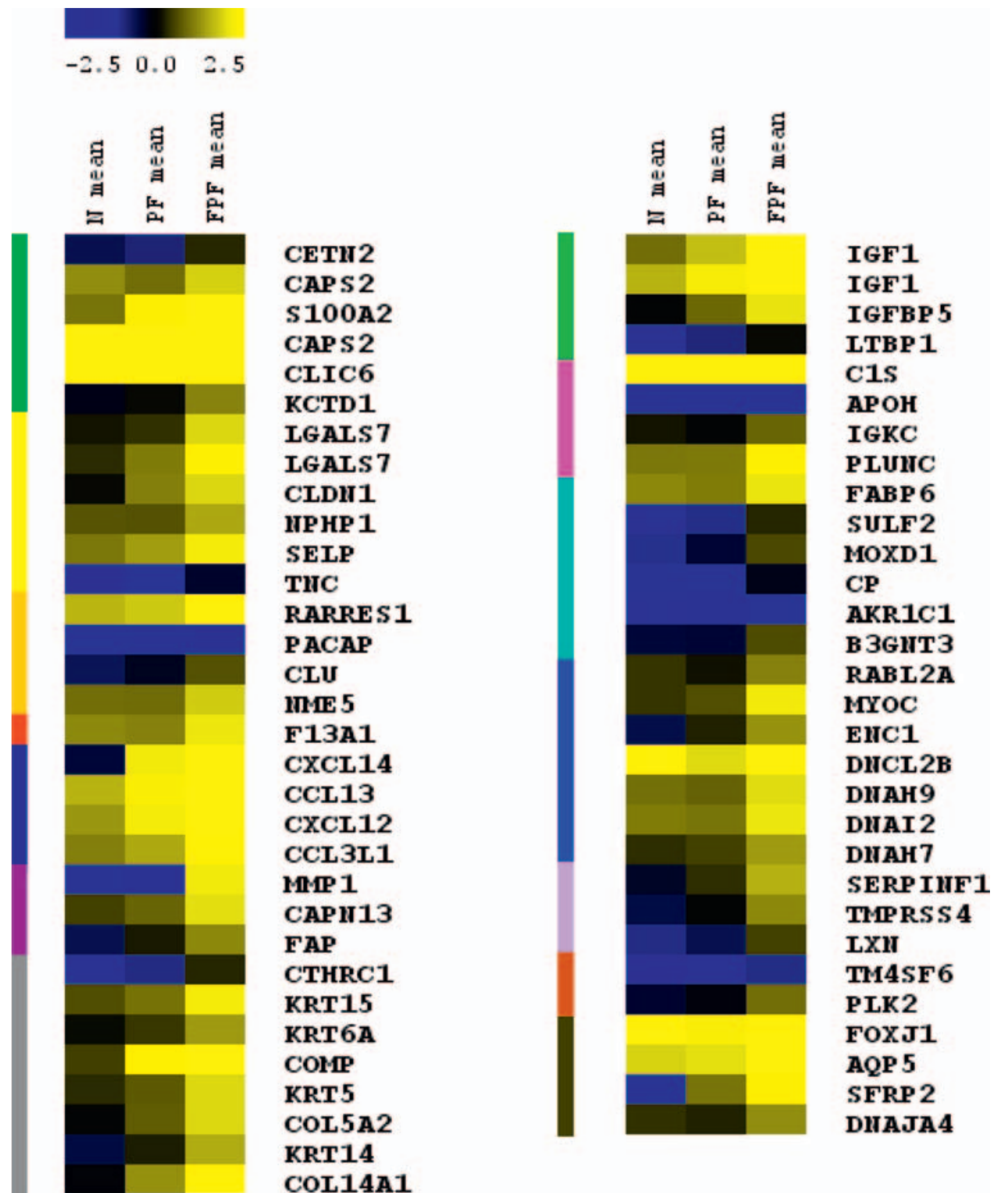


Figure 48.2 Mean Expression Ratios in normal, sporadic IIP, and familial IIP. Groups of 69 genes with known function that best differentiate familial IIP from sporadic IIP. Genes are arranged and color-coded according to their function: dark green – calcium/potassium ion binding and transport; yellow – cell adhesion; orange – cell proliferation and death; red – coagulation cascade; blue – cytokines, chemokines, and their receptors; purple – ECM degradation; gray – ECM structural component; bright green – growth factors and their receptors; magenta – immune response; turquoise – metabolism; navy – motor activity; lavender – proteases and inhibitors; brown – regulation of I-kappaB/NF-kappaB cascade; and black – other.

entities. Microarray-based expression profiling has the capability to improve diagnostic accuracy. A recent study by Selman and colleagues demonstrates that a distinct microarray expression signature exists for hypersensitivity pneumonitis, a clinical entity with well-recognized radiologic and pathologic overlap with IIP, when compared to IPF (Selman et al., 2006).

Linkage Studies in Pulmonary Fibrosis

The only published study performing genome-wide linkage analysis in FIP comes from Finland (Hodgson et al., 2006). Using 6 pedigrees, NPL scores of 1.7 on chromosome 3 (marker D3S1278) and chromosome 4 (marker D4S424), and NPL score of 1.6 on chromosome 13 (D13S265) were obtained. On chromosome 4, a shared haplotype was identified among 8 of 24 multiplex families. A candidate gene located in the region of interest, ELMOD2, was further investigated by resequencing of exons and exon/intron boundaries. No mutations in ELMOD2 in these locations were identified. RT-PCR and *in situ* hybridization demonstrated decreased levels of ELMOD2 mRNA in six cases of sporadic IPF compared to controls. Genome-wide linkage analysis of the 111 US families is in progress, and preliminary results indicate linkage (LOD score >3.0) on chromosome 11p15 not identified in the Finland study (unpublished data).

2009 UPDATE

IPF is the most common of the IIPs and continues to be a devastating and poorly understood disease. While IIPs most often occur as sporadic cases, familial forms represent about 5–10% of IPF cases. Genetic studies of both familial and sporadic IPF are providing new insights into potential disease mechanisms. Recently, both sporadic and familial IPF have been associated with telomere shortening and mutations in telomerase genes.

Two groups independently reported an association with mutations in TERT, the reverse transcriptase component of telomerase, in both familial and sporadic IPF (Armanios et al., 2007; Tsakiri et al., 2007). Variants in TERTC, the gene encoding the RNA component of telomerase, have also been identified. These studies were from a relatively small number of cases in the United States. Recently, abnormalities in TERT have been associated with sporadic IPF from a genome-wide association study (GWAS) in larger cohort of patients from Japan. A GWAS using 159 patients with IPF and 934 controls tested 214,508 tagged SNPs and showed that SNP rs2736100, located in intron 2 of TERT, had a p value of 2.8×10^{-6}. In a replication set of 83 IPF patients and 535 controls, a p value of 3.6×10^{-3} was obtained (Mushiroda et al, 2008). This TERT intron 2 SNP is located within a LD block (D' and $r^2 > 0.8$) within TERT, suggesting that rs2736100 may be associated with some other unidentified variation within TERT in this population of IPF patients from Japan.

Novel missense mutations were identified in TERT in patients from the United States with dyskeratosis congenita who had pulmonary fibrosis (Armanios et al., 2007). These investigators also sequenced TERT in 73 kindreds with familial IPF and identified 8% of probands heterozygous for TERT mutations and one proband heterozygous for a mutation in TERTC. Telomere length in these affected individuals was reduced to below the 10th percentile in lymphocytes. Whole-genome linkage analysis was performed in two large families with pulmonary fibrosis that showed evidence for linkage to chromosome 5 in a region containing TERT, and variants in TERT were found in these families as well (Taskiri et al., 2007).

Recently, the possible role of telomere shortening has been generalized to sporadic and familial IPF patients without identified mutations in TERT/TERTC (Cronkhite et al., 2008). Using a modified Southern blot technique, restriction fragment length methods and a quantitative polymerase chain reaction method to measure telomere length of genomic DNA, these investigators found telomere length less than the 10th percentile ($p = 2.6 \times 10^{-8}$) in a significantly higher proportion of probands with familial pulmonary fibrosis (24%) and sporadic cases (23%) in which no coding mutations in TERT or TERTC had been found. Pulmonary affected status was significantly associated with telomerase restriction fragment lengths even after controlling for age, sex and ethnicity ($p = 6.1 \times 10^{-11}$). Overall, 25% of sporadic IPF and 37% of familial IPF had telomere lengths below the 10th percentile. These studies were performed in circulating lymphocytes, and it will be important to replicate these results in future studies in lung cells, particularly in the alveolar epithelium.

To date, studies of the genetics of pulmonary fibrosis have identified mutations in surfactant proteins and more recently in telomerase, as susceptibility alleles for the development of pulmonary fibrosis. Multiple lines of evidence – including GWAS in sporadic IPF, family linkage studies and measurements of telomere length in sporadic IPF – all suggest a role for telomere shortening and susceptibility to pulmonary fibrosis. The association of the risk of developing pulmonary fibrosis and telomere shortening may explain in part the late onset of these diseases. In addition, there are environmental influences in pulmonary fibrosis, such as cigarette smoking, that are known to cause telomere shortening, thus providing a potential gene–environment connection. These data also provide additional support for the hypothesis that apoptosis of the alveolar epithelium predisposes to the development of pulmonary fibrosis.

2009 UPDATE REFERENCES

Armanios, M., Chen, J.J., Cogan, J.D., Alder, J.K., Ingersoll RG, Markin C, et al. (2007). Telomerase mutations in families with idiopathic pulmonary fibrosis. *N Engl J Med* 356, 1317–1326.

Cronkhite, J.T., Xing, C., Raghu, G., Chin, K.M., Torres, F., Rosenblatt, R.L. and Garcia, C.K. (2008). Telomere shortening in familial and sporadic pulmonary fibrosis. *Am J Respir Crit Care Med* 178(7), 729–737.

Mushiroda, T., Wattanapokayakit, S., Takahasi, A., Nukiwa, T., Kudoh, S., Ogura, T., Taniguchi, H., Kubo, M., Kamatani, K., Nakamura, Y. and the Pirfenidone Clinical Study Group. *J Med Genet* 45, 654–656.

Tsakiri, K.D., Cronkhite, J.T., Kuan, P.J., Xing, C., Raghy, G., Weissler, J.C. et al. (2007). Adult-onset pulmonary fibrosis caused by mutations in telomerase. *Proc Natl Acad Sci U S A* 104, 7552–7557.

REFERENCES

Adelman, A.G., Chertkow, G. and Hayton, R.C. (1966). Familial fibrocystic pulmonary dysplasia: A detailed family study. *Can Med Assoc J* 95, 603–610.

Akahoshi, M., Ishihara, M., Remus, N., Uno, K., Miyake, K., Hirota, T., Nakashima, K., Matsuda, A., Kanda, M., Enomoto, T. et al. (2004). Association between IFNA genotype and the risk of sarcoidosis. *Hum Genet* 114, 503–509.

Alia, P., Mana, J., Capdevila, O., Alvarez, A. and Navarro, M.A. (2005). Association between ACE gene I/D polymorphism and clinical presentation and prognosis of sarcoidosis. *Scand J Clin Lab Invest* 65, 691–697.

Armaniaos, M.Y. et al. (2007). Telomerase mutations in families with idiopathic pulmonary fibrosis. *N Engl J Med* 356(13), 1317–1326.

Auwerx, J., Boogaerts, M., Ceuppens, J.L. and Demedts, M. (1985). Defective host defence mechanisms in a family with hypocalciuric hypercalcaemia and coexisting interstitial lung disease. *Clin Exp Immunol* 62, 57–64.

Baughman, R.P., Lower Elyse, E. and Du Bois, R.M. (2003). Sarcoidosis. *Lancet* 361, 1111–1118.

Berlin, M., Fogdell-Hahn, A., Olerup, O., Eklund, A. and Grunewald, J. (1997). HLA-DR predicts the prognosis in Scandinavian patients with pulmonary sarcoidosis. *Am J Respir Crit Care Med* 156, 1601–1605.

Bitterman, P.B., Rennard, S.I., Keogh, B.A., Wewers, M.D., Adelberg, S. and Crystal, R.G. (1986). Familial idiopathic pulmonary fibrosis: Evidence of lung inflammation in unaffected members. *N Engl J Med* 314, 1343–1347.

Bogunia-Kubik, K., Tomeczko, J., Suchnicki, K. and Lange, A. (2001). HLA-DRB1*03, DRB1*11 or DRB1*12 and their respective DRB3 specificities in clinical variants of sarcoidosis. *Tissue Antigens* 57, 87–90.

Bogunia-Kubik, K., Koscinska, K., Suchnicki, K. and Lange, A. (2006). HSP70-hom gene single nucleotide (+2763 G/A and +2437 C/T) polymorphisms in sarcoidosis. *Int J Immunogenet* 33, 135–140.

Bonanni, P.P., Frymoyer, J.W. and Jacox, R.F. (1965). A family study of idiopathic pulmonary fibrosis. A possible dysproteinemic and genetically determined disease. *Am J Med* 39, 411–421.

Buck, A.A. and Mc, K.V. (1961). Epidemiologic investigations of sarcoidosis. III. Serum proteins; syphilis; association with tuberculosis: familial aggregation. *Am J Hyg* 74, 174–188.

Chase, M. and Siltzbach, L.E. (1967). Concentration of the active principle responsible for the Kveim reaction. *La Sarcoidose* 196, 150–160.

Chibbar, R., Shih, F., Baga, M., Torlakovic, E., Ramlall, K., Skomro, R., Cockcroft, D.W. and Lemire, E.G. (2004). Nonspecific interstitial pneumonia and usual interstitial pneumonia with mutation in surfactant protein C in familial pulmonary fibrosis. *Mod Pathol* 17, 973–980.

Corsini, E., Luster, M.I., Mahler, J., Craig, W.A., Blazka, M.E. and Rosenthal, G.J. (1994). A protective role for T lymphocytes in asbestos-induced pulmonary inflammation and collagen deposition. *Am J Respir Cell Mol Biol* 11, 531–539.

Depinho, R.A. and Kaplan, K.L. (1985). The Hermansky-Pudlak syndrome: Report of three cases and review of pathophysiology and management considerations. *Medicine* 64, 192–202.

Foley, P.J., McGrath, D.S., Puscinska, E., Petrek, M., Kolek, V., Drabek, J., Lympany, P.A., Pantelidis, P., Welsh, K.I., Zielinski, J. et al. (2001). Human leukocyte antigen-DRB1 position 11 residues are a common protective marker for sarcoidosis. *Am J Respir Cell Mol Biol* 25, 272–277.

Gardner, J., Kennedy, H.G., Hamblin, A. and Jones, E. (1984). HLA associations in sarcoidosis: A study of two ethnic groups. *Thorax* 39, 19–22.

Grunewald, J., Eklund, A. and Olerup, O. (2004). Human leukocyte antigen class I alleles and the disease course in sarcoidosis patients. *Am J Respir Crit Care Med* 169, 696–702.

Harrington, D., Major, M., Rybicki, B., Popovich, J., Jr., Maliarik, M. and Iannuzzi, M.C. (1994). Familial analysis of 91 families. *Sarcoidosis* 11, 240–243.

Harris, J., Waltuck, B. and Swenson, E. (1969). The pathophysiology of the lungs in tuberous sclerosis. A case report and literature review. *Am Rev Respir Dis* 100, 379–387.

Headings, V.E., Weston, D., Young, R.C., Jr. and hackney, R.L., Jr. (1976). Familial sarcoidosis with multiple occurrences in eleven families: A possible mechanism of inheritance. *Ann NY Acad Sci* 278, 377–385.

Hodgson, U., Laitinen, T. and Tukiainen, P. (2002). Nationwide prevalence of sporadic and familial idiopathic pulmonary fibrosis: Evidence of founder effect among multiplex families in Finland. *Thorax* 57, 338–342.

Hodgson, U., Pulkkinen, V., Dixon, M., Peyrard-Janvid, M., Rehn, M., Lahermo, P., Ollikainen, V., Salmenkivi, K., Kinnula, V., Kere, J. et al. (2006). ELMOD2 is a candidate gene for familial idiopathic pulmonary fibrosis. *Am J Hum Genet* 79, 149–154.

Hughes, E.W. (1964). Familial interstitial pulmonary fibrosis. *Thorax* 19, 515–525.

Ina, Y., Takada, K., Yamamoto, M., Morishita, M., Senda, Y. and Torii, Y. (1989). HLA and sarcoidosis in the Japanese. *Chest* 95, 1257–1261.

Ishihara, M., Ohno, S., Ishida, T., Ando, H., Naruse, T., Nose, Y. and Inoko, H. (1994). Molecular genetic studies of HLA class II alleles in sarcoidosis. *Tissue Antigens* 43, 238–241.

Javaheri, S., Lederer, D.H., Pella, J.A., Mark, G.J. and Levine, B.W. (1980). Idiopathic pulmonary fibrosis in monozygotic twins. The importance of genetic predisposition. *Chest* 78, 591–594.

Kanazawa, N., Okafuji, I., Kambe, N., Nishikomori, R., Nakata-Hizume, M., Nagai, S., Fuji, A., Yuasa, T., Manki, A., Sakurai, Y.

et al. (2005). Early-onset sarcoidosis and CARD15 mutations with constitutive nuclear factor-kappaB activation: Common genetic etiology with Blau syndrome. *Blood* 105, 1195–1197.

Kriegova, E., Melle, C., Kolek, V., Hutyrova, B., Mrazek, F., Bleul, A., du Bois, R.M., von Eggeling, F. and Petrek, M. (2006). Protein profiles of bronchoalveolar lavage fluid from patients with pulmonary sarcoidosis. *Am J Respir Crit Care Med* 173, 1145–1154.

Kruit, A., Grutters, J.C., Ruven, H.J., van Moorsel, C.H., Weiskirchen, R., Mengsteab, S. and van den Bosch, J.M. (2006). Transforming growth factor-beta gene polymorphisms in sarcoidosis patients with and without fibrosis. *Chest* 129, 1584–1591.

Lawson, W.E., Grant, S.W., Ambrosini, V., Womble, K.E., Dawson, E.P., Lane, K.B., Markin, C., Renzoni, E., Lympany, P., Thomas, A.Q. et al. (2004). Genetic mutations in surfactant protein C are a rare cause of sporadic cases of IPF. *Thorax* 59, 977–980.

Lee, H.L., Ryu, J.H., Wittmer, M.H., Hartman, T.E., Lymp, J.F., Tazelaar, H.D. and Limper, A.H. (2005). Familial idiopathic pulmonary fibrosis: Clinical features and outcome. *Chest* 127, 2034–2041.

Lyons, D.J., Donald, S., Mitchell, D.N. and Asherson, G.L. (1992). Chemical inactivation of the Kveim reagent. *Respiration* 59, 22–26.

Magi, B., Bini, L., Perari, M.G., Fossi, A., Sanchez, J.C., Hochstrasser, D., Paesano, S., Raggiaschi, R., Santucci, A., Pallini, V. et al. (2002). Bronchoalveolar lavage fluid protein composition in patients with sarcoidosis and idiopathic pulmonary fibrosis: A two-dimensional electrophoretic study. *Electrophoresis* 23, 3434–3444.

Makle, S.K., Pardee, N. and Martin, C.J. (1970). Involvement of the lung in tuberous sclerosis. *Chest* 58, 538–540.

Maliarik, M., Chen, K., Major, M., Sheffer, R., Popovich, J.J., Rybicki, B. and Iannuzzi, M. (1998a). Analysis of HLA-DPB1 polymorphisms in African-Americans with sarcoidosis. *Am J Respir Crit Care Med* 158, 111–114.

Maliarik, M.J., Rybicki, B.A., Malvitz, E., Sheffer, R.G., Major, M., Popovich, J., Jr. and Iannuzzi, M.C. (1998b). Angiotensin-converting enzyme gene polymorphism and risk of sarcoidosis. *Am J Respir Crit Care Med* 158, 1566–1570.

Marshall, R., Puddicombe, A., Cookson, W. and Laurent, G. (2000). Adult familial cryptogenic fibrosing alveolitis in the United Kingdom. *Thorax* 55, 143–146.

Martin, T.M., Doyle, T.M., Smith, J.R., Dinulescu, D., Rust, K. and Rosenbaum, J.T. (2003). Uveitis in patients with sarcoidosis is not associated with mutations in NOD2 (CARD15). *Am J Ophthalmol* 136, 933–935.

Martinetti, M., Tinelli, C., Kolek, V., Cuccia, M., Salvaneschi, L., Pasturenzi, L., Semenzato, G., Cipriani, A., Bartova, A. and Luisetti, M. (1995). "The sarcoidosis map": A joint survey of clinical and immunogenetic findings in two European countries. *Am J Respir Crit Care Med* 152, 557–564.

McGrath, D.S., Foley, P.J., Petrek, M., Izakovicova-Holla, L., Kolek, V., Veeraraghavan, S., Lympany, P.A., Pantelidis, P., Vasku, A., Wells, A.U. et al. (2001). Ace gene I/D polymorphism and sarcoidosis pulmonary disease severity. *Am J Respir Crit Care Med* 164, 197–201.

McKusick, V.A. and Fisher, A.M. (1958). Congenial cystic disease of the lung with progressive pulmonary fibrosis and carcinomatosis. *Ann Inter Med* 48, 774–790.

Moura, M., Carre, P., Larios-Ramos, L., Didier, A. and Leophonte, P. (1990). [Sarcoidosis and heredity. 3 familial cases]. *Rev Pneumol Clin* 46, 28–30.

Mrazek, F., Holla, L.I., Hutyrova, B., Znojil, V., Vasku, A., Kolek, V., Welsh, K.I., Vacha, J., du Bois, R.M. and Petrek, M. (2005). Association of tumour necrosis factor-alpha, lymphotoxin-alpha and HLA-DRB1 gene polymorphisms with Lofgren's syndrome in Czech patients with sarcoidosis. *Tissue Antigens* 65, 163–171.

Musk, A., Zilko, P., Manners, P., Kay, P. and Kamboh, M. (1986a). Genetic studies in familial fibrosing alveolitis. Possible linkage with immunoglobulin allotypes (gm). *Chest* 89, 206–210.

Musk, A.W., Zilko, P.J., Manners, P., Kay, P.H. and Kamboh, M.I. (1986b). Genetic studies in familial fibrosing alveolitis. Possible linkage with immunoglobulin allotypes (Gm). *Chest* 89, 206–210.

Niimi, T., Sato, S., Tomita, H., Yamada, Y., Akita, K., Maeda, H., Kawaguchi, H., Sugiura, Y. and Ueda, R. (2000a). Lack of association with interleukin 1 receptor antagonist and interleukin-1beta gene polymorphisms in sarcoidosis patients. *Respir Med* 94, 1038–1042.

Niimi, T., Tomita, H., Sato, S., Akita, K., Maeda, H., Kawaguchi, H., Mori, T., Sugiura, Y., Yoshinouchi, T. and Ueda, R. (2000b). Vitamin D receptor gene polymorphism and calcium metabolism in sarcoidosis patients. *Sarcoidosis Vasc Diffuse Lung Dis* 17, 266–269.

Nogee, L.M., Wert, S.E., Proffit, S.A., Hull, W.M. and Whitsett, J.A. (2000). Allelic heterogeneity in hereditary surfactant protein B (SP-B) deficiency. *Am J Respir Crit Care Med* 161, 973–981.

Nogee, L.M., Dunbar, A.E., 3rd, Wert, S.E., Askin, F., Hamvas, A. and Whitsett, J.A. (2001). A mutation in the surfactant protein C gene associated with familial interstitial lung disease. *N Engl J Med* 344, 573–579.

Ortiz, L.A., Lasky, J., Hamilton, R.F., Jr., Holian, A., Hoyle, G.W., Banks, W., Peschon, J.J., Brody, A.R., Lungarella, G. and Friedman, M. (1998). Expression of TNF and the necessity of TNF receptors in bleomycin- induced lung injury in mice. *Exp Lung Res* 24, 721–743.

Pabst, S., Baumgarten, G., Stremmel, A., Lennarz, M., Knufermann, P., Gillissen, A., Vetter, H. and Grohe, C. (2006). Toll-like receptor (TLR) 4 polymorphisms are associated with a chronic course of sarcoidosis. *Clin Exp Immunol* 143, 420–426.

Polakoff, P.L., Horn, B.R. and Scherer, O.R. (1979). Prevalence of radiographic abnormalities among Northern California shipyard workers. *Ann NY Acad Sci* 33, 333–339.

Riccardi, V. (1981). Von Recklinghausen neurofibromatosis. *N Engl J Med* 305, 1617–1627.

Rossi, G., Szapiel, S., Ferrans, V. and Crystal, R. (1987). Susceptibility to experimental interstitial lung disease is modified by immune- and non-immune related genes. *Am Rev Respir Dis* 135, 448–455.

Rossman, M.D., Thompson, B., Frederick, M., Maliarik, M., Iannuzzi, M.C., Rybicki, B.A., Pandey, J.P., Newman, L.S., Magira, E., Beznik-Cizman, B. et al. (2003). HLA-DRB1*1101: A significant risk factor for sarcoidosis in blacks and whites. *Am J Hum Genet* 73, 720–735.

Rottoli, P., Magi, B., Perari, M.G., Liberatori, S., Nikiforakis, N., Bargagli, E., Cianti, R., Bini, L. and Pallini, V. (2005). Cytokine profile and proteome analysis in bronchoalveolar lavage of patients with sarcoidosis, pulmonary fibrosis associated with systemic sclerosis and idiopathic pulmonary fibrosis. *Proteomics* 5, 1423–1430.

Rybicki, B.A., Iannuzzi, M.C., Frederick, M.M., Thompson, B.W., Rossman, M.D., Bresnitz, E.A., Terrin, M.L., Moller, D.R., Barnard, J., Baughman, R.P. et al. (2001a). Familial aggregation of sarcoidosis. A case-control etiologic study of sarcoidosis (ACCESS). *Am J Respir Crit Care Med* 164, 2085–2091.

Rybicki, B.A., Kirkey, K.L., Major, M., Maliarik, M.J., Popovich, J., Jr., Chase, G.A. and Iannuzzi, M.C. (2001b). Familial risk ratio of sarcoidosis in African-American sibs and parents. *Am J Epidemiol* 153, 188–193.

Rybicki, B.A., Hirst, K., Iyengar, S.K., Barnard, J.G., Judson, M.A., Rose, C.S., Donohue, J.F., Kavuru, M.S., Rabin, D.L., Rossman, M.D. et al. (2005a). A sarcoidosis genetic linkage consortium: The sarcoidosis genetic analysis (SAGA) study. *Sarcoidosis Vasc Diffuse Lung Dis* 22, 115–122.

Rybicki, B.A., Walewski, J.L., Maliarik, M.J., Kian, H. and Iannuzzi, M.C. (2005b). The BTNL2 gene and sarcoidosis susceptibility in African Americans and Whites. *Am J Hum Genet* 77, 491–499.

Sabounchi-Schutt, F., Astrom, J., Hellman, U., Eklund, A. and Grunewald, J. (2003). Changes in bronchoalveolar lavage fluid proteins in sarcoidosis: A proteomics approach. *Eur Respir J* 21, 414–420.

Sato, H., Grutters, J.C., Pantelidis, P., Mizzon, A.N., Ahmad, T., Van Houte, A.J., Lammers, J.W., Van Den Bosch, J.M., Welsh, K.I. and Du Bois, R.M. (2002). HLA-DQB1*0201: A marker for good prognosis in British and Dutch patients with sarcoidosis. *Am J Respir Cell Mol Biol* 27, 406–412.

Schneider, E., Epstein, C., Kaback, M. and Brandes, D. (1977). Severe pulmonary involvement in adult Gaucher's disease: Report of three cases and review of the literature. *Am J Med* 63, 475–480.

Schurmann, M., Reichel, P., Muller-Myhsok, B., Schlaak, M., Muller-Quernheim, J. and Schwinger, E. (2001). Results from a genome-wide search for predisposing genes in sarcoidosis. *Am J Respir Crit Care Med* 164, 840–846.

Seitzer, U., Gerdes, J. and Muller-Quernheim, J. (2002). Genotyping in the MHC locus: Potential for defining predictive markers in sarcoidosis. *Respir Res* 3, 6.

Selikoff, I.J., Lilis, R. and Nicholson, W.J. (1979). Asbestos disease in United States shipyards. *Ann NY Acad Sci* 330, 293–311.

Selman, M., Lin, H.M., Montano, M., Jenkins, A.L., Estrada, A., Lin, Z., Wang, G., DiAngelo, S.L., Guo, X., Umstead, T.M. et al. (2003). Surfactant protein A and B genetic variants predispose to idiopathic pulmonary fibrosis. *Hum Genet* 113, 542–550. Epub Sep 6.

Selman, M. et al. (2006). Gene expression profiles distinguish idiopathic pulmonary fibrosis from hypersensitivity pneumonitis. *Am J Respir Crit Care Med* 173, 188–198.

Siltzbach, L.E. (1961). The Kveim test in sarcoidosis. A study of 750 patients. *JAMA* 178, 476–482.

Solliday, N.H., Williams, J.A., Gaensler, E.A., Coutu, R.E. and Carrington, C. (1973). Familial chronic interstitial pneumonia. *Am Rev Respir Dis* 108, 193–204.

Song, Z., Marzilli, L., Greenlee, B.M., Chen, E.S., Silver, R.F., Askin, F.B., Teirstein, A.S., Zhang, Y., Cotter, R.J. and Moller, D.R. (2005). Mycobacterial catalase-peroxidase is a tissue antigen and target of the adaptive immune response in systemic sarcoidosis. *J Exp Med* 201, 755–767.

Spagnolo, P., Renzoni, E.A., Wells, A.U., Copley, S.J., Desai, S.R., Sato, H., Grutters, J.C., Abdallah, A., Taegtmeyer, A., du Bois, R.M. et al. (2005). C-C chemokine receptor 5 gene variants in relation to lung disease in sarcoidosis. *Am J Respir Crit Care Med* 172, 721–728.

Statement on sarcoidosis (1999). Joint Statement of the American Thoracic Society (ATS), the European Respiratory Society (ERS) and the World Association of Sarcoidosis and Other Granulomatous Disorders (WASOG) adopted by the ATS Board of Directors and by the ERS Executive Committee, February 1999. *Am J Respir Crit Care Med* 160, 736–755.

Steele, M.P., Speer, M.C., Loyd, J.E., Brown, K.K., Herron, A., Slifer, S.H., Burch, L.H., Wahidi, M.M., Phillips, J.A., 3rd, Sporn, T.A. et al. (2005). Clinical and pathologic features of familial interstitial pneumonia. *Am J Respir Crit Care Med* 172, 1146–1152.

Swaye, P., van Ordstrand, H.S., McCormack, L.J. and Wolpaw, S.E. (1969). Familial Hamman-Rich syndrome. *Chest* 55, 7–12.

Swider, C., Schnittger, L., Bogunia-Kubik, K., Gerdes, J., Flad, H., Lange, A. and Seitzer, U. (1999). TNF-alpha and HLA-DR genotyping as potential prognostic markers in pulmonary sarcoidosis. *Eur Cytokine Netw* 10, 143–146.

Takada, T., Suzuki, E., Ishida, T., Moriyama, H., Ooi, H., Hasegawa, T., Tsukuda, H. and Gejyo, F. (2001). Polymorphism in RANTES chemokine promoter affects extent of sarcoidosis in a Japanese population. *Tissue Antigens* 58, 293–298.

Terry, R., Sperry, W. and Brodoff, B. (1954). Adult lipoidosis resembling Neimann-Pick disease. *Am J Pathol* 30, 263–286.

Thomas, A.Q., Lane, K., Phillips, J., 3rd, Prince, M., Markin, C., Speer, M., Schwartz, D.A., Gaddipati, R., Marney, A., Johnson, J. et al. (2002). Heterozygosity for a surfactant protein C gene mutation associated with usual interstitial pneumonitis and cellular nonspecific interstitial pneumonitis in one kindred. *Am J Respir Crit Care Med* 165, 1322–1328.

Tsukahara, M. and Kajii, T. (1983). Interstitial pulmonary fibrosis in two sisters. Possible autosomal recessive inheritance. *Jinrui Idengaku Zasshi* 28, 263–267.

Valentonyte, R., Hampe, J., Huse, K., Rosenstiel, P., Albrecht, M., Stenzel, A., Nagy, M., Gaede, K.I., Franke, A., Haesler, R. et al. (2005). Sarcoidosis is associated with a truncating splice site mutation in BTNL2. *Nat Genet* 37, 357–364.

Warshamana, G.S., Pociask, D.A., Sime, P., Schwartz, D.A. and Brody, A.R. (2002). Susceptibility to asbestos-induced and transforming growth factor-beta1-induced fibroproliferative lung disease in two strains of mice. *Am J Respir Cell Mol Biol* 27, 705–713.

Whitsett, J.A. and Weaver, T.E. (2002). Hydrophobic surfactant proteins in lung function and disease. *N Engl J Med* 347, 2141–2148.

Wiman, L.G. (1972). Familial occurrence of sarcoidosis. *Scand J Respir Dis Suppl* 80, 115–119.

Yang, I.V., et al. (2005). Gene expression profiling distinguishes familial and non-familial forms of pulmonary fibrosis. Proceedings of the American Thoracic Society, A242, May.

Zorzetto, M., Bombieri, C., Ferrarotti, I., Medaglia, S., Agostini, C., Tinelli, C., Malerba, G., Carrabino, N., Beretta, A., Casali, L. et al. (2002). Complement receptor 1 gene polymorphisms in sarcoidosis. *Am J Respir Cell Mol Biol* 27, 17–23.

CHAPTER

Peptic Ulcer Disease

John Holton

INTRODUCTION

Peptic ulcer disease (PUD) comprises both duodenal ulcer (DU) and gastric ulcer (GU). Duodenal ulcers occur most often in the first part of the duodenum or in the pre-pyloric region of the stomach (antrum). Gastric ulcers are most frequently seen on the lesser curve of the stomach at the junction of the body and antrum (angularis) (Figure 49.1). Acute stress ulcers involve the body of the stomach and are often multifocal and transient. Histologically, the ulcer is a break in the mucosa with loss of epithelial cells, exposure of the basement membrane and involvement of the muscularis mucosae.

Ulcers develop because of an imbalance between the normal protective attributes of the stomach and the potentially damaging secretions in the lumen of the stomach. This imbalance may be caused by a number of factors, the principal one being colonization by *Helicobacter pylori* (Figure 49.2). Ulceration may also occur associated with a number of other conditions such as Crohns disease, vascular insufficiency, hypersecretory states such as gastrinoma (Zollinger Ellison syndrome), antral G cell hyperplasia, mastocytosis and multiple endocrine neoplasms (MEN-1). Acute stress ulcers are caused by excess alcohol use, non-steroidal anti-inflammatory agents (NSAIDS), burns, trauma to the central nervous system, cirrhosis, chronic pulmonary disease, renal failure, radiation and chemotherapy.

Pre-*Helicobacter* Management of PUD

Prior to the isolation of *H. pylori*, ulceration was thought to be caused by psychological stress and excess spicy foods. It was recognized that smoking was a contributory factor and that excess acid in the stomach was the primary pathology. The treatment was mainly surgical. The first operation was carried out in 1882 with antral resection and duodenogastrostomy (Billroth I) followed a few years later by antral resection and gastrojejunostomy (Billroth II). The aim of these surgical procedures was to reduce the antral phase of acid production. In 1943, Dragstedt introduced truncal vagotomy to reduce the cephalic phase of acid production. However, vagotomy also reduced the drainage from the stomach and had to be combined with a pyloroplasty in order to allow free drainage from the stomach. Frequently both these procedures (antral resection, either Billroth I or II; vagotomy with pyloroplasty) were combined. In the 1970s highly selective vagotomy was introduced, which specifically reduced acid secretion. Side effects of these surgical procedures included diarrhea, a dumping syndrome and anemia. In addition to surgical procedures, in 1976 long-term pharmacological acid suppression with H_2-receptor antagonists was introduced to ameliorate symptoms and allow ulcer healing. However, if the medication was stopped, the ulcer frequently recurred.

Isolation of *H. pylori* and Post-*Helicobacter* Management

Curved or spiral shaped bacteria had been seen in the stomach of various animals and humans as far back as 1888 and had been thought of as commensals and of no significance. Since this period, however, two technical advances set the ground for the subsequent isolation of *H. pylori*: in the 1970s the development of endoscopes allow stomach biopsies to be taken, and the

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

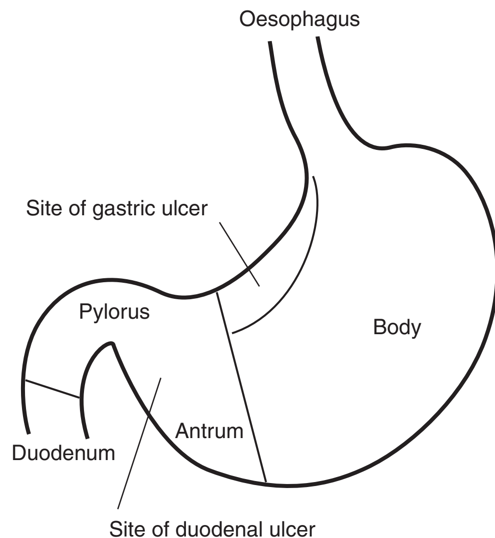


Figure 49.1 This diagram shows the usual location of gastric and duodenal ulcers. Duodenal ulcers are found in the antrum of the stomach or the first part of the duodenum. Gastric ulcers are located at the junction of the antrum and fundus or body of the stomach.

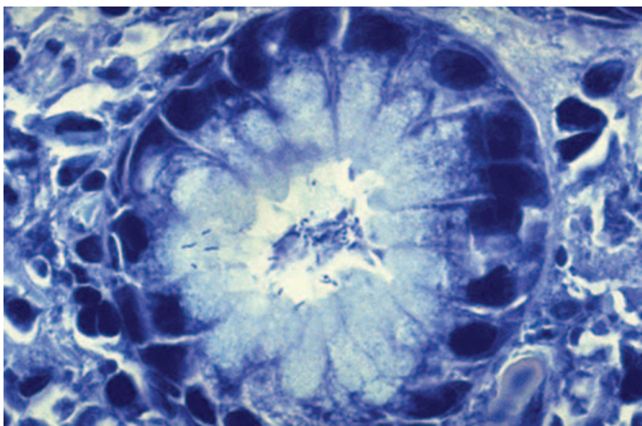


Figure 49.2 This is a Giemsa stained section of stomach showing *Helicobacter pylori* located within a gastric gland, Some of the bacteria are located on the surface of the epithelial cells.

recognition of cultural conditions necessary for the isolation of microaerophilic bacteria. In 1982–1983, Warren, a histopathologist from Perth, Australia, had noted an association between the presence of these organisms and gastritis. He and his colleague Marshall, a gastroenterologist, studied a large series of endoscopic biopsies demonstrating the almost perfect association between gastritis and the presence of this curved bacterium (Warren and Marshall, 1983) that was initially called a campylobacter-like organism. Additionally, they were successful in isolating the organism on artificial media by culturing it under microaerobic conditions similar to that used to isolate the gastrointestinal pathogen *Campylobacter jejuni*. The organism was initially called *Campylobacter pylori* but in 1989 was transferred to a separate Genus called *Helicobacter* (Goodwin, 1994). The initial suggestion that gastritis and ulceration may be caused by this organism was met with scepticism, leading both Marshall and Morris to separately swallow a culture of the organism in order to fulfill Koch's Postulates. Both individuals developed gastritis, which resolved on eradication of the organism. In 1997 the genome sequence of one strain of the organism was determined (Tomb et al., 1997). Eventually, the causal nature of at least 90% of cases of PUD as being *H. pylori* was widely accepted, and both Warren and Marshall received the Nobel Prize for Medicine and Physiology in 2005 (Megraud, 2005).

The recognition that *H. pylori* is the principal cause of PUD has led to a dramatic change in the management of disease: from surgery or lifelong H_2-receptor antagonists to a short course of an acid suppressive (usually a proton pump inhibitor, PPI) combined with two antibiotics – usually amoxycillin and clarithromycin or amoxicillin and metronidazole (European Helicobacter Study Group, 2005).

PATHOPHYSIOLOGY OF ULCER FORMATION

The two most common causes of PUD are infection with *H. pylori* and taking NSAIDS.

Helicobacter-Induced Ulcers

H. pylori causes damage to the host by five principal mechanisms:

1. *H. pylori* causes direct damage to the protective mucus barrier and the epithelial cell layer by inhibiting mucus secretion, secreting phospholipase A (which disrupts the micellar structure of the mucus) and generating ammonium which, in addition to disrupting the protective mucus layer, is also directly cytotoxic for the gastric epithelial cells. *H. pylori* also secretes a vacuolating cytotoxin which is directly cytotoxic (Harris et al., 1996; Kubota et al., 2004) and injects the cagA protein (see later) into the epithelial cell by means of a type IV secretion apparatus. This latter protein disrupts intracellular signaling pathways in the epithelial cell which in part leads to secretion of the chemokine (IL-8).
2. *H. pylori* induces a Th-1 mediated acute inflammatory response by stimulating IL-8 secretion from the epithelial cells and recruiting granulocytes into the lamina propria. Products from *H. pylori* induce the release of free oxygen radicals from the granulocytes, thus causing bystander damage to the epithelial cells (Brisslert et al., 2005).
3. *H. pylori* inhibits the negative feedback loop in the control of acid regulation by means of its lipopolysaccharide inhibiting the production of somatostatin. This leads to continuous production of gastrin (and thus acid) and a characteristic hypergastrinaemia (Moss et al., 1992).

4. *H. pylori* affects the balance of cell division/apoptosis, thus affecting ulcer development and healing as well as having pathological implications for the long-term complication of colonization by *H. pylori* – gastric adenocarcinoma (Cai et al., 2005).
5. During infection with *Helicobacter pylori* auto-antibodies are produced, which target the acid secreting mechanism of the parietal cells and eventually may lead to loss of parietal cells and thus loss of the ability to produce acid (Lo et al., 2005).

NSAIDS-Induced Ulcers

Non-steroidal anti-inflammatory agents damage the stomach by two mechanisms, a local mechanism and a systemic effect on mucus production.

1. NSAIDS are both weak acids and lipophilic, and as such the hydrophobic surfactant layer on the apical surface of the epithelial cells and in the mucus does not present such a barrier to penetration. Once they have penetrated the protective mucus, NSAIDs cause direct damage to the epithelial cell by altering cell permeability and damaging mitochondria, leading to apoptosis (Nagano et al., 2005).
2. When absorbed into the blood, NSAIDS inhibit the enzyme cyclo-oxygenase 1 (COX-1), which produces prostaglandins from arachidonic acid metabolism. The effect of prostaglandins is to increase blood flow in the stomach, increase mucus and bicarbonate production and inhibit acid production. Reduction in prostaglandin secretion diminishes three of the major defense mechanisms of the stomach: bicarbonate and mucus production and blood flow (Gambero et al., 2005).

Ulcer Formation

The net effect of these various insults is the loss of protective mucus and death of the epithelial barrier, resulting in exposure of the damaged epithelium and underlying lamina propria to the effects of acid and proteolytic enzymes resulting in an ulcer (Figures 49.3 and 49.4). If *Helicobacter* spreads from the antrum to give a pan-gastritis, the acid-producing cells suffer damage, and eventually an atrophic gastritis develops with a decrease in acid production and an increase in intraluminal *p*H. Atrophic gastritis and the subsequent intestinal metaplasia that develops are premalignant risk factors that in a percentage of individuals will lead onto the development of gastric adenocarcinoma. Similar damage results from NSAID usage but by a different mechanism.

THE HELICOBACTER GENOME

Genome Structure and Diversity

In 1997, the genome sequence of *H. pylori* 26695 was sequenced by The Institute for Genomic Research (TIGR) (Tomb et al., 1997) and 2 years later that of another strain, J99, was sequenced by AstraZeneca in collaboration with Genome Therapeutics Corporation, United States of America.

Strain 26695 has 1667867 base pairs (bp) an average GC content of 39% and 1590 predicted coding sequences, of which 1091 were annotated. Well developed systems for motility, iron scavenging, restriction and modification and outer membrane proteins (OMPs)/adhesions were noted with a relatively restricted repertoire of metabolic and biosynthetic systems or regulatory networks. An abundance of homopolymeric tracts was found, underscoring the potential for antigenic variation by slipped strand mispairing. In this strain there were five areas with a different GC ratio, one of which has been identified as a pathogenicity island (PAI), and two others in relation to the presence of insertion sequences (IS). Two ISs were identified with 13 copies of one (IS605) and 4 copies of the other (IS606). In all 95 paralogous gene families (16% of the total genome) are present, the largest family with 32 members related to porins/OMP. Three secretion pathways are present: the SecA-dependent one found in many bacteria, a Type IV secretion system associated with the PAI and related to that of *Agrobacterium* and a flagellar export pathway. The large number of homopolymeric tracts, dinucleotide repeats and paralogous families explains the potential for the high degree of variation known to occur in the *H. pylori* genome, reflecting its ability to survive and evolve in its restricted ecological niche. Although placed in the gamma Proteobacteria it also carries genes more closely related to non-Proteobacteria testifying to lateral gene transfer at some point in the organism's evolutionary history.

The second *H. pylori* isolate to be sequenced allowed a comparison of both J99 and 26695 (Alm et al., 1999). *H. pylori* J99 has 1643831 bp, which is smaller than that of 26695 by 24036 bp. The number of predicted coding sequences in J99 is 1495, which is fewer than the revised number of predicted coding sequences in 26695 (now estimated to be 1552 rather than the original 1590). Each strain has unique genes not found in the other strain. J99 has 89 unique genes, of which 56 are of unknown function whilst 26695 has 117 unique genes with 91 of unknown function. There are a predicted 1406 genes common to both strains. J99 has fewer complete ISs. Like 26695, J99 carries a complete PAI and also like strain 26695 has a large number of paralogous gene families. One region of both genomes carries 46–48% of the strain-specific genes and has been termed the plasticity zone (PZ). In J99 this PZ is continuous but in strain 26695 it is split into two sections separated by about 600 kb. Allelic variation occurs in both strains with differences found mainly at the third codon position. Moderate macrogenome diversity also exists between the strains, with about 85% of orthologs having the same neighbor at both sides, 10% with the same neighbor on one side and a strain-specific gene on the other and about 3% flanked by strain-specific genes. Despite these genomic variations the proteome of both strains are very similar. Following the sequencing of the second *H. pylori* genome (J99), it was necessary to re-annotate both genomes to a common classification (Boneca et al., 2003).

H. pylori has one of the highest allelic diversity of any organism, and comparison of both sequenced genomes indicates a high degree of size variation between orthologs due to a variety of

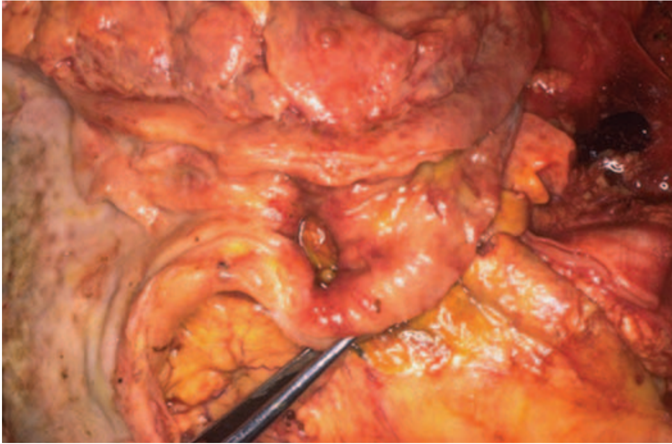


Figure 49.3 This is a photograph of the macroscopic appearance of a duodenal ulcer in the antrum of the stomach.

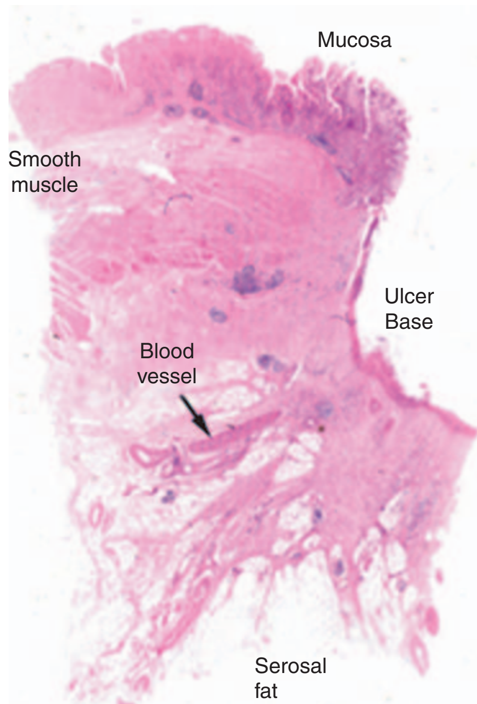


Figure 49.4 This is a photomicrograph of a hematoxylin and eosin stained section of stomach illustrating an ulcer that has eroded the epithelium exposing the underlying tissues.

mechanisms. The diversity is generated by frequent recombination, a high mutation rate, slipped strand mispairing, frame shifts and the presence of pseudogenes (Tomb et al., 1997; Suerbaum et al., 1998; Wang et al., 1999). Much of the diversity is localized within the PAI or PZ. On a macrogenomic scale, however, the sequences of the two genomes indicate a similar organization.

The population genomic structure of *H. pylori* is regarded as panmictic rather than clonal (Go et al., 1996) although isolates from an individual or family group are clonal (Han et al., 2000). A study of 23 isolates using electrophoresis of 16 enzymes indicated an allelic diversity ranging from 2–11 alleles with no obvious clustering of disease-related strains (Hazell et al., 1997). A whole-genome microarray analysis of 15 strains utilizing 1660 unique sequences derived from the strains J99 and 26695 showed a high degree of genetic diversity between the strains (Salama et al., 2000). One thousand two hundred and eighty one sequences were common to all the strains and 362 were missing from at least one of the isolates. At a conservative estimate, about 12–18% of the genes of an individual isolate were strain-specific. This genetic diversity is probably a reason for the ability of *H. pylori* to become a chronic pathogen in the stomach despite a vigorous local and systemic immune response. In part, it explains the different clinical outcomes of infection with *H. pylori*, and it also reflects the continuing evolution of the organism over decades within the human stomach of an individual. Several studies (Han et al., 2000; Israel et al., 2001; Lundin et al., 2005) have demonstrated the presence of related sub-strains of *H. pylori* within the stomach of a single individual. In one study the person who was infected with the original sequenced isolate, J99, remained infected despite appropriate therapy, and after 6 years additional cultures were isolated from the same patient. The new isolates were subjected to random amplified polymorphic DNA-PCR (RAPD-PCR), sequencing of unlinked genes and micro-array analysis. The results demonstrated that all the isolates were related to the original strain, although the isolates varied in genetic content from the original strain and co-isolates by about 3%. Current isolates did not hybridize with some open reading frames (ORFs) from the original strain with a loss of 0.28–1.52% ORFs and yet some ORF sequences from current isolates hybridized with specific sequences from the other strain, *H. pylori* 26695, of which the whole genome sequence is known. A second study confirmed the presence of closely related sub-clones isolated from a single individual (Bjorkholm et al., 2001) and additionally went on to demonstrate the genetic and phenotypic stability of these two clones in an experimental system using transgenic mice expressing human Lewis b, the ligand for the BabA adhesion on the organism. Of the two original isolates, one had lost the PAI and, although this isolate could colonize germ-free transgenic mice along with the PAI positive isolate, only the PAI positive isolate could colonize conventional mice harboring a normal flora. Over a 3- or 10-month period of study, no nucleotide substitutions or deletions were detected in either strain indicating that in this model system either there was insufficient selective pressure or time for divergence to occur.

Allelic Diversity and Disease

H. pylori colonizes over half the population of the world. In some countries the colonization rate is as high as 90% even by the age of 10 years. In others it is about 30% overall, with only 5% being colonized within the first decade. This geographical difference in prevalence is caused principally by variation in availability of a good public health infrastructure and prevailing social conditions. Of those colonized, the vast majority are

without symptoms although they do have gastritis. A small proportion, about 5–20%, will develop PUD and an even smaller percentage, about 1%, which develop gastric cancer. The difference in clinical outcome is the results of the interaction of three factors: environmental, host susceptibility and strain variation.

Strain Variation and Virulence

As mentioned, *H. pylori* possesses a number of virulence characteristics that are not present or not expressed in all strains. Some strains have acquired a PAI that carries 31 different genes, some of which relate to the synthesis of a Type IV secretion system, and a gene (*cagA*) coding for a protein that the Type IV secretion system transfers to gastric epithelial cells – the cytotoxin associated gene product (CagA). Nearly all strains carry another locus *vacA*, separate from the PAI, that secretes a vacuolating cytotoxin (VacA). This latter product, however, is secreted in varying amounts depending on the presence of allelic diversity within the *vacA* gene and possibly the presence of the cagA PAI. Strains can thus be broadly divided into two main types: Type I that co-express *cagA* and *vacA* and Type II that do not (Xiang et al., 1995). The clinical significance of this is that Type I strains are more often isolated from patients with serious gastroduodenal disease such as PUD and gastric cancer.

Vacuolating Cytotoxin

About 50% of isolates produce a vacuolating cytotoxin. The *vacA* gene encodes a 139–140 kDa protein which is cleaved into a 33 amino acid signal sequence and a 90 kDa toxin. The *C*-terminal domain of the toxin comprises 14–16 β strands that inserts into the bacterial membrane and acts as an autotransporter (Type V secretion system) for the toxin (Fischer et al., 2001) that is secreted. After secretion, the toxin is further cleaved into a 34 kDa *N*-terminal region and a 58 kDa *C*-terminal region that nevertheless remain associated. The toxin monomer is activated by acid, undergoing a conformational change and then oligomerizes to form a 900 kDa protein, as it associates with lipid rafts in the cytoplasmic membrane of epithelial cells (Lupetti et al., 1996). In the cytoplasm of the epithelial cell, its mode of action is to disrupt the endocytic pathway leading to enlargement of late endosomes. The toxin forms anion-selective channels in membranes allowing the accumulation of weak bases, stimulated by the acidification of the vacuole by V-ATPase. Additionally, transfection experiments show that the p34 peptide becomes associated with the mitochondria releasing cytochrome C from the mitochondria and thus activating the caspase cascade (Galmich et al., 2000). When given directly onto gastric epithelium, the cytotoxin induces damage causing necrosis and ulceration. A further consequence of the action of the toxin is the inhibition of processing of antigens in antigen-presenting cells, thus affecting T-cell responses.

Although most strains carry the *vacA* gene, there is a large diversity in the sequence of the mid region (MR) and the leader sequence (LS). This diversity has consequences in relation to secretion of toxin, binding to cell types and geographical distribution of strains. The general structure of the VacA protein is illustrated in Figure 49.5. There are four variations of the LS designated S1a,S1b,S1c and S2 and several variations of the MR designated m1a, m1b, m1T, m2a m2b and chimeric types, for example, m1b-m2 (Atherton et al., 1995; Pan et al., 1998; Strobel et al., 1998; van Doorn et al., 1998; van Doorn et al., 1999; Wang et al., 1998). The primers used to detect these variations are given in Table 49.1.

Strains with an S1 LS have a different cleavage site from that of S2, and these strains are more often found associated with PUD than S2-containing strains. They produce high levels of cytotoxin and are more frequently found in cagA positive strains (Atherton, 1997). The MR is also an independent marker for cytotoxin activity *in vitro* with m1 stains being more cytotoxic than m2 strains. The MRs are responsible for binding to cells and depending upon the cell type m2 strains are just as toxic as m1 strains (Ji et al., 2000; Pagliaccia et al., 1998). Various combinations of the LS and MR alleles have been isolated from patients, giving strains with different phenotypes of cytotoxin activity. Strains with the S1/m1 alleles appear to produce more cytotoxin compared to S2/m2 strains, which apparently do not produce detectable cytotoxin activity, although the gene is transcribed. This lack of association between *H. pylori* S1/m2 strains in countries with high levels of PUD and gastric cancer is explicable by the lack of binding of m2 strains to cell types used in testing cytotoxic activity. However, there may also be a modulation of toxin production based on levels of transcription of the vacA gene (Forsyth et al., 1998), which could affect the phenotype of the strain and clinical outcomes.

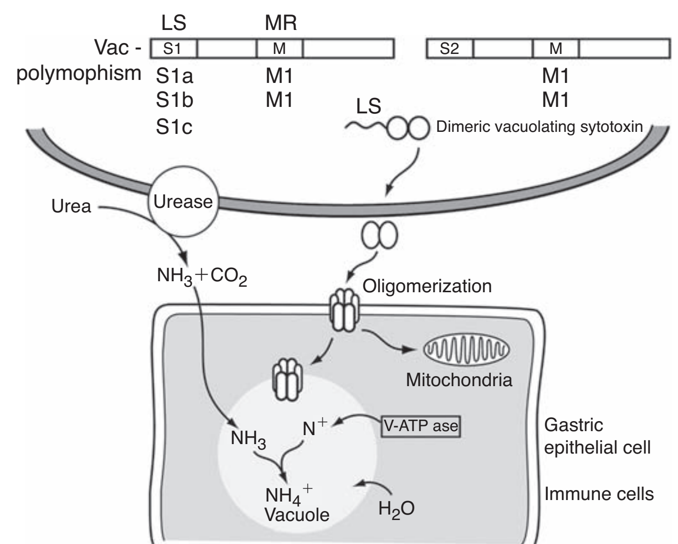


Figure 49.5 This diagram illustrates the polymorphisms found in the *vac A* gene in the LS and mid region (MR) of the gene. It also illustrates the secretion and oligomerization of the vacA protein and its action on the epithelial/immune cell to disrupt normal endosomal trafficking leading to the production of large vacuoles and its effect on the mitochondria leading to apoptosis.

TABLE 49.1 Primers for some polymorphic regions of *Helicobacter pylori*

Primer	Polymorphic locus	References
5′GTCAGCATCACACCGCAAC3′ 5′CTGCTTGAATGCGCCAAAC3′	vacA s1aSS1-F VA1-R	Atherton et al. (1995)
5′AGCGCCATACCGCAAGAG3′	vacAs1bSS3-F/VA1-R	Atherton et al. (1995)
5′CTCTCGCTTTAGTGGGGYT3′	vacA s1c S1C-F/VA1-R	Yamazaki et al. (2005)
5′ATGGAAATACAACAAACACAC3′	vacA s2 VA1-F/VA1-R	Atherton et al. (1995)
5′GGTCAAAATGCGGTCATGG3′ 5′CCATTGGTACCTGTAGAAAC3′	vacA m1aVA3-F VA3-R	Atherton et al. (1995)
5′GGAGCCCCAGGAAACATTG3′ 5′CTGCTTGAATGCGCCAAAC3′	vacA m2 VA4-F VA4-R	Atherton et al. (1995)
5′CAATCTGTCCAATCAAGCGAG3′ 5′GCGTCAAAATAATTCCAAGG3′	vacA m1/m2 VAG-F VAG-R	Atherton et al. (1999)
5′GGCCCCAATGCAGTCATGGAT 3′ 5′GCTGTTAGTGCCTAAAGAAGCAT3′	vacA m1b VAm-F3 Vam-R3	Pan et al. (1998)
5′GTGTTTTTAACCAAAGTATC3′ 5′CTATAGCCASTYTCTTTGCA3′	IceA1 iceA1F iceA1R	van Doorn et al. (1998)
5′CAACGATAATGGCACAAACT3′ 5′GTCGTATCAATAACAGAAGTTG3′	Type I hopQ OP5136 OP4829	Cao et al. (2002)
5′GATAACAGGCAAGCTTTTGAGG3′ 5′CTGCAAAAGATTGTTTGGCAGA3′	cagA CAGAF CAGAR	van Doorn et al. (1998)
5′GGCAATGGTGGTCCTGGAGCTAGGC3′ 5′GGAAATCTTTAATCTCAGTTCGG3′	5′end cagA cagA5 cagA2	Mukhopadhyay et al. (2000)
5′AGGATTTCAGCAAGGTAACGCAAGC3′ 5′TAAGATTTTTGGAAACCACCTTTTGTAT3′	3′end cagA cagA-F40481 cagA-R41660	Mukhopadhyay et al. (2000)
5′ACCCTAGTCGGTAATGGGTTA3′ 5′GTAATTGTCTAGTTTCGC3′	3′end cagA CAG1 CAG2	Yamaoka et al. (1998b)
5′AATCCAAAAAGGAGGAAAAACATGAAA3′ 5′TGTTAGTGATTTCGGTGTAGGACA3′	BabA2	Rad et al. (2002)

Cag PAI and the CagA Gene Product

CagA is a marker for the cag PAI: a 40 Kb DNA fragment containing 31 genes, of which 6 encode a Type IV secretion system, by which the CagA protein is injected into host cells. Once inside the gastric epithelial cell, it is phosphorylated at a tyrosine residue, present in EPIYA motifs, by Src-family kinases and binds to the SH2 domain of SHP-2 phosphatase, thus deregulating its activity (Hatakeyama, 2003). The immediate result of this is the alteration of cell morphology, with the prolonged activation of Erk signaling and the production of the 'humming bird' phenotype. The cell cycle is also interfered with, caused by an increasing expression of cyclin D3, the phosphorylation of the tumor suppressor Rb and the expression of the transcription factor c-jun, leading to the increase in cell proliferation of gastric epithelial cells, probably caused by a deregulation of the G1/S checkpoint of the cell cycle (De Luca et al., 2003). These effects may have long-term consequences in relation to gastric carcinogenesis. Additionally, CagA recruits both scaffolding protein ZO-1 and the junctional adhesion protein to its site of attachment (inducing loss of villi and pedestal formation) and affects epithelial barrier functions (Amieva et al., 2003). Finally, engagement of the Type IV secretion apparatus leads to secretion of IL-8 from the epithelial cell with consequent recruitment of granulocytes (Fischer et al., 2001).

The presence of the cag PAI is not an "all or none" phenomenon. Several studies have shown that the PAI can exist as a single contiguous entity or can be divided by an insertion element (IS 605) or chromosomal DNA into two parts; cag I and cag II. Various deletions of the cag PAI have also been recognized (Censini et al., 1996; Covacci et al., 1997). The insertion

element is an additional variable as it may be completely absent or it may be present in variable numbers located in different parts of the genome (Figure 49.6).

The cagA protein is also variable in size (120–140 kDa) due to a different number of repeats of sequences present in the 3′-end of the gene. Analysis of this region of the cagA gene revealed four types designated A-D based on the nucleotide sequence and organization of the locus (Yamaoka et al., 1999). These repeats are illustrated in Figure 49.6. Geographic differences can also be found in the 3′ region of the genome. Studies of European strains compared to East Asian strains have identified nucleotide differences leading to two amino acid sequences denoted as (1) Western cagA specific Sequence (WSS) and (2) East Asian cagA specific sequence (EASS) (Azuma et al., 2002; Yamaoka et al., 1999), namely:

1. FPLKRHDKVDDLSKVGRSVSPEPIYATIDDLG
2. ESSAINRKIDRINKIASAGKGVGGFSGAGRSASPEPIY ATIDFDEANQAG

Differences of nucleotide sequence also exist in the less variable, 5′ end, of the *cagA* gene. Two types have been identified, cagA1 and cagA2, the former found in Europe, America and Australia whilst the latter was found in isolates from East Asia. Isolates with the cagA2 strains were principally vacA S1c, and cagA1 was more commonly linked to vacA m1 strains (van Doorn et al., 1999b). Potential additional complexity is provided by the presence of EPIYA motifs in the cagA protein, whose number and phosphorylation status may vary.

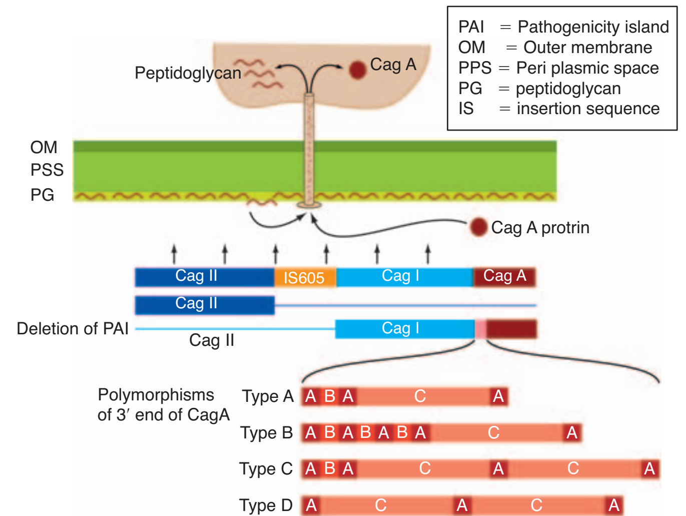


Figure 49.6 This diagram illustrates the PAI and its deletions and the polymorphisms found with in the *cagA* gene based upon the variation in the number of repeats of sequences A, B and C. The diagram also illustrates the inoculation of both bacterial peptidoglycan from the cell wall of *Helicobacter pylori* and the cagA protein into a gastric cell leading to disruption of intercellular signaling. The materials are transferred into the eukaryotic cell by a Type IV secretion system which is coded for by genes on the pathogenicity island.

Other Virulence Markers

Several other virulence characteristics have also been identified. Some of these characteristics, such as the urease enzyme and the flagella, are important for establishing colonization. Urease also acts as a toxin by hydrolyzing urea to produce ammonia, which is cytotoxic, and by diminishing the viscous nature of the mucus layer, which is protective.

Several adhesins have been identified, the best characterized one being BabA, which binds the Lewis b blood group antigen expressed on gastric epithelial cells. Some strains have a non-functional pseudogene (BabA1) whilst other strains only have one functional gene, BabA2, and others have a separate functional allele, BabB. BabA has several polymorphisms located in the MR of the protein leading to variation in degree of binding to Lewis b. Strains expressing a chimeric protein BabA2/BabA1 have also been identified that show phase variation of expression of the protein, and this reported heterogeneity may have clinical implications (Backstrom et al., 2004; Hennig et al., 2004). Other adhesins are SabA (which binds sialyl Lewis x and which is found in inflamed tissue), AlpA and AlpB, HpA and HopZ. Less information is available concerning the other adhesions. The ligands for AlpA/B are unknown. HpA is the coding sequence for an hemaglutinin that binds *N*-acetylneuraminyl-lactose and HopZ exists as two alleles and its expression is possible regulated by slipped strand mispairing.

Other virulence factors that have been identified are: an OMP HopQ that is present as two alleles Type I and Type II, an outer inflammatory protein OipA associated with increased expression of IL-8, and a locus called *ice*A (*i*nduced by *c*ontact with *e*pithelium), which has two alleles *ice*A1 and *ice*A2, the former encoding a putative restriction endonuclease (Peek et al., 1998; Yamaoka et al., 2000).

Whole Genome Expression Studies

In order to understand the changes in genome expression following infection, analysis of genome expression by microarrays have been performed. In one study, the effects of *in vitro* acid stress were investigated (Ang et al., 2001). A microarray containing 1534 ORFs from strain 26695 was prepared and the genome expression at *pH* 7.2 and 5.5 analyzed. Four hundred and forty five ORFs were expressed under both *pH* conditions, 80 had increased expression under acid conditions and 4 had reduced expression. Of those that had increased expression, one group coded for proteins recognized to be involved in the acid-stress response, one group had no obvious connection to acid stress and one group was of unknown function.

A second study (Graham et al., 2002) investigated the *in vivo* whole genome expression from the human stomach. In order to overcome the instability of RNA isolated from tissue and the comparatively low numbers of bacteria found in the stomach, the method of selective capture of transcribed sequences (SCOTS)

combined with microarray analysis was used. In all, 14 genes were differentially expressed *in vivo* compared to *in vitro* grown bacteria, including the VirB4 homolog (part of the Type IV secretion apparatus). The majority of the genes expressed in the human stomach encoded *H. pylori*-specific proteins of unknown function.

This approach of analyzing *in vivo* gene expression is likely to be valuable in identifying disease-related genes, and these results encourage further studies in this area.

Geographic Distribution and PUD

Studies of the genome sequence of the hypervariable regions of the *H. pylori* genome isolated in different geographical locations show that different strains have geographical predominance. These differences can be used to analyze population migrations and also suggest that *H. pylori* may have been transmitted to the New World from European invaders. In Northern and Eastern Europe S1a alleles predominate; in France and North America S1a and S1b are equal in number; in Iberia and South America S1b predominate; and in East Asia S1c is the major strain type. The MR alleles are equally distributed in these regions except for South America where m1 predominates. These studies also confirm that East Asian strains are quite distinct from European isolates (Kersulyte et al., 2000; van der Ende et al., 1998; Yamaoka et al., 2000b).

Markers for PUD

Genome Analysis

The high level of genetic variation within the genome of *H. pylori* explains, in part, the variation in clinical outcome of infection and why the majority of individuals colonized by the organism remain asymptomatic whilst other will develop serious gastroduodenal disease including PUD. Type I strains (presence of S1 *vacA* alleles and production of the vacuolating cytotoxin and presence of the cag PAI) are more often isolated from patients with PUD compared to Type II strains, which do not have the cag PAI and are not phenotypically cytotoxic but possess the vacA S2 allele. Similarly the presence of BabA2, oipA, iceA1 and Type I HopQ alleles have all been linked to PUD in some studies.

In Europe and North America, S1a/m1, cagA positive strains are more common in patients with PUD compared to non-ulcer dyspepsia (NUD), but in China nearly all strains are vacA and cagA positive and are equally distributed between PUD and NUD (Pan et al., 1998). In Korea, S1c strains are more common in PUD (Choe et al., 2002), and in a Japanese study the S1a/m1a WSS 3′ cag region was significantly associated with PUD whereas the S1c/m1b ESS 3′ cag region were not (Yamazaki et al., 2005). The Far East genotype, however, was correlated with gastric atrophy in a study in Japan (Azuma et al., 2004). In Brazil, cagA was correlated with DU but the subtypes of the 3′ cagA region were not correlated with disease groups (Rota et al., 2001). On the other hand, in Malaysia in a study of the three different ethnic groups (Malay, Chinese, Indian) no correlation could be found between cagA, or vacA and PUD (Tan et al., 2005).

In Europe, the association of VacA (s1), cagA, BabA2 or Lewis expression by Hp (Thoreson et al., 2000) is correlated with PUD. The association is more so in strains that are positive for all three alleles (Olfat et al., 2005), but in Brazil a study could not show any correlation between BabA2 and PUD although the locus was more common in chronic gastritis (Gatti et al., 2005). A large study of 827 strains from four countries, however, found that of the three loci, only BabA2 could be considered as a marker for separating PUD from gastritis although none were particularly valuable as predictive markers (Yamaoka et al., 1999; Yamaoka et al., 2002). A further locus identified as a marker for PUD is iceA (van Doorn et al., 1998), although once again a study from Japan could not confirm this (Ito et al., 2000). These conflicting results demonstrate that other bacterial markers may be important and that interaction with host factors are also important in disease outcome.

A study of two strains (both caA^{+}vacAs1, iceA1) isolated from either DU (G1.1) or GU (B128) were investigated by microarray and for the induction of inflammation in a gerbil model (Israel et al., 2001). The GU isolate induced more inflammation, IL-8 secretion and apoptosis compared to the DU isolate. The microarray analysis demonstrated that the DU-inducing strain had a large deletion covering cag6-cag23 (which includes cagE). This study emphasizes both the importance of the cag PAI in pathogenesis of serious disease and also the utility of microarray analysis in identifying potential markers for specific clinical outcomes.

A study of three genes in the PZ (JHP940; JHP947; HP986) from 200 Brazilian patients identified JHP947 as being associated more strongly with DU and gastric carcinoma (99%) as opposed to gastritis (44%) (Santos et al., 2003), but again, as with other studies, it was not specific for PUD.

Proteome Analysis

An immunoproteomic approach using the strain 26695 and sera from 24 *H. pylori* positive patients (Haas et al., 2002) yielded 310 antigenic protein spots, of which some were differentially recognized in patients with gastritis or PUD. Further work by this group (Krah et al., 2003) identified peptides within a subset of spots using MALDI-TOF. The protein GroEL was identified as a component in 15 of the spots and was recognized antigenically.

A proteome analysis of *H. pylori* isolates from two patients (Pereira et al., 2006), one with gastritis and the other with DU showed that four spots were uniquely associated with each condition and six spots had variable expression. Further work is required to confirm this study and to identify the proteins.

Serological Markers of PUD

A number of serological studies have identified low molecular weight proteins as possible markers of PUD. A study of 108 patients with gastric cancer, PUD or NUD showed that antibodies to a 26 kDa protein were more common in PUD and that when combined with sero-prevalence to the cagA protein the sensitivity and specificity was 76% and 62% respectively (Kuo et al., 2003). A similar study of 156 patients from Thailand

(Vilaichone et al., 2003) showed that antibodies to a 35 kDa protein were associated with GU. A study of 80 Japanese patients with different clinical conditions showed the 35 kDa antigen was more common (97%) in GU or DU compared to gastritis (70%) (Yamaoka et al., 1998).

HUMAN POLYMORPHISM AND PUD

The first demonstration that polymorphisms in the human genome could affect the clinical outcome of *Helicobacter* infection was provided by the increases risk of developing gastric adenocarcinoma in patients who had certain polymorphisms in IL-1 (El-Omar et al., 2000). It was subsequently shown that a synergistic effect was seen between bacterial and human polymorphisms (Rad et al., 2003). Thus, the VacA s1 genotype was linked to a greater risk of developing severe gastroduodenal disease, especially if in combination with other 'high risk' host markers such as the host IL-1β genotype IL-1β–511T/-31C and the genotype of its receptor IL-1RN*2.

When examined singly, in relation to the development of DU in 278 individuals, the alleles IL-1B + 3954 and IL-RN*2 were not associated with the development of DU; however, there was a strong allelic link between the two loci in patients with DU (Garcia-Gonzales et al., 2001). On the other hand, a study of 315 individuals showed that the presence of IL-1RN*2 was a risk factor, with an odds ratio (OR) of 22.6 in the presence of colonization by *Helicobacter* (Hsu et al., 2004). A smaller study of 215 individuals could not confirm that any polymorphisms in IL-1B, IL-1RN, TNF-A TNF-B were related to the development of DU (Garcia-Gonzales et al., 2005). However, when looked at in combination: IL-1RN*2; IL-1B-31; IL-1B-511C; IL-1B + 3954C; TNF-haplotypeE was found with a slight (non-significant) increase in the patients with DU. An earlier study by the same group showed that genotype LTA (lymphotoxin, TNF-B) NcoI 2.2 and TNF-I were both more common in GU than DU (Lanas et al., 2001). Further confirmation of tumor necrosis factor (TNF) polymorphisms associated with PUD was provided (Lu et al., 2005), showing that the alleles TNF-A-1031C and TNF-A -863A have an OR of 2.4 for the loci individually and 6.0 in combination.

A significant association between IL-8 A/T heterozygote and duodenal ulceration (Gyulai et al., 2004) or gastric ulceration (Ohyauchi et al., 2005) has been identified compared to the wild type T/T allele. Studies of polymorphisms in IL-6, IL-12 and CD14 have not shown any association (Garcia-Gonzales et al., 2005; Gyulai et al., 2004; Lobo Gatti et al., 2005). Finally, associations between CD11c exon 15 and intron 31 G/A in combination show an OR of 2.4 for GU disease (Hellmig et al., 2005) and polymorphisms in the myloperoxidase (−468 A/A) locus show an OR of 8.7 for DU when infected with *H. pylori* (Hsu et al., 2005). An inability to mount an effective immune response may be more likely to lead to chronic infection and thus indirectly to PUD. A study of microsatellite polymorphisms in the T-cell receptor V β6 locus has identified a genotype (TCRBV6S1B/B) that is not able to eliminate *Helicobacter* and thus is more susceptible to chronic infection although in the cohort studied this genotype did not correlate with clinical outcome (Kunstmann et al., 2000). A study of 20 single nucleotide polymorphisms (SNPs) in the genes for matrix metalloproteases 1, 3, 7 and 9 in 599 isolates from infected patients showed a strong association between GU and a promoter variants of MMP 7 and 9 (Hellmig et al., 2006).

Polymorphisms in genes related to the action of NSAIDS have been identified although relationships to clinical outcome and ulcer formation have not been so far linked. Cyclo-oxygenase 1 (COX-1) converts arachidonic acid to prostaglandin H2, and heterozygosity for the polymorphisms A-842G/C50T are more likely to show a greater inhibition of prostaglandin synthesis by aspirin than is the homozygous (Halushka et al., 2003).

GENOMICS IN THE MANAGEMENT OF DISEASE

Cytochrome P450 (CYP) represents a group of enzymes involved in metabolism of drugs, which are found principally in the liver. The enzymes are coded for by at least 50 genes with a number of pseudogenes identified. In one study it was found that nearly 60% of over 300 drugs were metabolized by CYP (Bertz and Granneman, 1997). Polymorphisms in these enzymes are recognized to be important in treatment of diseases, as some polymorphisms may metabolize different drugs at different rates thus having different clinical outcomes, that is, success or failure. *H. pylori* is treated with PPI, for example, omeprazole, and a combination of clarithromycin and metronidazole (or amoxicillin). CYP2C19 is important in metabolism of PPIs and CYP3A4 in metabolism of clarithromycin. In one study (Sapone et al., 2003) of 143 patients with *H. pylori* infection, lack of eradication was correlated with homozygous extensive metabolizers (HomEM) (CYP2C19*1/*1) or heterozygous extensive metabolizers (HetEM) (CYP2C19*1/*2 or *1/*3). All patients with CYP2C19*2/*2 had their organism successfully eradicated. Those individuals who were HetEM in CYP2C19 and carried the polymorphisms CYP3A4*1B and CYP3A4*2 were more likely to have *Helicobacter* eradicated compared to the wild type CYP3A4, suggesting some synergistic action between the two cytochrome P450s. A cost analysis of genotyping patients for CYP2C19 polymorphisms prior to commencing eradication therapy in the United States showed a savings ranging from $495–$2125 for every ulcer prevented (Lehmann et al., 2003).

Interleukin-1beta is also an important determinant of successful eradication of *Helicobacter* using triple therapy (Furtura et al., 2004; Take et al., 2003). In a study of 336 patients, IL-1B-511T/T had the highest eradication rate (94.7%) compared to IL-1B-511C/C with 77.3%, with the heterozygote C/T having an intermediate eradication rate. This is mediated by the intragastric *p*H (related to the IL-1 polymorphism) as this affects the activity of particularly clarithromycin in the stomach.

An additional *Helicobacter*-associated factor that affects the eradication rate is resistance to clarithromycin. Resistance is related to polymorphisms within the 23S ribosome subunit to which clarithromycin binds. Resistant isolates have A2124G or A2143G polymorphisms, and, if present, the eradication rate is only 48% compared to 87% in the absence of these polymorphisms (Furtura et al., 2005).

Results of a study including 684 subjects showed a reduced risk of developing DU if they carried the TGFb + 869C/C polymorphism and that the T/C polymorphism was involved in susceptibility to developing DU (Garcia-Gonzalez et al., 2006). A correlation between susceptibility to developing DU and NOD1 A796A was present in 20% of patients with DU compared to 6% in controls (Hofner et al., 2007). In children with DU, the G238A of the TNFα gene was present in 31% of cases also carrying the strain marker iceA1, compared to 1.6% in controls (Wilschanski et al., 2007). In adults, iceA1 compared to iceA2 was also more frequent in DU cases (Caner et al., 2007). In rats, real-time PCR and RT-PCR were used to show that of 8000 genes studied in relation to cysteamine-induced ulceration, 40 genes had marked changes in expression, suggesting the interaction of many genes in the development of ulceration (Deng et al., 2007).

The effect of host polymorphisms on eradication of *H. pylori* showed that MDR1 T3435T polymorphism had an eradication rate of 67% when given lansoprazole + amoxicillin + clarithromycin, compared to C/C or C/T which had eradication rates of 81–82% (Furtura et al., 2007).

The effects of host polymorphisms in the cytochrome P450 2C9 (CYP2C9) locus on ulceration and bleeding in association with NSAID use were shown in a study of 26 patients compared to 52 controls, all of the study population being *H. pylori* negative. Higher frequencies of bleeding were found in patients with CYP2C9*1/*3 (34/5.8%) and CYP2C9*1/*2 (27/15.4%) compared to controls (Pilotto et al., 2007).

Genomics in Relation to Diagnosis

A number of molecular techniques have been applied to the detection, typing and assessment of eradication of *H. pylori*. Detection of markers of virulence and antibiotic resistance has been achieved largely by the polymerase chain reaction (PCR) with appropriate primers. The use of real-time PCR allows direct detection and quantification of the target of interest and the development of robotic workstations can increase throughput of analytes. Specimens used for the detection and genotyping of *Helicobacter* has largely been gastric biopsies although stool specimens have also been used. A number of different methods of typing *H. pylori* have been used such as ribotyping, RAPD-PCR, PCR-RFLP and pulse field gel electrophoresis (PFGE) (Ge and Taylor, 1998) in both coding and non-coding areas (Bereswell et al., 2000).

2009 UPDATE

Host–pathogen interaction in *Helicobacter pylori*-related disease continues to be investigated by genomic techniques for polymorphisms in both bacterial and host loci that may act as biomarkers for specific clinical outcomes. Ongoing studies using proteomic analysis of host–pathogen interactions (Wu et al., 2008) and sequence analysis of *H. pylori* isolates (Mcclain et al., 2009) are likely to lead to the identification of disease-related biomarkers.

Microbial Polymorphisms

The role of the *VacA* locus continues to be investigated in various geographical locations. In Italy, the incidence of the VacA s1m1 genotype is increasing despite the falling incidence of PUD, raising questions of the relevance of s1m1 as a biomarker for PUD (De Francesco et al., 2009). In Iran, VacA s1m2 is predominant in PUD (Selehi et al., 2008), although another study from Iran could find no correlation between VacA genotypes and clinical outcome (Hussein et al., 2008). In Iraq, the vacA i1 locus and the dupA locus were both correlated with PUD, but this was not the case in Iran (Hussein et al., 2008). Similarly, in Italy the vacA i1 locus was associated with PUD (Basso et al., 2008). In Iran and Iraq (Selehi et al., 2008), Egypt (Said Essa et al., 2008), and Italy (Basso et al., 2008), the cagA locus was correlated with PUD. In Brazil, a study of restriction fragment length polymorphism in the *ureC* and *ureB* loci in *H. pylori* could demonstrate no differences in the patterns between PUD or gastritis (Roesler et al., 2008). In a Chinese population, the *dupA* locus was positively associated with DU and inversely with GU and gastric cancer (Zhang et al., 2008).

Host Polymorphisms

Various studies of polymorphisms in genes associated with gastric physiology, antigen processing and inflammation have been reported. A study from India has demonstrated that a certain Cox-2 polymorphism is positively associated with PUD (and gastric cancer) (Saxena et al., 2008) and a subsequent study has confirmed this and that other Cox-2 polymorphisms are associated with gastritis or with a lower risk of gastric atrophy (Achyut et al., 2009). A study from the same group on PPARγ demonstrated that carriage of the G allele was associated with PUD and gastric cancer compared to NUD (Prasad et al., 2008). In Brazil, polymorphisms with TLR2, 4 and 5 were not found to be associated with DU in children, although a TLR4 polymorphism was associated with higher levels of IL-8 and IL-10 and colonization by cagA strains (Moura et al., 2008). A study from Korea of polymorphisms in IL-10, IL-8 and IL-6 demonstrated they were significantly associated with GU (and gastric cancer) and that a synergistic effect on the development of either

of these conditions occurred with some of the polymorphisms and others were less frequent than controls in patients with DU (Kang et al., 2008b). A study from Mexico demonstrated that a TLR4 polymorphism was found more often in patients with DU compared to gastritis and expressed lower levels of IL-1β, IL-6, IL-8 and Groα, yet higher levels of TNFα, IL-10, MCP-1 and MIP-1α (Trejo-de la et al., 2008). In Turkey, the role of polymorphisms in IL-1β and IL-1RN was investigated showing as association with PUD and gastric cancer or NUD depending on the polymorphism (Erzin et al., 2008). A study from Japan (Tahara et al., 2008) on the role of a polymorphism in NADPH oxidase did not show any relationship between PUD and controls, although it was associated with less intestinal metaplasia. In Chinese patients, a study of a polymorphism of histamine *N*-methyltransferase did not show any significant difference in allele frequency between DU patients and controls (Hailong et al., 2008). A study of the multidrug resistant transporter (MDR1) polymorphism in patients with PUD or gastric cancer compared to controls could not demonstrate any association (Sugimoto et al., 2008). In patients with PUD, the role of the SHP-2 gene in relation to gastric atrophy was investigated. Previous studies in Japanese subjects demonstrated that the AA genotype was associated with a lower risk of gastric atrophy, whereas in the current study in Uzbekistan, the GG genotype was linked to a lower risk of atrophy (Hamijima et al., 2008).

Management of *H. pylori*

Polymorphisms within CYP2C19 affect the plasma levels of PPIs being lower in normal homozygotes and heterozygotes (fast metabolizer) compared to mutant homozygotes (slow metabolizer) (Chaudry et al., 2008). The role of MDR1 C3435T and CYP2C19 genotypes on the eradication of *H. pylori* was investigated in Korean patients treated with pantoprazole + clarithromycin + amoxycillin (Oh et al., 2008). No effect on the eradication rates was noted. In a further study from Korea, however, the eradication rates were higher in the slow metabolizer groups compared to the fast metabolizer group (Kang et al., 2008a). A new oligonucleotide array was reported for the investigation of clarithromycin resistance in *H. pylori*. The technique was reported to be both rapid and accurate compared to conventional methods (Chen et al., 2008).

PUD and NSAIDS

The role of polymorphisms in CYP2C9 and NSAID-induced gastropathy has been investigated, but could not show that CYP2C9*2 and CYP2C9*3 are involved (Ma et al., 2008). Polymorphisms of endothelial nitric oxide synthase have been investigated in relation to gastric bleeding in persons taking low-dose aspirin. The results demonstrate that persons with a lower frequency of GPIIIa A2 have a reduced risk of bleeding (Piazuelo et al., 2008).

2009 UPDATE REFERENCES

Achyut, B.R., Ghoshal, U.C., Moorchung, N. and Mittal, B. (2009). Role of cyclooxygenase-2 functional gene polymorphisms in *Helicobacter pylori* induced gastritis and gastric atrophy. *Mol Cell Biochem* 321, 103–109.

Basso, D., Zambon, C.F., Letley, D.P., Strangres, A., Marchet, A., Rhead, J.L., Schiavon, S., Guariso, G., Ceroti, M., Nitti, D. et al. (2008). Clinical relevance of *Helicobacter pylori cagA* and *vacA* gene polymorphisms. *Gastroenterology* 135, 91–99.

Chaudry, A.S., Kochhar, R. and Kohli, K.K. (2008). Genetic polymorphism of CYp2C19 and therapeutic response to proton pump inhibitors. *Indian J Med Res* 127, 521–530.

Chen, S., Li, Y. and Yu, C. (2008). Oligonucleotide microarray: A new method for screening the 23S rRNA gene of *Helicobacter pylori* for single nucleotide polymorphisms associated with clarithromycin resistance. *J Gastroenterol Hepatol* 23(1), 126–131.

De Francesco, V., Margiotta, M., Zullo, A., Hassan, C., Giogio, F., Zotti, M., Stoppino, G., Bastianelli, A., Diterlizzi, F., Verderosa, G. et al. (2009). *Helicobacter pylori* vacA arrangement and related diseases: A retrospective study over a period of 15 years. *Dig Dis Sci* 54, 97–102.

Erzin, Y., Koksal, V., Altun, S., Dobrucali, A., Asian, M., Erdamar, S., Goksel, S., Dirican, A. and Locazeybek, B. (2008). Role of host interleukin 1 beta gene (IL-1B) and interleukin 1 receptor antagonist (IL-1RN) polymorphisms in clinical outcomes in *Helicobacter pylori* positive Turkish patients with dyspepsia. *J Gastroenterol* 43, 705–710.

Hailong, C., Mei, Q., Zhang, L. and Xu, J. (2008). C314T polymorphism in histamine N-methyltransferase gene and susceptibility to duodenal ulcer in Chinese population *Clin Chim Acta* 389, 51–54.

Hamijima, N., Rahimov, B., Malikov, Y., Adbiev, S., Ahn, K.S., Bahramov, S., Kawai, S., Nishio, K., Naito, M. and Goto, Y. (2008). Association between a PTPN11 polymorphism and gastric atrophy – opposite in Uzbekistan to that in Japan. *Asian Pac J Cancer Prev* 9, 217–220.

Hussein, N.R., Mohammadi, M., Talebkhan, Y., Doraghi, M., Letly, D.P., Muhammad, M.K., Afgent, R.H. and Atherton, J.C. (2008). Differences in virulence markers between *Helicobacter pylori* strains from Iraq and those from Iran: Potential importance of regional differences in *H. pylori*-associated disease. *J Clin Microbiol* 46, 1774–1779.

Kang, J.M., Kim, N., Lee, D.H., Park, Y.S., Kim, J.S. Chang, I.J, Song, I.S. and Jung, H.C. (2008a). Effects of the CYP2C19 polymorphism on the eradication rate of *Helicobacter pylori* infection by 7 day triple therapy with regular proton pump inhibitor dosage. *J Gastroenterol Hepatol* 23, 1287–1291.

Kang, J.M., Kim, N., Lee, D.H., Park, J.H., Lee, M.K., Kim, J.S., Jung, H.C. and Song, I.S. The effects of genetic polymorphisms of IL-6, IL-8 and Il-10 on *Helicobacter* induced gastroduodenal disease in Korea. *J Clin Gastroenterol* 2008b; [Epub ahead of print].

Ma, J., Yang, X.Y., Qiao, L., Liang, L.Q. and Chen, M.H. (2008). CYP2C9 polymorphism in non-steroidal anti-inflammatory drug-induced gastropathy. *J Dig Dis* 9, 79–83.

Mcclain, M.S., Shaffer, C.L., Israel, D.A., Peek, R.M. and Cover, T.L. (2009). Genomic sequence analysis of *Helicobacter pylori* strains associated with gastric ulceration and gastric cancer. *BMC Genomics* 10, 3.

Moura, S.B., Almeida, L.R., Guerra, J.B., Rocha, G.A., Camargos Rocha, A.M., Melo, F.F., Correa-Oliveira, R., Bittencourt, P., Carvalho, S.D. and Magalhaes Queiroz, D.M. (2008). Toll-like receptor (TLR2,TLR4 and TLR5) gene polymorphisms and *Helicobacter pylori* infection in children with and without duodenal ulcer. *Microbes Infect* 10, 1477–1483.

Oh, J.H., Dong, M.S., Choi, M.G., Yoo, H.W., Lee, S.B., Park, Y.I. and Chung, I.S. Effects of CYP2C19 and MDR1 genotype on the eradication rate of *Helicobacter pylori* infection by triple therapy with pantoprazole, amoxycillin and clarithromycin. *J Gastroenterol Hepatol* 2008 September 24; [Epub ahead of print].

Piazuelo, E., Fuentes, J., Garcfa-Gonzalez, M.A., Jimenez, P. and Lanas, A. (2008). A case control study of the association between polymorphisms of the endothelial nitric oxide synthase and glycoprotein IIIa genes and upper gastrointestinal bleeding in users of low-dose asprin. *Clin Ther* 30, 121–130.

Prasad, K.N., Saxena, A., Ghoshal, U.C., Bhagat, M. R. and Krishnani, N. (2008). Analysis of Pro12Ala PPAR gamma polymorphism and *Helicobacter pylori* infection in gastric adenocarcinoma and peptic ulcer disease. *Ann Oncol* 19, 1299–1303.

Roesler, B.M., de Oliveira, T.B., Bonon, S.H., Monici, L.T., Zeitune, J.M. and Costa, S.C. Restriction fragment length polymorphisms of urease C and urease B genes of *Helicobacter pylori* strains isolated from Brazilian patients with peptic ulcer and chronic gastritis. *Dig Dis Sci*2008 November 13; [Epub ahead of print].

Said Essa, A., Alaa Eldeen Nouh, M., Mohammed Ghaniam, N., Graham, D.Y. and Said Sabry, H. (2008). Prevalence of cagA in relation to clinical presentation of *Helicobacter pylori* infection in Egypt. *Scand J Infect Dis* 40, 730–733.

Saxena, A., Prasad, K.N., Ghoshal, U.C., Bhagat, M.R., Krishnani, N. and Husain, N. (2008). Polymorphism of -765G.C Cox-2 is a risk factor for gastric adenocarcinoma and peptic ulcer disease in addition to *H. pylori* infection: A study from North India. *World J Gastroenterol* 14, 1498–1503.

Selehi, Z., Abadi, A.S., Ismail, P.B., Kqueen, C.Y., Jelodar, M.H. and Kamalidehghan, B. Evaluation of *Helicobacter pylori vacA* genotypes in Iranian patients with peptic ulcer disease. *Dig Dis Sci* 2008 12 December; [Epub ahead of print].

Sugimoto, M., Furuta, T., Shirai, N., Kodaira, C., Nishino, M., Yamade, M., Ikuma, M., Sugimura, H., Ishizaki, T. and Hishida, A. (2008). MDR1 C3435T polymorphism has no influence on developing *Helicobacter pylori* infection related gastric cancer and peptic ulcer disease in Japanese. *Life Sci* 83, 301–304.

Tahara, T., Shibata, T., Wang, F., Nakamura, M., Okubo, M., Yoshioka, D., Sakata, M., Hirata, I. and Arisawa, T. Association of polymorphisms of the p22PHOX component of NADPH oxidase in gastroduodenal disease in Japan. *Scand J Gastroenterol* 2008 December 17; [Epub ahead of print].

Trejo-de la, O.A., Torres, J., Perez-Rodriguez, M., Camorlinga-Ponce, M., Luna, L.F., Abdo-Francis, J.M., Lazcano, E. and Maldonado-Bernal, C. (2008). TLR4 single nucleotide polymorphisms alter mucosal cytokine and chemokine patterns in Mexican patients with *Helicobacter pylori* associated gastroduodenal diseases. *Clin Immunol* 129, 333–340.

Wu, M.S., Chow, L.P., Lin, J.T. and Chiou, S.H. (2008). Proteomic identification of biomarkers related to *Helicobacter pylori*-associated gastroduodenal disease: Challenges and opportunities. *J Gastroenterol Hepatol* 23, 1657–1661.

Zhang, Z., Zheng, Q., Chen, X., Xiao, S., Liu, W. and Lu, H. (2008). The *Helicobacter pylori* duodenal ulcer promoting gene *dupA* in China. *BMC Gastroenterol* 8, 49.

ACKNOWLEDGEMENTS

I would like to thank Prof Sampurna Roy of Calcutta University, India for permission to use Figures 49.1–49.4.

REFERENCES

Alm, R.A., Ling, L.L., Moir, D.T., King, B.L., Brown, E.D., Doig, P.C., Smith, D.R., Noonan, B., Guild, B.C., deJonge, B.L. et al. (1999). Genomic-sequence comparison of two unrelated isolates of the gastric pathogen *Helicobacter pylori*. *Nature* 397, 176–180.

Amieva, M.R., Vogelmann, R., Covacci, A., Tompkins, L.S., Nelson, W.J. and Falkow, S. (2003). Disruption of the epithelial apical-junctional complex by *Helicobacter pylori* CagA. *Science* 300, 1430–1434.

Ang, S., Lee, C-Z., Peck, K., Sindici, M., Matrubutham, U., Gleeson, M.A. and Wang, J-T. (2001). Acid-induced gene expression in *Helicobacter pylori*: Study in genomic scale by microarray. *Infect Immun* 69, 1679–1686.

Atherton, J.C., Cao, P., Peek, R.M., Tummuru, M.K.R., Blaser, M.J. and Cover, T.L. (1995). Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori*. *J Biol Chem* 270, 17771–17777.

Atherton, J.C. (1997). The clinical relevance of strain types of *Helicobacter pylori*. *Gut* 40, 701–703.

Atherton, J.C., Cover, T.L., Twells, R.J., Morales, M.R., Hawkey, C.J. and Blaser, M.J. (1999). Simple and accurate PCR-based system for typing vacuolating cytotoxin alleles of *Helicobacter pylori*. *J Clin Microbiol* 37, 2979–2982.

Azuma, T., Yamakama, A., Yamazaki, S., Fukuta, K., Ohtani, M., Ito, Y., Dojo, M., Yamazaki, Y. and Kuriyama, M. (2002). Correlation between variations in the 3′ region of the gagA gene in *Helicobacter pylori* and disease outcome. *J Infect Dis* 186, 1621–1630.

Azuma, T., Yamakawa, A., Yamazaki, S., Ohtani, M., Ito, Y., Muramatsu, A., Suto, H., Yamazaki, Y., Keida, Y., Higashi, H. et al. (2004). Distict diversity of the cag pathogenicity island among *Helicobacter pylori* strains in Japan. *J Clin Microbiol* 42, 2508–2517.

Backstrom, A., Lundberg, C., Kersulyte, D., Berg, D.E., Boren, T. and Arnqvist, A. (2004). Metastability of *Helicobacter pylori* bab adhesin gene and dynamics of Lewis b antigen binding. *Proc Natl Acad Sci USA* 101, 16923–16928.

Bereswell, S., Schonenberger, R., Thies, C., Stahler, F., Strobel, S., Pfefferle, P., Wille, L. and Kist, M. (2000). New approaches for genotyping of *Helicobacter pylori* based on amplification of polymorphisms in intergenic DNA regions and at the insertion site of the cag pathogenicity island. *Med Microbiol Immunol* 189, 105–113.

Bertz, R.J. and Granneman, G.R. (1997). Use of *in vitro* and *in vivo* data to estimate the likelyhood of metabolic pharmacokinetic interactions. *Clin Pharmacokinet* 32, 210–258.

Bjorkholm, B., Lundin, A., Sillen, A., Guillemin, K., Salama, N., Rubio, C., Gordon, J.I., Falk, P. and Engstrand, L. (2001). Comparison of genetic divergence and fitness between two subclones of *Helicobacter pylori*. *Infect Immun* 69, 7832–7838.

Boneca, I.G., de Reuse, H., Epinat, J.C., Pupin, M., Labigne, A. and Moszer, I. (2003). A revised annotation and comparative analysis of *Helicobacter pylori* genomes. *Nuc Acid Res* 31, 1704–1714.

Brisslert, M., Enarsson, K., Lundin, S., Karlsson, A., Kusters, J.G., Svennerholm, A.M., Backert, S. and Quiding-Jarbrink, M. (2005). *Helicobacter pylori* induced neutrophil transendothelial migration: Role of bacterial HP-NAP. *FEMS Microbiol Lett* 249, 95–103.

Cai, X., Stoicov, C., Li, H., Carlson, J., Whary, M., Foc, J.G. and Houghton, J. (2005). Overcoming Fas-mediated apoptosis accelerates *Helicobacter*-induced gastric cancer in mice. *Cancer Res* 65, 10912–10920.

Caner, V., Yilmaz, M., Yonetci, N., Zencir, S., Karagenc, N., Kaleli, I. and Bagci, H. (2007). *H. pylori* iceA alleles are disease specific virulence factors. *World J Gastroenterol* 13, 2581–2585.

Cao, P. and Cover, T.C. (2002). Two different families of hopQ alleles in *Helicobacter pylori*. *J Clin Microbiol* 40, 4504–4511.

Censini, S., Lange, C., Xiang, Z., Crabtree, J.E., Ghiara, P., Borodovsky, M., Pappuoli, R. and Covacci, A. (1996). Cag, a pathogenicity island of *Helicobacter pylori* encodes type 1 specific and disease associated virulence factors. *Proc Natl Acad Sci USA* 93, 14648–14653.

Choe, Y.H., Kim, P.S., Lee, D.H., Kim, Y.S., Shin, Y.W., Hwang, T.S., Kim, H.J., Song, S.U. and Choi, M.S. (2002). Diverse vacA allelic types of *Helicobacter pylori* in Korea and clinical correlation. *Yonsei Med J* 43, 351–356.

Covacci, A., Falkow, S., Berg, D.E. and Rappuoli, R. (1997). Did the inheritance of pathogenicity island modify the virulence of *Helicobacter pylori*. *Trends Microbiol* 5, 205–208.

De Luca, A., Baldi, A., Russo, P., Todisco, A., Altucci, L., Giardullo, N., Pasquale, L., Iaquinto, S., D'Onofrio, V., Parodi, M.C. et al. (2003). Coexpression of *Helicobacter pylori's* proteins CagA and HspB induces cell proliferation in AGS gastric epithelial cells, independently from the bacterial infection. *Cancer Res* 63, 6350–6356.

Deng, X., Szabo, S., Khomenko, T., Jadus, M.R., Yoshida, M. and Chen, L. (2007). Detection of duodenal ulcer-associated genes in rats. *Dig Dis Sci*, Aug 4 Epub PMID 17676396.

El-Omar, E.M., Carrington, M., Chow, W.H., McColl, K.E., Bream, J.H., Young, H.A., Herrera, J., Lissowska, J., Yuan, C.C., Rothman, N. et al. (2000). Interleukin-1 polymorphisms associated with increased risk of gastric cancer. *Nature* 404, 398–402.

European Helicobacter Study Group (2005). Maastricht 3 Consensus Report, http://www.helicobacter.org.

Fischer, W., Buhrdorf, R., Gerland, E. and Haas, R. (2001). Outer membrane targeting of passenger proteins by the vacuolating cytotoxin autotransporter of *Helicobacter pylori*. *Infect Immun* 69, 6769–6775.

Fischer, W., Puls, J., Buhrdorf, R., Gebert, B., Odenbreit, S. and Haas, R. (2001). Systematic mutagenesis of the *Helicobacter pylori* cag pathogenicity island: Essential genes for CagA translocation in host cells and induction of interleukin-8. *Mol Microbiol* 42, 1337–1348.

Forsyth, M.H., Atherton, J.C., Blaser, M.J. and Cover, T.L. (1998). Heterogeneity of levels of vacuolating cytotoxin gene (vac A) transcription among *Helicobacter pylori* strains. *Infect Immun* 66, 3088–3094.

Furtura, T., Shirai, N., Xiao, F., El-Omar, E.M., Rabkin, C.S., Sugimura, H., Ishizaki, T. and Ohashi, K. (2004). Polymorphism of interleukin-1 beta affects the eradication rate of *Helicobacter pylori* by triple therapy. *Clin Gastroenterol Hepatol* 2, 22–30.

Furtura, T., Sagehashi, Y., Shirai, N., Sugimoto, M., Nakamura, A., Kodaira, M., Kenmotsu, K., Nagano, M., Egashira, T., Ueda, K. et al. (2005). Influence of CYP2C19 polymorphism and *Helicobacter* genotype determined from gastric tissue samples on response to triple therapy for *H. pylori* infection. *Clin Gastroenterol Hepatol* 3, 564–573.

Furtura, T., Sugimoto, M., Shirai, N., Matsushita, F., Nakajima, H., Kumagai, J., Senoo, K., Kodaira, C., Nishino, M., Yamade, M. et al. (2007). Effect of MDR1 C3435T polymorphism on cure rates of *Helicobacter pylori* infection by triple therapy with lansoprazole, amoxycillin and clarithromycin in relation to CYP2C19 genotypes and 23SrRNA genotypes of *H. pylori*. *Aliment Pharmacol Ther* 26, 693–703.

Galmich, A., Rassow, J., Boye, A., Cagnol, S., Chambard, J-C., Contamin, S., de Thillot, V., Just, I., Ricci, V., Solcia, E. et al. (2000). The *N*-terminal 34 kDa fragment of *Helicobacter pylori* vacuolating cytotoxin targets mitochondria and induces cytochrome C release. *EMBO J* 19, 6361–6370.

Gambero, A., Becker, T.L., Zago, A.S., de Oliveira, A.F. and Pedrazzoli, J. (2005). Comparative study of anti-inflammatory and ulcerogenic activities of different cyclo-oxygenase inhibitors. *Inflammopharmacology* 13, 441–454.

Garcia-Gonzales, M.A., Lanas, A., Santolaria, S., Crusius, J.B., Serrano, M.T. and Pena, A.S. (2001). The polymorphic IL-1B and IL-1RN genes in the aetiopathogenesis of peptic ulcer. *Clin Expt Immunol* 125, 368–375.

Garcia-Gonzales, M.A., Lanas, A., Wu, J., Benito, R., Santolaria, S., Crusius, B. and Pena, S. (2005). Lack of association of IL-12 p40 gene polymorphism with peptic ulcer disease. *Hum Immunol* 66, 72–76.

Garcia-Gonzales, M.A., Savelkoul, P.H., Benito, R., Santolaria, S., Crusius, J.B., Pena, A.S. and Lanas, A. (2005). No allelic variant association of the IL-1 and TNF gene polymorphisms in the susceptibility of duodenal ulcer disease. *Int J Immunogenet* 32, 299–306.

Garcia-Gonzalez, M.A., Strunk, M., Piazuelo, E., Benito, R., Santolaria, S., Jimenez, P., Sopena, F., Pascual, C., Simon, M.A., Sainz, R. et al. (2006). TGFB1 gene polymorphisms: Their relevance in the suceptibility to *Helicobacter pylori*-related disease. *Genes Immun* 7, 640–646.

Gatti, L.L., Fagundes e Souza, E.K., Leite, K.R., Bastos, E.L., Vicentini, L.R., Silva, L.C., Smith, M., de, A. and Payao, S.L. (2005). cagA vacA alleles and bavA2 genotypes of *Helicobacter pylori* associated with gastric disease in Brazilian adult patients. *Diag Microbiol Infect Dis* 51, 231–235.

Ge, Z. and Taylor, D.E. (1998). *Helicobacter pylori* – Molecular genetics and diagnostic typing. *Br Med Bull* 54, 31–38.

Go, M.F., Kapur, V., Graham, D.Y. and Musser, J.M. (1996). Population genetic analysis of *Helicobacter pylori* by multilocus enzyme electrophoresis: Extensive allelic diversity and recombinational population structure. *J Bacteriol* 178, 3934–3938.

Goodwin, C.S. (1994). How Helicobacter acquired its name, and how it overcomes gastric defence mechanisms. *J Gastroenterol Hepatol* 9(Suppl. 1), S1–S3.

Graham, J.E., Peek, R.M., Krishna, . and Cover, .L. (2002). Global analysis of *Helicobacter pylori* gene expression in human gastric mucosa. *Gastroenterology* 123, 1637–1648.

Gyulai, Z., Klausz, G., Tiszai, A., Lenart, Z., Kasa, I.T., Lonovics, J. and Mandi, Y. (2004). Genetic polymorphism of interleukin 8(IL-8) is associated with *Helicobacter pylori* induced duodenal ulcer. *Eur Cytokine Netw* 15, 353–358.

Haas, G., Karaali, G., Ebermayer, K., Metzger, W.G., Lamer, S., Zimny-Arndt, U., Diescher, S., Goebel, U.B., Vogt, K., Roznowski, A.B. et al. (2002). Immunoproteomics of *Helicobacter pylori* infection and related gastric disease. *Proteomics* 2, 313–324.

Halushka, M.K., Walker, L.P. and Halushka, P.V. (2003). Genetic variation in cyclooxygenase 1: Effects on response to aspirin. *Clin Pharmacol Ther* 73, 122–130.

Han, S.R., Zschausch, E.H-C., Meyer, H.G.W., Schneider, T., Loos, M., Bhakdi, S. and Maeurer, M.J. (2000). *Helicobacter pylori* clonal population structure and restricted transmission within families revealed by molecular typing. *J Clin Microbiol* 38, 3646–3651.

Harris, P.R., Cover, T.L., Crowe, D.R., Orenstein, J.M., Graham, M.F., Blaser, M.J. and Smith, P.D. (1996). *Helicobacter pylori* cytotoxin induces vacuolation of primary mucosal epithelial cells. *Infect Immun* 64, 4867–4871.

Hatakeyama, M. (2003). *Helicobacter pylori* CagA – A potential bacterial oncoprotein that functionally mimics the mammalian Gab family of adaptor proteins. *Microbes Infect* 5, 143–150.

Hazell, S.L., Andrews, R.H., Mitchell, H.M. and Daskalopoulous, G. (1997). Genetic relationship among isolates of *Helicobacter pylori*: Evidence for the existence of a *Helicobacter pylori*-species complex. *FEMS Microbiol Lett* 150, 27–32.

Hellmig, S., Mascheretti, S., Renz, J., Frenzel, H., Jelschen, F., Rehbein, J.K., Folsch, U., Hampe, J. and Schreiber, S. (2005). Haplotype analysis of the CD11 gene cluster in patients with chronic *Helicobacter pylori* infection and gastric ulcer disease. *Tissue Antigens* 65, 271–274.

Hellmig, S., Ott, S., Rosenstiel, P., Folsch, U.R., Hampe, J. and Schreiber, S. (2006). Genetic varients in matrix metalloproteinase genes are associated with development of gastric ulcer in *H. pylori* infection. *Am J Gastroenterol* 101, 29–35.

Hennig, E.E., Mernaught, R., Edl, J., Cao, P. and Cover, T.L. (2004). Heterogeneity among *Helicobacter pylori* strains in expression of the outer membrane protein BabA. *Infect Immun* 72, 3429–3435.

Hofner, P., Gyulai, Z., Kiss, Z.F., Tiszai, A., Tiszlavicz, L., Toth, G., Szoke, D., Molnar, B., Lonovics, J., Tulassay, Z. et al. (2007). Genetic polymorphism of NOD1 and IL-* but not polymorphism of TLR4 genes are associated with *Helicobacter pylori*-induced duodenal ulcer and gastritis. *Helicobacter* 12, 124–131.

Hsu, P.I., Li, C.N., Tseng, H.H., Lai, K.H., Hsu, P.N., Lo, G.H., Lo, C.C., Yeh, J.J., Ger, L.P., Hsiao, M. et al. (2004). The interleukin-1 RN polymorphism and *Helicobacter pylori* infection in the development of duodenal ulcer. *Helicobacter* 9, 605–613.

Hsu, P.I., Jwo, J.J., Tseng, H.H., Lai, K.H., Lo, G.H., Lo, C.C., Wu, C.J., Chuah, S.K., Hwang, I.R., Chen, J.L. et al. (2005). Association of the myloperoxidase – 468G > A polymorphism with gastric inflammation and duodenal ulcer risk. *World J Gastroenterol* 11, 2796–2801.

Israel, D.A., Salama, N., Arnold, C.N., Moss, A.S., Ando, T., Wirth, H-P., Tham, K.T., Camorlinga, M., Blaser, M.J., Falkow, S. et al. (2001). *Helicobacter pylori* strain-specific differences in genetic content identified by microarray influence host inflammatory responses. *J Clin Invest* 107, 611–620.

Israel, D.A., Salama, N., Krishna, U., Rieger, U.M., Atherton, J.C., Falkow, S. and Peek, R.M. (2001). *Helicobacter pylori* genetic diversity within the gastric niche of a single human host. *Proc Natl Acad Sci USA* 98, 14625–14630.

Ito, Y., Azuma, T., Ito, S., Suto, H., Miyaji, H., Yamazaki, Y., Kato, T., Kohli, Y., Keida, Y. and Kuriyama, M. (2000). Sequence analysis and clinical significance of the iceA gene from *Helicobacter pylori* in Japan. *J Clin Microbiol* 38, 483–488.

Ji, X., Fernandez, T., Burroni, D., Pagiaccia, C., Atherton, J.C., Reyrat, J-M., Rappuoli, R. and Telford, J.L. (2000). Cell specific *Helicobacter pylori* cytotoxin is determined by a short region in the polymorphic mid-region. *Infect Immun* 68, 3754–3757.

Kersulyte, D., Mukhopadhyay, A.K., Velapatino, B., Su, W., Pan, Z., Garcia, C., Hernandez, V., Valdez, Y., Mistry, R.S., Gilman, R.H. et al. (2000). Different genotypes of *Helicobacter pylori* from different human populations. *J Bacteriol* 182, 3210–3218.

Krah, A., Schmidt, F., Becher, D., Schmidt, M., Albrecht, D., Rack, A., Buttner, K. and Jungblutt, PR. (2003). Analysis of automatically generated peptide mass fingerprints of cellular proteins and antigens from *Helicobacter pylori* 26695 separated by two-dimentional electrophoresis. *Mol Cell Proteomic* 2, 1271–1283.

Kubota, Y., Kato, K., Dairaku, N., Koike, T., Iijima, K., Imatani, A., Sekine, H., Ohara, S., Matsui, H. and Shimosegawa, T. (2004). Contribution of glutamate synthetase to ammonia-induced apoptosis in gastric mucosal cells. *Digestion* 69, 140–148.

Kunstmann, E., Hardt, C., Elitok, E., Harder, M., Suerbaum, S., Peitz, U., Schmiegel, W. and Epplen, J. (2000). The non-functional allele TCRBV6S1B is stronly associated with *Helicobacter pylori* infection. *Infect Immun* 68, 6493–6495.

Kuo, C.H., Wu, D.C., Lu, C.Y., Su, Y.C., Yu, F.J., Lee, Y.C., Wu, I.C., Lin, S.R., Liu, C.S., Jan, C.M. et al. (2003). Low moleculsar weight protein of *Helicobacter pylori* and its relation to gastroduodenal diseaese. *Hepatogastroenterology* 50, 897–901.

Lanas, A., Garcia-Gonzales, M.A., Santolaria, S., Crusius, J.B., Serrano, M.T., Benito, R. and Pena, A.S. (2001). TNF and LTA gene polymorphism reveal different risk in gastric and duodenal ulcer patients. *Genes Immun* 2, 415–421.

Lehmann, D.F., Medicis, J.J. and Franklin, P.D. (2003). Polymorphisms and the pocketbook: The cost effectiveness of cytochrome P450 2C19 genotyping in the eradication of *Helicobacter pylori* infection associated with duodenal ulcer. *J Clin Pharmacol* 43, 1316–1323.

Lo, C.C., Hsu, P.I., Lo, G.H., Lai, K.H., Tseng, H.H., Lin, C.K., Chan, H.H., Tsai, W.L., Chen, W.C. and Peng, N.J. (2005). Implications of anti-parietal cell antibodies and anti-*Helicobacter pylori* antibodies in histological gastritis and patient outcome. *World J Gastroenterol* 11, 4715–4720.

Lobo Gatti, L., Zanbaldi Tunes, M., de Labio, R.W., Silva, L.C., de Arruda Cardoso Smith, M. and Marques Payao, S.L. (2005). Interleukin 6 polymorphism and *Helicobacter pylori* infection in Brazilian adult patients with chronic gastritis. *Clin Expt Med* 5, 112–116.

Lu, C.C., Sheu, B.S., Chen, T.W., Yang, H.B., Hung, K.H., Kao, A.W., Chuang, C.H. and Wu, J.J. (2005). Host TNF-alpha -1031 and -863

promotor single nucleotide polymorphisms determine the risk of benign ulceration after *H pylori* infection. *Am J Gastroenterol* 100, 1274–1282.

Lundin, A., Bjorkholm, B., Kuperschmidt, I., Unemo, M., Nilsson, P., Andersson, D.I. and Engstrand, L. (2005). Slow genetic divergence of *Helicobacter pylori* strains during long term colonisation. *Infect Immun* 73, 4818–4822.

Lupetti, P., Heuser, J.E., Manetti, R., Massari, P., Lanzavecchia, S., Bellon, PL., Dallai, R., Rappuoli, R. and Telford, J.L. (1996). Oligomeric and subunit structure of the *Helicobacter pylori* vacuolating cytotoxin. *J Cell Biol* 133, 801–807.

Megraud, F. (2005). A humble bacterium sweeps this year's Nobel Prize. *Cell* 123, 975–976.

Moss, S.F., Legon, S., Bishop, A.E., Polak, J.M. and Calam, J. (1992). Effects of *Helicobacter pylori* on gastric somatostatin in duodenal ulcer disease. *Lancet* 340, 930–932.

Mukhopadhyay, A.K., Kersulyte, D., Jeong, J.-Y., Datta, S., Ito, Y., Chowdhury, A., Chowdhury, S., Santra, A., Bhattacharya, S.K., Azuma, T. et al. (2000). Distinctiveness of genotypes of *Helicobacter pylori* in Calcutta, India. *J Clin Microbiol* 182, 3219–3227.

Nagano, Y., Matsui, H., Muramatsu, M., Shimokawa, O., Shibahara, T., Yanaka, A., Nakahara, A., Matsuzaki, Y. and Tanka, N. (2005). Rebamipide significantly inhibits indomethacin induced mitochondrial damage, lipid peroxidation and apoptosis in gastric epithelial RGM-1 cells. *Dig Dis Sci* 50(Supp 1), S76–S83.

Ohyauchi, M., Imatani, A., Yonechi, M., Asano, N., Miura, A., Iijima, K., Koike, T., Sekine, H., Ohara, S. and Shimosegawa, T. (2005). The polymorphism interleukin 8 -251A/T influences the susceptibility of *Helicobacter pylori* related gastic disease in the Japanese population. *Gut* 54, 330–335.

Olfat, F.O., Zheng, Q., Oleastro, M., Voland, P., Boren, T., Karttunen, R., Engstrand, L., Rad, R., Prinz, C. and Gerhard, M. (2005). Correlation of the *Helicobacter pylori* adherence factor babA with duodenal ulcer disease in four European countries. *FEMS Immunol Med Microbiol* 44, 151–156.

Pagliaccia, C., de Bernard, M., Lupetti, P., Ji, X., Burroni, D., Cover, T.L., Papini, E., Rappuoli, R., Telford, J.L. and Reyrat, J.M. (1998). The m2 form of the *Helicobacter pylori* cytotoxin has cell type-specific vacuolating activity. *Proc Natl Acad Sci USA* 95, 10212–10217.

Pan, Z-J., Berg, D.E., van der Hulst, R.W.M., Su, W.W., Raudonikiene, A., Xiao, D.D., Dankert, J., Tytgat, G.N. and van der Ende, A. (1998). Prevalence of vacuolating cytotoxin production and distribution of distinct vacA alleles in *Helicobacter pylori* from China. *J Infect Dis* 178, 220–226.

Peek, R.M., Thompson, S.A., Donahue, J.P., Tham, K.T., Atherton, J.C., Blaser, M.J. and Miller, G.G. (1998). Adherance to gastric epithelial cells induces expression of a *Helicobacter pylori* gene, iceA that is associated with clinical outcome. *Proc Assoc Am Physician* 110, 531–544.

Pereira, D.R.B., Martins, D., Winck, F.V., Smolka, MB., Nishimura, N.F., Rabelo-Goncalves, E.M.A., Hara, N.H., Marangoni, S., Zeitune, J.M.R. and Novello, J.C. (2006). Comparative analysis of two dimentional electrophoresis maps ('-DE) of *Helicobacter pylori* from Brazilian patients with chronic gastritis and duodenal ulcer: A preliminary report. *Rev Inst Med Trop S Paulo* 48, 175–177.

Pilotto, A., Seripa, D., Franceschi, M., Scarcelli, C., Colaizzo, D., Grandone, E., Niro, V., Andriulli, A., Leandro, G., Di Mario, F. et al. (2007). Genetic susceptibility to nonsteroidal anti inflammatory drug related gastroduodenal bleeding : Role of cytochrome P450 2C9 polymorphisms. *Gastroenterology* 133, 465–471.

Rad, R., Gerhard, M., Lang, R., Schoniger, M., Rosch, T., Schepp, W., Becker, I., Wagner, H. and Prinz, C. (2002). The *Helicobacter pylori* blood group antigen binding adhesin facilitates bacterial colonisation and augments a non-specific immune response. *J Immunol* 168, 3033–3041.

Rad, R., Prinz, C., Neu, B., Neuhofer, M., Zeitner, M., Voland, P., Becker, I., Schepp, W. and Gerhard, M. (2003). Synergistic effect of *Helicobacter pylori* virulence factors and interleukin-1 polymorphisms for the development of severe histological changes in the gastric mucosa. *J Infect Dis* 188, 272–281.

Rota, C.A., Pereira-Lima, J.C., Blaya, C. and Nardi, N.B. (2001). Consensus and variable region PCR analysis of *Helicobacter pylori* 3′ region of cagA gene in isolates from individuals with and without peptic ulcer. *J Clin Microbiol* 39, 606–612.

Salama, N., Guillemin, K., McDaniel, T.K., Sherlock, G., Tomkins, L. and Falkow, S. (2000). A whole-genome microarray reveals genetic diversity among *H. pylori strains*. *Proc Natl Acad Sci USA* 97, 14668–14673.

Santos, A., Queiroz, M.M., Menard, A., Marais, A., Rocha, G.A., Oliveira, C.A., Margardia, A., Nogueira, M.F., Uzeda, M. and Megraud, F. (2003). New pathogenicity marker found in the plasticity region of the *Helicobacter pylori* genome. *J Clin Microbiol* 41, 1651–1655.

Sapone, A., Vaira, D., Trespidi, S., Perna, F., Gatta, L., Tampieri, A., Ricci, C., Cantelli-Forti, G., Miglioli, M., Biagi, G.L. et al. (2003). The clinical role of cytochrome P450 genotypes in *Helicobacter pylori* management. *Am J Gastroenterol* 98, 1010–1015.

Strobel, S., Bereswill, S., Balig, P., Allgaier, P., Sonntag, H-G. and Kist, M. (1998). Identification and analysis of new vacA genotypes of *Helicobacter pylori* in different patient groups in Germany. *J Clin Microbiol* 36, 1285–1289.

Suerbaum, S., Smith, J.M., Bapumia, K., Morelli, G., Smith, N.H., Kunstmann, E., Dyrek, I. and Achtman, M. (1998). Free recombination within *Helicobacter pylori*. *Proc Natl Acad Sci USA* 95, 12619–12624.

Take, S., Mizuno, M., Ishiki, K., Nagahara, Y., Yoshida, T., Inaba, T., Yamamoto, K., Okada, H., Yokota, K., Oguma, K. et al. (2003). Interleukin-1 beta gentic polymorphism influences the effect of cytochrome P 2C19 genotype on the cure rate of 1-week triple therapy for *Helicobacter pylori* infection. *Am J Gastroenterol* 98, 2403–2408.

Tan, H.J., Rizal, A.M., Rosmadi, M.Y. and Goh, K.L. (2005). Distribution of *Helicobacter pylori* cagA, cagE and vacA in different ethnic groups in Kuala Lumpa, Malaysia. *J Gastroenterol Hepatol* 20, 589–594.

Thoreson, A.-C.E., Hamlet, A., Celik, J., Bystrom, M., Nystrom, S., Olbe, L. and Svennerholm, A-M. (2000). Differences in surface exposed antigen expression between *Helicobacter pylori* strains isolated from duodenal ulcer patients and from asymptomatic subjects. *J Clin Microbiol* 38, 3436–3441.

Tomb, J.F., White, O., Kerlavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S., Dougherty, B.A. et al. (1997). The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 388, 539–547.

van der Ende, A., Pan, Z.-J., Bart, A., van der Hulst, R.M.W., Feller, M., Xiao, S.-D., Tytgat, G.N.J. and Dankert, J. (1998). cagA positive *Helicobacter pylori* populations in China and the Netherlands are distinct. *Infect Immun* 66, 1822–1826.

van Doorn, L.-J., Figueiredo, C., Sanna, R., Pena, S., Modolo, P., Ng, E.K.W., Atherton, J.C., Blaser, M.J. and Quint, W.G.V. (1998). Expanding alleleic diversity of *Helicobacter pylori* vacA. *J Clin Microbiol* 36, 2597–2603.

van Doorn, L.-J., Figueiredo, C., Sanna, R., Plaisier, A., Schneeberger, P., Boer, W.D. and Quint, W. (1998). Clinical relevance of the cagA, vacA and iceA status of *Helicobacter pylori*. *Gastroenterology* 115, 58–66.

van Doorn, L.-J., Figueiredo, C., Megraud, F., Pena, S., Midolo, P., de Magalhaes Queiroz, M., Carneiro, F., Vanderborght, B., Da Gloria, F., Pegado, M. et al. (1999). Geographic distribution of vacA allelic types of *Helicobacter pylori*. *Gastroenterology* 116, 823–830.

van Doorn, L.-J., Figueiredo, C., Sanna, R., Blaser, M.J. and Quint, W. G.V. (1999). Distinct variants of *Helicobacter pylori* cagA are associated with vacA sub-types. *J Clin Microbiol* 37, 2306–2311.

Vilaichone, R.K., Mahachai, V., Kositchaiwat, C., Graham, D.Y. and Yamaoka, Y. (2003). Relation between seroreactivity to low molecular weight *Helicobacter pylori*-specific antigens and disease presentation. *Clin Diagn Lab Immunol* 10, 1025–1028.

Wang, G., Humayun, M.Z. and Taylor, D.E. (1999). Mutation as an origin of genetic variability in *Helicobacter pylori*. *Trends Microbiol* 7, 488–493.

Wang, H.J., Kuo, C.-H., Yeh, A.A.M., Chang, P.C.L. and Wang, W.-C. (1998). Vacuolating toxin production in clinical isolates of *Helicobacter pylori* with different vacA genotypes. *J Infect Dis* 178, 207–212.

Warren, J.R. and Marshall, B.J. (1983). Unidentified curved bacilli on gastric epithelium in active chronic gastritis. *Lancet* 1, 1273–1275.

Wilschanski, M., Schlesinger, Y., Faber, J., Rudensky, B., Ohnona, F.S., Freier, S., Rahman, E., Refael, S. and Halle, D. (2007). Combination of *Helicobacter pylori* strain and tumor necrosis factor -alpha polymorphism of the host increases the risk of peptic ulcer disease in children. *J Pediatr Gastroenterol Nutr* 45, 199–203.

Xiang, Z., Censini, S., Bayeli, P.F., Telford, J.L., Figura, N., Rappuoli, R. and Covacci, A. (1995). Analysis of expression of CagA and VacA virulence factors in 43 strains of *Helicobacter pylori* reveals that clinical isolates can be divided into two major types and that cagA is not necessary for expression of the vacuolating cytotoxin. *Infect Immun* 63, 94–98.

Yamaoka, Y., Kodama, T., Graham, DY. and Kashima, K. (1998). Search for putative virulence factors of Helicobacter pylori: The low molecular weight (33–35 K) antigen. *Dig Dis Sci* 43, 1482–1487.

Yamaoka, Y., Kodama, T., Kashima, K., Graham, D.Y. and Sepulveda, A.R. (1998b). Variants of the 3′ region of the cagA gene in *Helicobacter pylori* isolates from patients with different *H. pylori* associated diseases. *J Clin Micro* 36, 2258–2263.

Yamaoka, Y., El-Zimaity, H.M.T., Gutierrez, O., Figura, N., Kim, J.K., Kodama, T., Rappuoli, R. and Covvaci, A. (1999). Relationship between CagA 3′ repeat region of *Helicobacter pylori*, gastric histology and susceptibility to low *p*H. *Gastroenterology* 117, 342–349.

Yamaoka, Y., Kodama, T., Gutierrez, O., Kim, J.G., Kashima, K. and Graham, D.Y. (1999). Relationship between *Helicobacter pylori* iceA, cagA and vacA status and clinical outcome: Studies in four different countries. *J Clin Microbiol* 37, 2274–2279.

Yamaoka, Y., Kwon, D.H. and Graham, D.Y. (2000). A M(r) 34,000 proinflammatory outer membrane protein (oipA) of *Helicobacter pylori*. *Proc Natl Acad Sci USA* 97, 7533–7538.

Yamaoka, Y., Osato, M.S., Sepulveda, A.R., Gutierrez, O., Figura, N., Kim, J.G., Kodama, T., Kashima, K. and Graham, D.Y. (2000). Molecular epidemiology of *Helicobacter pylori*: Separation of *H. pylori* from East Asian and non-Asian countries. *Epidemiol Infect* 124, 91–96.

Yamaoka, Y., Souchek, J., Odenbreit, S., Haas, R., Arnqvist, A., Boren, T., Kodama, T., Osato, M.S., Gutierrez, O., Kim, J.G. et al. (2002). Discrimination between cases of duodenal ulcer and gastritis on the basis of putative virulence factors of *Helicobacter pylori*. *J Clin Microbiol* 40, 2244–2246.

Yamazaki, S., Yamakawa, A., Okuda, T., Ohtani, M., Suto, H., Ito, Y., Yamazaki, Y., Keida, Y., Higashi, H., Hatakeyama, M. et al. (2005). Distinct diversity of vacA, cagA, cagE genes of *Helicobacter pylori* associated with peptic ulcer in Japan. *J Clin Microbiol* 43, 3906–3916.

RECOMMENDED RESOURCES

Websites

http://www.tigr.org/tigr-scripts/CMR2gene_table.spl?db=ghp

As given on the website: The Comprehensive Microbial Resource (CMR) is a free website used to display information on all of the publicly available, complete prokaryotic genomes.

http://www.helicobacter.org

The website of the European Helicobacter Study Group

http://genolist.pasteur.fr/PyloriGene

As given on the website: The purpose is to collate and integrate various aspects of the genomic information from *H. pylori*. PyloriGene provides a complete dataset of DNA and protein sequences derived from two different strains: 26695 and J99, linked to the relevant annotations and functional assignments.

http://ecocyc.org:1555/HPY/organism-summary?object=HPY

As given on the website: EcoCyc is a bioinformatics database that describes the genome and the biochemical machinery of *E. coli* K-12 MG1655. The long-term goal of the project is to describe the molecular catalog of the *E. coli* cell, as well as the functions of each of its molecular parts, to facilitate a system-level understanding of *E. coli*. EcoCyc is an electronic reference source for *E. coli* biologists, and for biologists who work with related microorganisms.

http://www.histopathology-india.net/PepUlc.htm

The gastroenterology website of Prof Sampurna Roy.

Book

Axon, A. (ed). *Best Practice & Research: Clinical gastroenterology*. vol. 21(2) *Helicobacter pylori*, Elsevier.

Description from the Elsevier website: In practical paperback format, each 200 page topic-based issue of *Best Practice & Research Clinical Gastroenterology* will provide a comprehensive review of current clinical practice and thinking within the specialty of gastroenterology.

Clinical (Metabolic Disease)

Section 7

50. Genomics in Pathogenesis of Cirrhosis

51. Genomic Medicine and Obesity

52. Diabetes

This page intentionally left blank

CHAPTER

Genomics in Pathogenesis of Cirrhosis

N.A. Shackel, K. Patel and J. McHutchison

INTRODUCTION

The liver has been called the "the custodian of the milieu interieur." Consistent with its many varied metabolic functions, the liver has a complex transcriptome and proteome. The normal liver has many diverse functions including synthesis of vitamins and proteins, bile production, immune defense, as well as metabolism of carbohydrates, lipids and toxins. Despite being only 2.5% of body weight, the liver receives 25% of the cardiac output that is essential for maintaining its metabolic and synthetic functions. The functional unit of the liver is the hepatic lobule, which is arranged in an organized repeating fashion around a central venule to form the intact organ. Liver injury is characterized by progressive fibrosis tissue deposition within the lobule, leading to eventual disruption of normal lobular architecture that is characteristic of cirrhosis (Figure 50.1) (Friedman, 2000). In the genomic era, the molecular classification of fibrosis and cirrhosis development is predominantly characterized by morphological changes.

Cirrhosis is the pathogenic hallmark of advanced liver injury. Cirrhosis is morphologically defined by distortion of hepatic architecture by dense bands of fibrosis "scar" leading to "islands" or nodules of hepatocytes (Friedman, 2000). The causes of cirrhosis are varied with many etiologies (Table 50.1). The development of cirrhosis is a premalignant condition, with virtually all cases of liver cancer (also known as hepatocellular carcinoma [HCC]) developing in individuals with prior cirrhosis. Various intrahepatic cell populations are essential in understanding the development of cirrhosis. The pivotal cell involved in fibrosis leading to cirrhosis is the hepatic stellate cell (HSC). However, the functional unit of the liver is the hepatocyte, which is the cell type from which HCC develops. The mechanisms of fibrosis leading to cirrhosis will be discussed in the context of understanding the pathogenic mechanisms by which this process evolves. Genomic medicine promises new insights into the pathogenesis of fibrosis and the promise of individualized predictive medicine aiming to prevent cirrhosis and the sequelae of liver failure and HCC.

FIBROSIS AND CIRRHOSIS

Despite the adoption of genomic medicine approaches in recent years, the hallmark of liver injury remain the morphological changes. In end-stage fibrosis there is "encapsulation" of islands of hepatocytes by bands of fibrosis, known as cirrhosis (Figure 50.1). The hepatic fibrotic response represents a wound healing response with morphological features of matrix remodeling, contraction and "scaring" (Eng and Friedman, 2000). The HSC is central to this fibrogenic process (Figure 50.1). The fibrogenic process can be regarded in stages based on morphology or on molecular pathogenic events (Friedman, 2000). Fibrosis of the liver has classically been regarded as an irreversible disease process. However, it is now clear that advanced fibrosis, possibly even cirrhosis, can significantly improve and possibly resolve completely in some cases (Benyon and Iredale, 2000; Corbett et al., 1993; Dufour et al., 1998; Tsushima, et al., 1999). Molecular studies of intrahepatic fibrogenesis in progressive injury indicated that this process is dynamic and characterized by distinct events associated with initiation, perpetuation and regression (Bataller

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

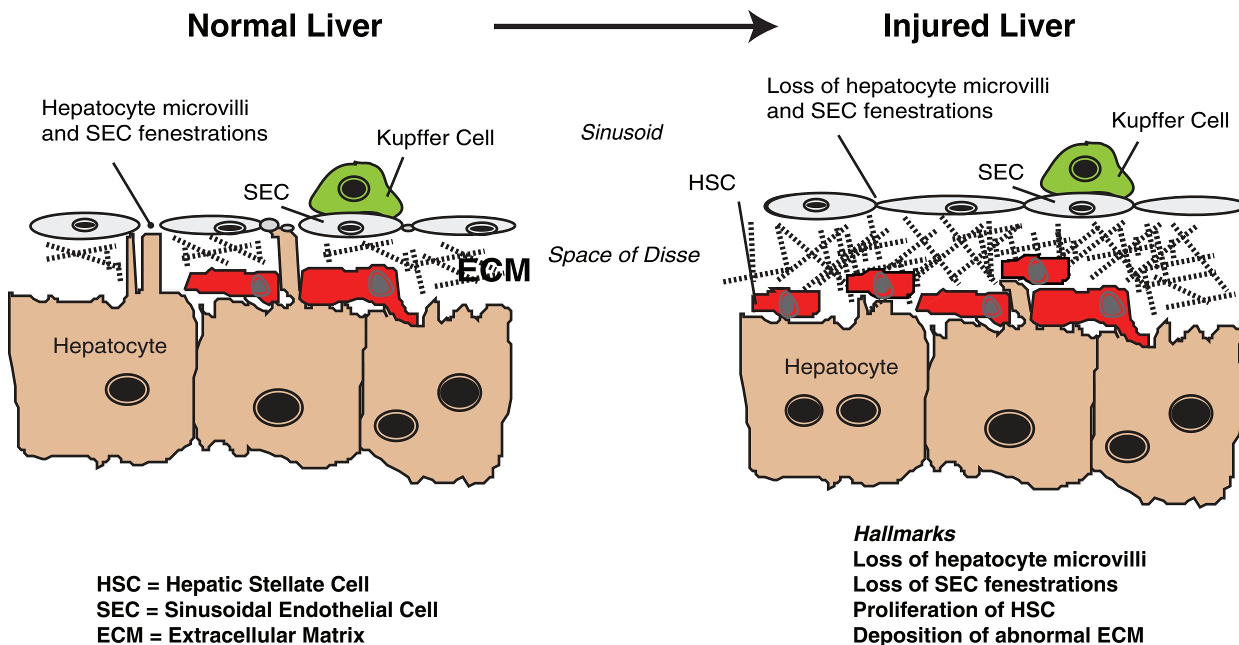


Figure 50.1 Liver fibrogenesis. In liver fibrosis development the quiescent stellate cell is transformed into an activated phenotype with ECM changes characterized by the deposition of a scar-like matrix within the space of Disse. This change is accompanied by a loss of hepatocyte microvilli and endothelial fenestration.

and Brenner, 2005). The remodeling of the ECM in fibrogenesis is clearly a dynamic process, and an improved understanding of the molecular pathways involved in this process will help in developing future targeted therapeutic agents.

Liver injury characterized by fibrosis is extremely rare in the absence of intrahepatic inflammatory changes. The evolution of both innate and adaptive immune response in chronic liver injury is an essential factor perpetuating and driving intrahepatic injury resulting in liver fibrogenesis. Therefore, the separate study of fibrogenic pathways characterized by ECM remodeling and intrahepatic inflammatory response is often completely arbitrary.

DIAGNOSIS OF CIRRHOSIS

An accurate assessment of cirrhosis and the preceding evolution of intrahepatic fibrosis is a frequent diagnostic challenge. Routine biochemical analysis, as well as clinical signs, are at best suggestive but not diagnostic of the grade of fibrosis or the development of cirrhosis. To-date, liver biopsy has been the principal approach to the assessment of fibrosis. However, liver biopsy is an invasive procedure that is prone to sampling error, significant observer variability and may be associated with morbidity and mortality (Siegel et al., 2005). This has led to the development of non-invasive measures of fibrosis including the appraisal of panels of profibrogenic serum markers and assessment of intrahepatic elasticity or 'stiffness' changes using modified ultrasonography (elastometry). Other imaging techniques of the liver with ultrasound, computer tomography or magnetic resonance imaging can provide indicators of cirrhosis, but interpretation is often operator-dependent and lacks sensitivity for accurate staging of advanced disease. However, imaging and biochemistry may be useful in providing indicators of worsening hepatic function from cirrhosis, such as ascites or the development of HCC. To-date there are no established molecular markers of fibrosis progression or cirrhosis.

GENETICS OF CIRRHOSIS

Cirrhosis pathogenesis is characterized by many conserved as well as divergent pathways involved in the development of progressive fibrosis. There are clearly a number of important host genetic factors that influence the development of fibrosis, rate of progression and the subsequent development of complications (Bataller et al., 2003). Older individuals and males are known to have significantly greater rates of liver fibrogenesis and subsequent development of cirrhosis (Poynard et al., 2005). Multiple candidate genes thought to be central to cirrhosis development have been identified using rodent models, and many human gene polymorphisms are likewise associated with fibrosis development (Table 50.2) (Bataller et al., 2003). Genetic polymorphisms have been implicated at all stages of fibrosis development including disease susceptibility (i.e., HCV and haptoglogin), injury (i.e., LPS interaction with CD14), immune responses (i.e., HLA-II alleles), activation of HSC (i.e., angiotensinogen) and increase in fibrogenic ECM (i.e., plasminogen and TGF-β) (Bataller et al., 2003). The single most comprehensive analysis of genetic polymorphisms

Table 50.1 Causes of cirrhosis and/or chronic liver disease

Infectious diseases
Brucellosis
Capillariasis
Echinococcosis
Schistosomiasis
Toxoplasmosis
Viral hepatitis (hepatitis B, C, D; cytomegalovirus; Epstein–Barr virus)
Inherited and metabolic disorders
α1-Antitrypsin deficiency
Alagille's syndrome
Biliary atresia
Familial intrahepatic cholestasis (FIC) types 1–3
Fanconi's syndrome
Galactosemia
Gaucher's disease
Glycogen storage disease
Hemochromatosis
Hereditary fructose intolerance
Hereditary tyrosinemia
Wilson's disease
Drugs and toxins
Alcohol
Amiodarone
Arsenicals
Oral contraceptives (Budd-Chiari)
Pyrrolidizine alkaloids (Veno-occlusive disease)
Other causes
Biliary obstruction (chronic)
Cystic fibrosis
Graft-versus-host disease
Jejunoileal bypass
Non-alcoholic steatohepatitis (NASH)
Primary biliary cirrhosis
Primary sclerosing cholangitis
Sarcoidosis
Causes of non-cirrhotic hepatic fibrosis
Idiopathic portal hypertension (Non-cirrhotic portal fibrosis, Banti's syndrome); three variants: Intrahepatic phlebosclerosis and fibrosis; Portal and splenic vein sclerosis and Portal and splenic vein thrombosis
Schistosomiasis ("*pipe-stem*" fibrosis with pre-sinusoidal portal hypertension)
Congenital hepatic fibrosis (may be associated with polycystic disease of liver and kidneys)

in fibrosis analyzed 24,832 putative functional, single nucleotide polymorphisms (SNPs) in 916 individuals with HCV infection. This study identified a missense mutation in the DEAD box polypeptide 5 (DDX5) gene associated with two POLG2 SNPs that are linked with an increased risk of advanced fibrosis (Huang et al., 2006). The same study identified another missense SNP in carnitine palmitoyltransferase 1A (CPT1A) associated with a decreased risk of fibrosis. The fibrosis association of these polymorphisms was validated in a separate cohort of 483 individuals (Huang et al., 2006). However, it is unclear what the function of these genes is in HCV induced liver injury. Further, it is unclear if the identified polymorphisms are important only in HCV liver injury or more generally in other types of liver injury.

Table 50.2 Candidate gene and polymorphisms involved in cirrhosis

Gene	Candidate gene	Gene polymorphism	Effect of protein on fibrosis
ADH		√	Unknown
ALDH		√	Unknown
Angiotensinogen		√	Increases
ApoE		√	Increases
CD14		√	Increases
CPT1A	√		Decreases
CTLA-4		√	Increases
CYP2E1		√	Unknown
DDX5		√	Increases
Fas	√		Increases
HFE		√	Unknown
HLA-II Haplotypes		√	Variable
IFN-γ	√		Decreases
IL-1 receptor		√	Increases
IL-10	√	√	Decreases
IL-13	√		Unknown
IL-1β		√	Increases
IL-6	√		Unknown
Leptin	√		Increases
MnSOD		√	Increases
NOS-2	√		Decreases
OB-R	√		Increases
Plasminogen	√		Decreases
SMAD-3	√		Increases
TAP2		√	Unknown
Telomerase	√		Decreases
TGF-β1	√	√	Increases
TIMP-1	√		Increases
TNF	√		Unknown

The identification of candidate genes and polymorphisms has seen a number of novel molecular approaches adopted in an attempt to treat liver disease (Prosser et al., 2006). These have included the use of caspase inhibitors, TGF-β blockade, TNF blockade, PDGF blockade and inhibition of Kupffer cell activity (Prosser et al., 2006). Further approaches have attempted to change the ECM composition in fibrosis via the administration of matrix metalloproteinases and plasminogen activator (Prosser et al., 2006). Although modern molecular approaches have identified multiple genes involved in the pathogenesis of cirrhosis, most of these findings have yet to lead to therapeutic agents in the clinical trial setting.

The identification of genetic loci associated with the development of progressive liver injury has focused on the underlying disease rather than the common pathway leading to the development of fibrosis and eventual cirrhosis (Table 50.3). The immunogenetics of autoimmune liver disease and viral hepatitis are partially characterized with a number of genetic loci being associated with disease severity and/or progression (Donaldson, 2004). However, the immunogenetics of other forms of liver injury is poorly characterized. To-date the precise association of many of these markers with liver disease severity remains to be determined. However, in the future the determination of the genetic haplotypes conferring susceptibility to fibrosis or predicting the progression of liver disease is likely to be incorporated into routine clinical practice.

THE LIVER TRANSCRIPTOME

Evolving genomic medicine approaches to liver disease require an understanding of the complexity of genome expression within the normal and diseased liver. There are estimated to be approximately 20–25,000 protein-encoding genes in the human genome. Further, there are an unknown number of functionally significant alternately spliced transcripts arising from these genes that may exceed 100,000 in number. How many of these mRNA transcripts are expressed in the liver is unknown. Methods of identifying and comparing organ transcriptomes are also uncommon. One method of inferring complexity is to examine GenBank human UniGene clusters of non-redundant gene sets (Yuan et al., 2001). These UniGene clusters are compiled from annotated and uncharacterized mRNA sequences and as a group represent a species' transcriptome (Yuan et al., 2001). Currently the human UniGene assembly of clusters (Build 196) has approximately 7 million sequences representing 83,896 non-redundant transcripts. Parsing key word searches[1], approximately 20% of transcripts (representing 16,950 clusters) were identified in liver tissue; this compares to brain (37%), lung (31%), kidney (25%), colon (22%) and heart (16%). Coulouarn et al. (2004) used a similar approach and identified 12,638 non-redundant clusters from liver tissue (UniGene Build 129). An alternate approach in which Serial Analysis of Gene Expression (SAGE) libraries were examined can also provide insights into the complexity of the liver transcriptome (Yamashita et al., 2000, 2004a, b). Two normal human liver SAGE libraries identified 15,496 and 18,081 unique transcripts from a total number of 66,308 and 125,700 tags, respectively (Yamashita et al., 2000, 2004a, b). However, in a SAGE comparison of multiple organs, 32,131 unique tags were identified (from a total of 455,325 tags) of which 56% were expressed in the liver compared to brain (75%), breast (81%) and colon (91%) (Yamashita et al., 2004a, b). Therefore, it is clear that the normal liver has a complex transcriptome expressing many thousands of mRNA transcripts.

Normal liver transcriptome expression varies to account for phenotype differences such as sex and age variation (Cao et al., 2001; Tadic et al., 2002; Yang et al., 2006). Microarray analysis of normal human liver by Yano et al. (2001) highlights the variability of the non-diseased liver transcriptome. A total of 2418 genes were examined in five normal patients, with only half of these transcripts being detected in four out of five patients. Further, only 27% of genes had co-ordinate expression in these non-diseased, apparently normal patients. Therefore, in addition to the liver having a complex transcriptome, there appears to be significant individual variability in transcript expression. This is further highlighted by the observation of Enard et al. (2002) that duplicate liver samples from the same individual differed by 12% (technical variation) but that intraspecies variation was as pronounced as interspecies variation in hepatic mRNA transcript expression when comparing chimpanzees and humans. Focused specialized arrays such as the Liverpool nylon array targeting the liver transcriptome have now been synthesized and include in excess of 10,000 target genes (Coulouarn et al., 2004). However, such approaches fail to detect differential gene expression for transcripts not expressed in normal liver that are subsequently

TABLE 50.3 Immunogenetics of liver disease

Disease	Dominant haplotype
Autoimmune liver disease	
AIH	DRB1*0301, DRB1*0401, DRB1*1501, DRB1*0405, DRB1*1301
PSC	MICA*008, DRB1*0301, DBQ1*0603, DBQ1*0602, DBQ1*0302, DBQ1*0303, MICA*002
PBC	DRB1*0803, DRB1*0801
Viral hepatitis	
HAV	DRB1*1301
HBV	DRB1*1301, DRB1*1302
HCV	DRB1*0101, DRB1*0301

AIH: auto-immune hepatitis; PSC: primary sclerosing cholangitis; PBC: primary biliary cirrhosis; HAV: hepatitis A virus; HBV: hepatitis B virus; HCV: hepatitis C virus.

[1]Parsing string used ("liver" or "hepatic") and "human" for UniGene Build 180.

expressed with the development of liver pathobiology. This is a particularly important consideration in genomic medicine, as transcriptomes in disease can markedly increase in complexity, especially in the presence of neoplastic transformation and inflammation (Feezor et al., 2005; Scriver, 2004).

THE LIVER PROTEOME

The human proteome is complex and variable depending on the organ or cell population being studied. Proteins can be subject to in excess of 100 different types of post-translational modifications (Cantin and Yates, 2004). Therefore, the estimated 25,000 genes in the human genome may give rise to greater than a million distinct proteins (Neverova et al., 2005; Righetti et al., 2005). Utilizing an approach of parsing GenBank with key words, we can gain insight into human organ proteome complexity[2]. Proteomes from human brain, liver, lung, kidney, bowel/colon, heart and serum express 40%, 13%, 10%, 18%, 6%, 7% and 6% respectively, of the proteins found in all of these organs. The number of GenBank human protein entries for each organ was: brain (53,091), liver (17,610), lung (12,770), kidney (23,283), bowel/colon (7749), heart (8961) and serum (8122). This highlights the complexity of the proteome within the liver and other solid organs.

Contained within a cell, approximately 90% of the cellular protein mass is due to the 100 most abundant proteins and a further 1200 proteins account for another 7% of the protein mass (Lefkovits et al., 2000, 2001). However, the remaining 3% of the protein mass includes 2800 proteins (over 50% of the different protein species) and frequently is below the detection limit of most proteomic detection methods (Lefkovits et al., 2000, 2001). Therefore, it is important to consider the frequency of protein expression within a homogeneous cellular sample (i.e., cell lines) compared to heterogeneous cellular sample (i.e., organs). This is an especially important consideration in genomic medicine, as the non-parenchymal cell subpopulation abundance in biopsy specimens is low and sample representation of the whole organ is often inaccurate, a problem known to occur in upward of 15% of liver biopsy samples (Ratziu et al., 2005; Regev et al., 2002).

Genomic medicine approaches often assume that changes in gene/mRNA expression reflect changes in corresponding protein expression. However, there are multiple examples where protein expression or function is not controlled by mRNA expression. Indeed, in the intact non-diseased liver tissue, approximately 25% of the changes in the mRNA transcript expression are not accompanied by changes in protein expression (Anderson and Seilhamer, 1997). Studies comparing mRNA and protein expression are rare in all organs, including the liver. Anderson et al showed a poor correlation of the liver tissue abundance of 19 proteins and corresponding mRNA transcripts (correlation coefficient of only 0.48) (Anderson and Seilhamer, 1997). Additionally, they isolated 50 abundant mRNA transcripts, of which 29 encoded secreted proteins (Anderson and Seilhamer, 1997). However, this result contrasted with the 50 most abundant proteins they isolated, as none were secreted (Anderson and Seilhamer, 1997). Overall, within mRNA transcriptomes compared to the corresponding proteomes, there is a bias in expression toward an over-representation of mRNA transcripts encoding secreted proteins, and high abundance mRNA transcripts such as G3PDH have been repeatedly demonstrated (Anderson and Seilhamer, 1997; Jansen and Gerstein, 2000; Miklos and Maleszka, 2001a, b; ter Kuile and Westerhoff, 2001). However, it is salient to remember that the protein expression in every cell is controlled by the transcriptome, although the relationship between individual gene transcripts and the corresponding protein expression may not at first be apparent.

[2]In November 2006, the GenBank protein entries were parsed with the key word *human* and one of the following: *brain, liver, lung, kidney, bowel or colon, heart, kidney and serum*. A total of 131,586 protein entries across all of these organs were identified. The percentage for each organ was then calculated to give an estimate of relative protein abundance.

DEVELOPMENT OF LIVER FIBROSIS

Liver fibrosis is characterized by activation of the HSC. Invariably, there is associated inflammation and an intrahepatic immune response driving the perpetuation of the pro-fibrogenic phenotype of the HSC. Therefore, fibrosis development can be examined by focusing on the pro-fibrogenic HSC or alternatively by studying the various types of liver injury, associated inflammatory response and the pathways of HSC activation.

Hepatic Stellate Cell

The HSC is the principal cell type mediating the matrix remodeling and degradation of fibrosis (Figure 50.2). The HSC normally constitutes approximately 5% of the intrahepatic cell mass in the non-diseased liver and increases to 10–15% in the diseased liver (Friedman, 2000; Mehal et al., 2001). The HSC is likely to have evolved from the neural crest origin in contrast to the endoderm origin of hepatic parenchymal cells, as suggested by the presence of markers such as glial fibrillary acidic protein, nestin and N-CAM (Levy et al., 1999; Nakatani et al., 1996; Niki et al., 1999; Vogel et al., 2000). The use of gene array analysis has led to the identification of a number of additional neural markers in cirrhosis including BDNF, GDNF and neuromodoluin, which are now thought to be previously unrecognized markers of HSC (Shackel et al., 2002). However, heterogeneity in the intrahepatic HSC population suggests that the intrahepatic stellate cells may not derive from a single embryonic source (Levy et al., 1999; Nakatani et al., 1996; Sell, 2001). The normal stellate cell exists in a quiescent state, and with injury it is transformed into a proliferative, fibrogenic and contractile myofibroblast,

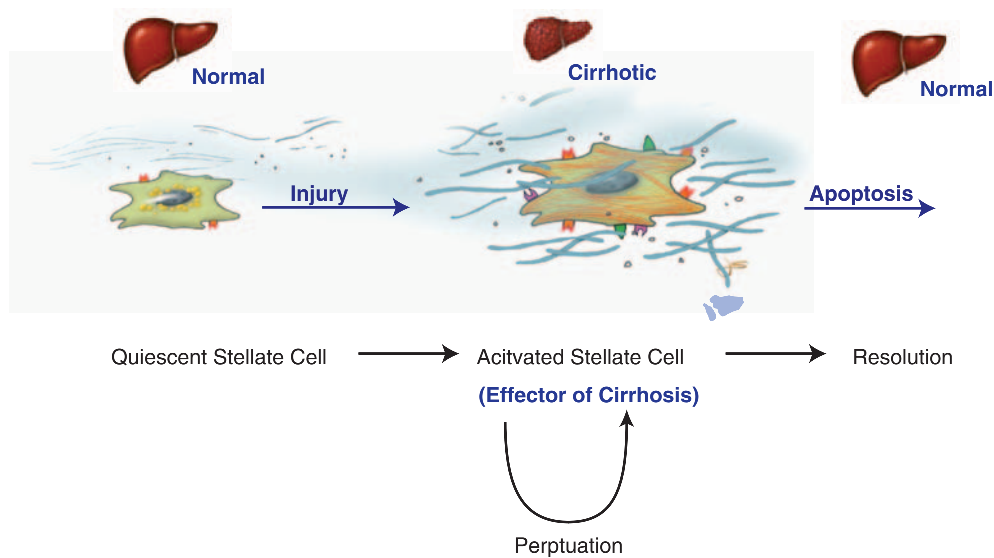


Figure 50.2 HSC activation. Intrahepatic injury is accompanied by activation of the HSC. This transformation sees the quiescent stellate cell, which stores retinoid, transformed into a profibrogenic cell that actively remodels the ECM leading to cirrhosis. Activation of the stellate cell leads to state in which the activated cell phenotype is perpetuated by both the initial injurious insult and also by autocrine cell activation. Resolution of fibrosis is due to removal of the activated stellate cell through the process of apoptosis.

a response known as stellate cell activation (Friedman, 2000). The activation of the HSC involves both an initiation and perpetuation of the activated cell phenotype. Additionally, the activated phenotype can resolve in the liver with cessation of injury. It is now clear that distinct molecular pathways are involved in the process of initiation, perpetuation and resolution of the activated HSC (Friedman, 2000).

Overall, fibrosis is characterized by a three- to six-fold increase in collagen and non-collagenous components of the ECM (Gressner, 1998). Activated stellate cells produce hyaluronan, fibronectin, entactin, tenescin, undulin, elastin and laminin (Gressner, 1998). The quiescent stellate cell produces predominantly type IV collagen, whereas the activated myofibroblast produces predominantly collagen type I as well as collagen type IV and III (Friedman, 2000; Gressner, 1998). Additionally, myofibroblasts produce proteoglycan, with chondroitin sulfate being the predominant proteoglycan, as well as LTBP sulfate and heparin sulfate (Gressner, 1998).

The phenotypic changes associated with HSC activation include matrix remodeling, proliferation and contractility. Upon activation the morphology of the HSC changes with a loss of intracellular vitamin A stores, development of more pronounced cytoplasmic processes and flattening of the cell (Friedman, 2000; Gressner, 1998). This is accompanied by increased HSC contractility, especially due to the production of NO and endothelin-1, which is an important determinant of the increased portal venous pressure during liver injury (Rockey, 1997). Proliferation of the activated HSC occurs in response to a number of growth factors, most of which signal through receptor tyrosine kinases (Ankoma Sey et al., 1998; Friedman, 2000). Growth factors such as PDGF, TGF-β and EGF are initially secreted from adjacent Kupffer cells, hepatocytes and cells of the inflammatory infiltrate (Callahan et al., 1985; Marra et al., 1994; Pinzani et al., 1996). Subsequently, an autocrine loop is established by the activated HSC that produces these growth factors, especially PDGF and TGF-β (Friedman, 2000; Callahan et al., 1985; Marra et al., 1994).

The dynamic nature of the fibrotic response is apparent given the documented resolution seen in many animal models of fibrosis and the improvement seen with treatment of some causes of human liver disease. The resolution of the activated HSC is unlikely to involve the "retrodifferentiation" or transition back to the quiescent phenotype but appears to involve activated HSC elimination by apoptosis (Benyon and Iredale, 2000). In animal models of fibrosis, there is a significant increase in the rate of activated HSC apoptosis. Activated HSC are more sensitive to Fas-ligand-induced apoptosis compared to quiescent HSC. Intrahepatic natural killer (NK) cells in response to liver injury upregulate Fas-ligand (FasL) expression (Benyon and Iredale, 2000; Moroda et al., 1997; Tsutsui et al., 1996). Persistence of the activated HSC despite increased Fas-ligand and Fas-ligand sensitivity seems to be due to the concurrent increase in proliferative factors, especially PDGF, that are anti-apoptotic (Benyon and Iredale, 2000; Simakajornboon et al., 2001; Staiger and Loffler, 1998). Similarly other autocrine factors may act to enhance activated HSC apoptosis including expression of IL-10 and members of the Bcl-2 family (Galle, 1997; Kanzler and Galle, 2000; Ockner, 2001). This pro- and anti-apoptotic response seen with HSC activation may be a means of autocrine regulation limiting "scar" formation.

Genomic Studies of HSCs

Genomic studies of HSC have led to the development of *ex vivo* cell culture systems and a better understanding of *in vivo* injury attributable to HSC. Microarray analysis has been utilized to identify differential gene expression in activated HSCs from *in vitro* culture models (Lee et al., 2004; Liu et al., 2004). In one study, a number of novel and previously recognized genes were identified, including MAPK pathway genes, osteopontin and ERK-1 (Lee et al., 2004; Qiang et al., 2006). Further, ERK-1 was subsequently shown in RNA interference experiments to be necessary for HSC proliferation as well as being anti-apoptotic with expression maintaining the activated cell phenotype (Qiang et al., 2006). Differentially expressed genes in murine-activated HSC included those involved in protein synthesis (RP16), cell-cycle regulation (Cdc7), apoptosis (Nip3) and DNA damage response (MAT1) (Liu et al., 2004). Further, genomic studies addressing the perpetuation of the activated HSC phenotype identified expression of the telomerase catalytic subunit (human telomerase reverse transcriptase [hTERT]) in human activated HSCs, which immortalizes these cells and maintains an activated HSC phenotype (Schnabl et al., 2002). Senescent HSCs expressed reduced levels of ECM proteins, including collagens, tenascin and fibronectin. Furthermore, maintenance of telomere length represents an important survival factor for activated human HSCs (Schnabl et al., 2003). Therefore, an immortalized human HSC line has been generated by infecting primary human HSCs with a retrovirus over-expressing hTERT (Schnabl et al., 2002). The hTERT positive HSCs do not undergo oncogenic transformation and exhibit morphologic and functional characteristics of *in vivo* activated HSCs. Subsequent Genechip™ and RT-PCR analysis showed that mRNA expression patterns in telomerase-positive HSCs are similar to those in primary *in vivo* activated human HSCs (Schnabl et al., 2002). Similarly, another two immortalized HSC lines LX-1 and LX-2 were characterized by microarray analysis and determined to have a gene expression profile similar to that of *in vivo* activated human primary HSC (Xu et al., 2005). These newly developed cell lines are proving to be valuable tools to study the biology of human HSCs. Importantly, although microarray and Genechip™ studies have demonstrated that these *ex vivo* HSC are similar to *in vivo* activated HSC with a greater than 70% similarity in transcript expression, a number of differences have been identified. Therefore, *ex vivo* cell culture of stable HSC may not be completely representative of *in vivo* injury highlighting the importance of intracellular interactions within the liver in determining transcriptome expression.

TRANSCRIPTOME ANALYSIS OF LIVER DISEASE

Presently there are hundreds of published gene array studies of human liver disease or studies that utilize human liver tissue. Most of these studies attempt to understand liver by examining mRNA transcript expression. There are few publications in human liver disease where gene expression is correlated with clinical outcome. Two of the most common causes of liver injury globally, viral hepatitis B and C, are discussed in more detail in Chapter 112. Unfortunately, many forms of liver injury cannot be studied in commonly used laboratory animal models and most animal studies focus on the pathways of fibrosis development.

Hepatitis B Virus Infection

Functional genomics studies of acute hepatitis B virus (HBV) infection in the chimpanzees has led to unique insights into how this virus evades the immune response and causes injury (Wieland et al., 2004). Initially, there is no apparent significant immune/inflammation-associated differential gene expression during the early phase of HBV infection and viral replication. Therefore, HBV infection acts in the initial phase as a "stealth virus" failing to induce a significant innate immune response (Wieland and Chisari, 2005; Wieland et al., 2004). Intrahepatic gene induction is first seen during the phase of attempted viral clearance. Gene expression during the early phase of infection was associated with T cell receptor and antigen presentation. Following this T cell effector function (granzymes), T cell recruitment (chemokines) and monocyte activation-associated gene expression has been demonstrated. A later phase of clearance was associated with the expression of B cell-related genes. Chronic HBV infection is characterized by gene expression profiles consistent with active inflammatory response (increased IRF-6 and CCL16), cell proliferation (increased cyclin-H and p53) and cellular apoptosis (14-3-3 interacting gene increased) (Honda et al., 2001, 2006).

Hepatitis C Virus Infection

Acute and chronic hepatitis C virus (HCV) infection has been studied using functional genomics techniques. These experiments have examined acute HCV infection in primates, as well as the sequelae of chronic infection. The results from microarray studies of acute HCV infection in the chimpanzee are intriguing. Acute HCV infection is characterized by a rapid (within 2 weeks) as well as a delayed induction (up to 6 weeks) of genes involved associated with the innate immune response (Bigger et al., 2001). Most of these genes are associated with interferon gene expression and are known as interferon response genes (IRG's; including ISG15, ISG16, CXCL9, CXCL10, Mx-1, stat-1, 2′5′-oligoadenylate synthetase and p27). Viral clearance appears to be associated with rapid induction of these IRG's. Overall HCV persistence appears to be associated with comparatively less induction of IRG's compared to viral clearance. Further, chronic HCV-related liver injury appears to be characterized by an IRG-associated chronic Th1 immune response, which is insufficient to clear the virus but is chronic and responsible for ongoing liver injury. The situation with interferon treatment of HCV infected individuals, which is aimed at viral eradication, is similar to acute infection as an immune response, characterized by a significant increase in IRG expression following treatment, is associated with a greater likelihood of a sustained long-term therapy response. Clearly, the immune response to HCV drives fibrogenesis,

as interferon administration is associated with a reduction intrahepatic inflammation and fibrosis even in the absence of a long-term virological clearance following treatment.

Alcoholic Liver Disease

Globally, alcohol is a leading cause of progressive liver fibrosis leading to cirrhosis. Intrahepatic gene profiling using microarrays in ethanol-fed baboons has identified increased expression of multiple annexin-related genes (including A1 and A2) that were not previously implicated in the progression of fibrosis in alcoholic liver disease (ALD; Seth et al., 2003). Further, the intrahepatic transcriptome profile in alcohol-associated liver injury is significantly different from other forms of liver disease (Deaciuc et al., 2004; Seth et al., 2003). Transcriptome profiling has allowed differentiation of alcoholic hepatitis from alcoholic steatosis. Genes known to be involved in alcohol injury such as alcohol dehydrogenases, acetaldehyde dehydrogenases, interleukin-8, *S*-adenosyl methionine synthetase, phosphatidylethanolamine *N*-transferase and several solute carriers have been shown to be differentially expressed in alcoholic hepatitis versus alcoholic steatosis. In alcoholic hepatitis, many novel differentially expressed genes have been identified, including claudins, osteopontin, CD209, selenoprotein, annexin A2 and genes related to bile duct proliferation (Seth et al., 2003). Differentially expressed genes involved cell adhesion, ECM proteins, oxidative stress and coagulation that were common to alcoholic hepatitis and end-stage ALD. Importantly, genes associated with fibrosis, cell adhesion, ECM remodeling are increased in human advanced ALD, consistent with the fibrotic nature of ALD. However, many of these genes are not specific to alcohol-induced liver injury and have been reported in other forms of liver cirrhosis such as primary biliary cirrhosis (PBC) (Shackel et al., 2001, 2002).

Non-Alcoholic Fatty Liver Disease

Non-alcoholic fatty liver disease (NAFLD) is the "western" liver disease related to obesity and form part of the metabolic syndrome characterized by increased BMI, hypertension and insulin resistance. Non-alcoholic steatohepatitis (NASH) is the clinicopathological syndrome in NAFLD in which lipid deposition within the liver is accompanied by inflammation and is widely studied using gene array analysis of mRNA transcript expression. The inflammation in NAFLD that results in NASH is of considerable importance as NASH, not NAFLD, is characterized by progressive injury and eventual cirrhosis. Studies have identified differentially expressed genes in end-stage NASH cirrhosis compared to other disease states (Sreekumar et al., 2003; Younossi et al., 2005a, b). Decreased expression of genes associated with mitochondrial function and increased expression of genes associated with the acute phase response are observed (Sreekumar et al., 2003). The latter increases were speculated to be associated with insulin resistance, a feature of NAFLD (Sreekumar et al., 2003). Further differential expression of genes involved in lipid metabolism, ECM remodeling, regeneration, apoptosis and detoxicification have all been observed in NASH following microarray analysis (Younossi, et al., 2005a, b).

Autoimmune Hepatitis

Autoimmune hepatitis (AIH) is an uncommon autoimmune disease affecting the hepatic lobule with a number of morphological similarities to the injury caused by HBV and HCV. The only available data on AIH are a comparison between HCV- and AIH-associated cirrhosis (Shackel et al., 2002). One of the key findings in this study was the observation that three inhibitors of apoptosis (IAP) genes were selectively differentially expressed in AIH. This is an intriguing finding. If this gene expression was identified to be in the intrahepatic lymphocyte population, then lack of apoptosis of such cells may be an important pathogenic pathway in the perpetuation of AIH. In this comparison to HCV, AIH was associated with an inflammatory gene pathway that consisted of a mix of Th1- and Th2-associated genes. Therefore, an intrahepatic Th1 immune response appears to be more fibrogenic, as evident from this work, HCV liver injury and animal models of liver injury (Shi et al., 1997).

Biliary Liver Injury

Viral hepatitis and alcoholic liver injury typically affect the hepatic lobule within the liver and are referred to classically "hepatitic" or "lobular" liver injury. However, a number of other types of liver injury are characterized by insults primarily to the bile ducts and are commonly referred to a "biliary" disease. Classically, biliary diseases include autoimmune-mediated bile duct injury of PBC and the poorly understood condition of bile ducts known as primary sclerosing cholangitis (PSC) in which inflammation, infection and fibrosis initially affects the bile ducts and then the liver parenchyma.

One of the major findings in a gene array examination of PBC end-stage liver disease was the identification of a subset of genes associated with the Wnt pathway (Shackel et al., 2001). In particular, Wnt13, Wnt5A and Wnt12 were shown to be differentially expressed. Other genes particularly upregulated in PBC included transcription initiation factor 250 kDa subunit (TAFII 250), PAX3/forkhead transcription factor and patched homolog (PTC). A consistent feature of the gene array analysis of PBC was the repeated identification of *Drosophila* genes homologs that were differentially expressed (Wnt genes, hedgehog pathway, notch pathway) (Shackel et al., 2001).

The only available data on PSC come in a comparison with PBC (Shackel et al., 2001). Compared with PBC, there were a far greater number of genes showing differential expression in PSC versus non-diseased liver (compared with PBC, and non-diseased liver). These include genes associated with epithelial biology (amphiregulin, bullous pemphigoid antigen), inflammation (T cell Secreted Protein P I-309, CTLA4), apoptosis-related genes (Bcl-2 interacting killer, Bcl-x, Death associated protein 3) and intracellular kinases such as CDK7 and JAK1.

A disease in which bile ducts are absent, biliary atresia (BA), has been studied comparing gene expression in embryonic versus perinatal forms of the disease (Zhang et al., 2004). Gene profiling clearly separated these two conditions. The most remarkable difference was in the expression of so-called regulatory genes.

In embryonic BA, 45% of differentially expressed genes were in this category versus 15% in the perinatal form. Included in these genes were imprinting genes, genes associated with RNA processing and cell cycle regulation that were not present in the perinatal form of BA. Overall, the results from genomic studies of biliary liver disease are consistent with distinct patterns of injury each associated with unique gene signatures.

The first clinically adopted genomic medicine approach in liver disease arose from the development of a gene array resquencing chip capable of diagnosis of a number of rare inherited inborn errors of metabolism resulting in neonatal hyperbilirubinemia (Liu et al., 2007; Watchko et al., 2002).

Hepatocellular Carcinoma

Liver injury characterized by cirrhosis is a premalignant condition. Despite the many diverse causes of liver injury, the development of HCC invariably occurs in the setting of cirrhosis. Indeed, the development of HCC without cirrhosis is extremely rare. Functional genomics methodologies are advancing the understanding of how fibrosis progression to cirrhosis is required for malignancy development. However, this sequence of events is still poorly understood.

Neoplastic proliferation in HCV-associated HCCs has been studied by microarray analysis. A plethora of potential novel tumor markers have been identified. These include the serine/threonine kinase 15 (STK15) and phospholipase A2 (PLA2G13 and PLA2G7) that were found to be increased in over half of the tumors identified (Smith et al., 2003). However, different studies identify different gene groups in HCV-associated HCC such as: cytoplasmic dynein light chain, hepatoma derived growth factor, ribosomal protein L6, TR3 orphan receptor and c-myc (Shirota et al., 2001). The clustering analysis in this study showed that the expression of 22 genes in HCC related to differentiation of the malignancy, with over half of these genes being transcription factors or related to cell development or differentiation (Shirota et al., 2001). Although many of these genes can be implicated in HCC development, they are often identified in gene sets obtained from end-stage diseased tissue. Therefore, whether these genes represent cause or effect is unknown. HBV-associated HCC has been studied by several groups (Iizuka et al., 2002, 2004; Kim et al., 2004a, b). Genes associated with cell proliferation, cell cycle, apoptosis and angiogenesis were dysregulated in HCC tissues. Increased expression of cyclin-dependent kinases was seen whilst several cell cycle negative regulators were decreased. Metastatic development has also been studied using gene arrays (Pan et al., 2003; Qin and Tang, 2004; Tang et al., 2004; Ye et al., 2003). Genes identified with metastatic development include osteopontin, and *in vivo* neutralizing antibodies to osteopontin block tumor invasion (Ye et al., 2003). A study of unsupervised gene profiling of patients with HCC revealed a set of genes associated with decreased survival including RhoC (Wang et al., 2004). These genes include a subset of pro-proliferative, anti-apoptotic genes as well as genes involved in ubiquitination and histone modification. Gene profiles in HCC have also demonstrated patterns of gene expression associated with tumor differentiation, vascular invasion as well as recurrence after surgery (Tang et al., 2004).

PROTEOMIC STUDIES OF LIVER DISEASE

Proteomic studies of liver disease have fallen into the following four groups: (1) discovery of previously unrecognized proteins in a cell population or disease state (Eng and Friedman, 2000; Kawada et al., 2001; Kristensen et al., 2000), (2) biomarker discovery (Schwegler et al., 2005; Seow et al., 2001), (3) hepatic toxicological prediction/profiling (Fella et al., 2005; Fountoulakis and Suter, 2002; Gao et al., 2004; Guzey and Spigset, 2002; Kaplowitz, 2004; Low et al., 2004; Meneses-Lorente et al., 2004; Merrick and Bruno, 2004; Nordvarg et al., 2004; Roelofsen et al., 2004) and (4) studies of known proteins or classes proteins (Greenbaum et al., 2002; Joyce et al., 2004) (see Tables 50.1 and 50.2). Proteomics has successfully been used in biodiscovery of proteins in hepatocytes, HSCs (Eng and Friedman, 2000; Kawada et al., 2001; Kristensen et al., 2000), HCC (Seow et al., 2001) and viral hepatitis (Comunale et al., 2004; Garry and Dash, 2003; Kim et al., 2003; Lu et al., 2004; Rosenberg, 2001; Scholle et al., 2004). In these studies tens to hundreds of proteins were identified. However, these studies are limited as they sample rather than profile the proteome. Further observed changes in protein expression may reflect weak associations rather than a direct role in the development of pathobiology. The biodiscovery approach has been successfully used in toxicological models and used to develop characteristic profiles of protein expression that may predict intrahepatic toxicology responses (Fella et al.,2005; Fountoulakis and Suter, 2002; Gao et al., 2004; Guzey and Spigset, 2002; Kaplowitz, 2004; Low et al., 2004; Merrick and Bruno, 2004; Meneses-Lorente et al., 2004; Nordvarg et al., 2004; Roelofsen et al., 2004).

This is an area of intense research focus for pharmaceutical companies as they strive to reduce development cost and aim to predict drug toxicity earlier in the drug development cycle. Biomarker discovery is another area receiving attention using proteomic methods. However, for years there has not been a newly FDA-approved serum marker as this is an immense research challenge. Biological sample protein concentrations vary by 12–15 orders of magnitude, and specific serum markers are likely to be expressed at nanomolar or lower concentrations. One approach to overcome these limitations is to use a combination of potential markers that are easier to detect but with each protein marker alone having a lower specificity but high sensitivity. This is an approach currently used in serum tests of hepatic fibrosis, and proteomic methods are being used to try and identify new serum markers of hepatic fibrogenesis (Henkel et al., 2005; Poon et al., 2005; Xu et al., 2004). One of the most promising approaches is the use of accurate mass tags (AMT) or suicide substrates that selectively and reproducibly target a sub-proteome (Bogdanov

and Smith, 2005; Greenbaum et al., 2002; Joyce et al., 2004; Pasa-Tolic et al., 2004). This has the added advantage of aiding prefractionation and increasing resolution of proteins as the tag can be captured on an affinity surface (Greenbaum et al., 2002; Joyce et al., 2004).

Hepatitis B Virus Infection

In contrast to gene array experiments, there are a number of studies that use proteomics on sera to examine different stages of chronic HBV infection. In one study, altered proteomic profiles were identified for haptoglobin beta and alpha 2 chain, apolipoprotein A-1 and A-1V, alpha-1 antitrypsin, transthyretin and DNA topisomerase 11 beta (He et al., 2003). Some of these proteins are amongst the most abundant serum proteins secreted by the liver and are generally associated with acute phase inflammatory responses. What was apparent in this study was that different isoforms of some of these proteins showed distinct changes in HBV infection itself and differed at times between patients with low inflammatory scores versus high inflammatory scores. Some examples include a decrease in cleaved haptoglobin beta peptides and ApoA-1 fragments in patients with higher inflammatory scores. In comparison, some alpha-1-antitrypsin fragments were increased in patients with higher inflammatory scores. An alternate approach studied serum protein profiles and correlated this with disease severity using a SELDI Protein Chip analysis and artificial neural network (ANN) models (Poon et al., 2005). They found 6 fragments with a positive and 24 with a negative prediction of fibrosis stage and subsequently developed a fibrosis index with excellent precise values for significant fibrosis and cirrhosis based on the Ishak fibrosis score. The inclusion of clinical biochemical parameters such as ALT, bilirubin, total protein, hemoglobin and INR strengthened the accuracy of their predictive model (see Chapter 112).

Hepatitis C Virus Infection

Proteomic methodologies have been applied to a number of aspects of HCV-related liver injury. These include the study of HCV-related HCC development in which overexpression of alpha enolase was identified and correlated with poorly differentiated HCC (Kuramitsu and Nakamura, 2005; Takashima et al., 2005). The response of hepatocyte cell lines to IFN gamma treatment has uncovered over 54 IFN response genes including many novel targets, an approach that may pave the way for novel therapies. Examination of protein extracts that bind to the HCV IRES has identified a number of novel protein targets such as Ewing Sarcoma breakpoint 1 region protein EWS and TRAF-3. The final aspect of HCV liver injury receiving attention is the study of potential biomarkers such as heat shock protein HSP-70 associated with HCV infection progression to HCC (Takashima et al., 2003) (see Chapter 112).

AIH, PBC and PSC Associated Liver Disease

There are few published proteomic studies addressing the pathophysiology of these diseases. Cholangiocarcinoma that is associated with PSC has been studied using proteomics techniques (Koopmann et al., 2004). Using tandem mass spectroscopy, Koopman and colleagues identified Mac-2-binding protein (Mac-2BP) as a diagnostic marker in biliary carcinoma. The diagnostic accuracy of serum Mac-2BP expression in biliary carcinoma was superior to the established marker CA19-9. This study highlights the progression of proteomic research in liver disease; a focus initially on malignancy and biomarker discovery is followed by studies of pathophysiology.

ALD and NAFLD

Proteomic studies of alcoholism have, like the gene array studies, been an eclectic mix of research examining HCC development associated with alcohol, studies of hepatocyte alcohol-related biology and neural aspects of alcohol addiction. Studies of the intrahepatic toxic effects using proteomics have helped outline toxicology profiles that can be used for screening as well as trying to understand the alcohol-associated liver injury. Mitochondrial ethanol hepatoxicity is thought to involve modification of protein thiol redox state. Using 2D gel proteomic studies, Venkatraman et al. (2004) were able to demonstrate a decrease in the reduced thiols on aldehyde dehydrogenase and glucose regulated protein 78 (Venkatraman et al., 2004). The change in aldehyde dehydrogenase-reduced thiols was accompanied by a reduction in specific activity of the enzyme. The term "alcoholomics" has been coined to refer to the study of those proteins (i.e., the sub-proteome) that are directly or indirectly affected by alcohol.

There has been to date only a single study evaluating proteomic profiling in NAFLD (Younossi et al., 2005a, b). This study used SELDI-TOF MS to profile serum samples from 91 patients with NAFLD and 7 obese controls. Twelve unique protein peaks were identified that associated with NALFD (4 associated with steatosis, 4 with steatosis with non-specific inflammation and 4 with NASH). Unfortunately, although the peak mass was shown, SELDI-TOF MS lacks the accuracy required to give mass determination enabling equivocal protein identification.

Hepatocellular Carcinoma

To date, biomarker profiling has predominantly focused on studies of malignant tissue (Liu et al., 2005; Wiesner, 2004; Wong et al., 2004). In one study, HCC development in chronic HBV infection was characterized by a significant decrease in a fragment of complement-3 and an isoform of apolipoprotein A-1 (Steel et al., 2003). In tissue studies, expression of variants of aldehyde dehydrogenase and tissue ferritin light chain has been identified in HCC, but not surrounding tissues. Similar studies have also identified fructose-bisphosphatase arginosuccinate synthetase and cathepsin B pre-protein as downregulated in HCC tissues. In an extensive study using laser capture microscopy, Li et al. (2004) identified 261 proteins differentially expressed between HCC and non-HCC hepatocytes. Kinases in the Eph family were identified along the Ras-like family of Rho proteins. In addition a DEAD box polypeptide was downregulated whilst three members of the spliceosome and heterogeneous

nuclear ribonucleoprotein K were upregulated. Also SELDI-TOF MS has examined the sera of 82 patients with cirrhosis (38 without and 44 with HCC). An algorithm including the six highest scoring peaks allow the prediction of HCC in over 90% of cases (Paradis et al., 2005). The highest discriminating peak was a C terminal peptide of vitronectin.

Proteomic Analysis of Blood As a Marker of Liver Disease: "Next Generation" Liver Function Tests?

Several studies have evaluated proteomic analysis of serum protein as a diagnostic test to assess the severity of liver disease and in particular for non-invasive assessment of liver fibrosis. These studies are in their infancy, with basic methodologies still unresolved. However, the early studies show the technique has promise. In a pilot study of 46 patients with chronic hepatitis B, an ANN model was derived from the proteomic fingerprint and used to derive a fibrosis index (Poon et al., 2005). The ANN fibrosis index strongly correlated with Ishak scores and stages of fibrosis. The area under the ROC curve for significant fibrosis (Ishak score >2) and cirrhosis (Ishak score >4) were both >0.90. Inclusion of International Normalized Ratio (INR), total protein, bilirubin, alanine aminotransferase and hemoglobin in the ANN model improved the predictive power, giving accuracies >90% for the prediction of fibrosis and cirrhosis. Another study found that pre-treatment of serum proteins to remove *N*-glycosylation enhanced the resolution of serum polypeptide profiles (Comunale et al., 2004). This technique has the potential to improve diagnostic serum proteomics.

Chen et al. (2004) developed a method of glycoproteomic analysis in an attempt to discover serum markers that can assist in the early detection of HBV-induced liver cancer. The authors showed that woodchucks diagnosed with HCC have dramatically higher levels of serum-associated core alpha-1,6-linked fucose. One glycoprotein, Golgi Protein 73 (GP73), was found to be elevated and hyper-fucosylated in the serum of animals and humans with a diagnosis of HCC. Serum profiling was used to distinguish HCC from earlier stages of HCV-related liver disease (Schwegler et al., 2005). The proteomic model distinguished chronic HCV from HCV to HCC with moderate sensitivity and specificity. Inclusion of known serum markers alpha feto-protein, des-gamma carboxyprothrombin, along with GP73, significantly improved the diagnostic accuracy of detecting HCC.

PROTEOMICS IN OTHER LIVER DISEASE

The metalloproteome is defined as the set of proteins that have metal-binding capacity by being metalloproteins or having metal-binding sites. The copper and zinc metalloproteomes were defined in human hepatoma lines (She et al., 2003). Although the gene for Wilson disease has been identified, the mechanisms by which excess copper leads to oxidative stress, acute liver failure or cirrhosis are not fully understood. Using an *in vitro* model of copper loading, novel copper-binding proteins were isolated using proteomic techniques (Roelofsen et al., 2004).

Although there has been limited proteomic analysis of liver tissue in models of disease, liver-associated pathobiological processes have been examined. One study of the liver aging process identified 85 differentially expressed proteins comprised of antioxidation glucose/amino acid metabolism signal transduction and cell cycle systems (Cho et al., 2003). In aging, the antioxidation system showed a large increase in glutathione peroxidase and a decrease in glutathione-*S*-transferase. Similarly, levels of t-glycludic enzymes were decreased in the aging animal. Furthermore, levels of proteins associated in signal transduction/apoptosis, for example, cathepsin B, were decreased in the aging process. However, it is unclear if the identified genes explain the increased rate of fibrosis progression seen with increasing age.

Proteomics has been used to also identify genes associated with LPS-induced liver injury. Proteins such as TRAIL receptor 2 were downregulated in the liver of LPS treated mice, whilst TNFAIP1 was significantly upregulated. Four different proteins were novel in the fatty liver proteome (aconitase succinate dehydrogenase, propanol Co-A carboxylase alpha chain and 3-hydroxyanthrilate 3-4-dioxygenase). LPS is thought to be an important mediator of injury in a number of conditions such as alcohol-related liver injury.

2009 UPDATE

Recent advances in high-throughput genotyping technology have led to an increasing number of genome-wide association studies to assess disease progression and determine risks for development of HCC in chronic liver disease. The importance of selection of adequate samples and controls for functional genomic studies was highlighted by a study that noted differential gene expression patterns in inflammatory response pathways among 28 histologically normal liver tissue samples obtained by either percutaenous or surgical techniques (Asselah et al., 2008). Emerging data indicate that changes in hepatic miRNA expression may occur during fibrosis progression and development of cirrhosis, and miRNAs that represent targets of cell division, mitosis, and G_1-S transition may determine outcomes in HCC (Jiang et al., 2008). Liver tumors appear to have distinct miRNA expression profiles dependent on malignancy, risk factors, and oncogene/tumor suppressor gene alterations (Ladeiro et al., 2008). Emerging tools for large-scale epigenetic analysis have also been applied in relation to development of HCC. A study using methylated CpG island amplification microarray techniques identified 719 prominent genes of hypermethylation

in HCC tissue, with distinct patterns of gene methylation noted in progression from adjacent normal liver tissue to HCC (Gao et al., 2008). Another study using real-time PCR for quantitative DNA methylation in 77 matched HCC and noncancerous liver tissues noted higher levels of methylation in HCC corresponding noncancerous liver tissues compared to 22 normal liver tissue samples. Methylation causes disruption of a variety of genes and pathways during hepatocarcinogenesis, such as RB-related (*p16*, *RIZ1*), p53-related (*HIC-1*, *Reprimo*), WNT/APC (*APC*, *SFRP2*, *CDH1*), receptor-tyrosine kinase-associated (*RASSF1A*, *RASSF2*, *SOCS-1*), transforming growth factor beta signaling (*RUNX3*), and apoptosis-related pathways (*CASP8*). This study indicates that chronic HCV infection may act as a powerful epimutagen based on findings of significant association with the progression of age-related methylation compared to HBV infection (Nishida et al., 2008). At this stage, there are no epigenomic data available in relation to fibrosis or responses to antiviral therapy.

A DNA-mediated annealing, selection, extension, and ligation (DASL) 6000-gene expression profiling method applied to paraffin-embedded tissue from patients with HCC has been shown to yield high quality data, even in tissue stored for up to 24 years. A 186-gene expression signature from adjacent noncancerous tissue was noted to correlate with increased survival in a validation cohort. In this study, the poor-prognosis gene signature included inflammatory pathway genes related to IFN signaling and activation of nuclear factor-κB and tumor necrosis factor-α (Hoshida et al., 2008).

Proteomic profiling using 2D-gel electrophoresis or SELDI-TOF has been evaluated in relation to developing biomarkers of fibrosis in HCV infection, and demonstrate differential protein expression for advanced stage fibrosis (White et al., 2007), with only minimal incremental diagnostic benefit compared to validated serum marker panels of fibrosis (Morra et al., 2007). Quantitative proteomic profiling using iTRAQ in liver tissues from HCC patients indicated differential expression of newly identified proteins in this disease state, including fibroleukin, interferon induced 56 kDa protein, milk fat globule-EGF factor 8, and myeloid-associated differentiation marker (Chaerkady et al., 2008). Serum protein peaks identified using SELDI-TOF MS may also accurately detect HCC compared to established tumor markers such as α-fetoprotein (AFP), AFP-L3, and prothrombin induced by vitamin K absence-II (PIVKA-II) (Zinkin et al., 2008).

There is minimal, reproducible data using "omics" platforms in nonviral mediated chronic liver disease such as nonalcoholic fatty liver disease. Advances in high-throughput functional genomic and quantitative proteomic techniques will likely result in enhanced biomarker and target discovery efforts in other chronic liver disease.

2009 UPDATE REFERENCES

Asselah, T., Bieche, I., Laurendeau, I., Martinot-Peignoux, M., Paradis, V., Vidaud, D., Valla, D.C., Bedossa, P., Marcellin, P. and Vidaud, M. (2008). Significant gene expression differences in histologically "Normal" liver biopsies: Implications for control tissue. *Hepatology* 48, 953–62.

Chaerkady, R., Harsha, H.C., Nalli, A., Gucek, M., Vivekanandan, P., Akhtar, J., Cole, R.N., Simmers, J., Schulick, R.D., Singh, S. et al. (2008). A quantitative proteomic approach for identification of potential biomarkers in hepatocellular carcinoma. *J Proteome Res* 7, 4289–4298.

Gao, W., Kondo, Y., Shen, L., Shimizu, Y., Sano, T., Yamao, K., Natsume, A., Goto, Y., Ito, M., Murakami, H. et al. (2008). Variable DNA methylation patterns associated with progression of disease in hepatocellular carcinomas. *Carcinogenesis* 29, 1901–1910.

Hoshida, Y., Villanueva, A., Kobayashi, M., Peix, J., Chiang, D.Y., Camargo, A., Gupta, S., Moore, J., Wrobel, M.J., Lerner, J. et al. (2008). Gene expression in fixed tissues and outcome in hepatocellular carcinoma. *N Engl J Med* 359, 1995–2004.

Jiang, J., Gusev, Y., Aderca, I., Mettler, T.A., Nagorney, D.M., Brackett, D.J., Roberts, L.R. and Schmittgen, T.D. (2008). Association of microRNA expression in hepatocellular carcinomas with hepatitis infection, cirrhosis, and patient survival. *Clin Cancer Res* 14, 419–27.

Ladeiro, Y., Couchy, G., Balabaud, C., Bioulac-Sage, P., Pelletier, L., Rebouissou, S. and Zucman-Rossi, J. (2008). MicroRNA profiling in hepatocellular tumors is associated with clinical features and oncogene/tumor suppressor gene mutations. *Hepatology* 47, 1955–1963.

Morra, R., Munteanu, M., Bedossa, P., Dargere, D., Janneau, J.L., Paradis, V., Ratziu, V., Charlotte, F., Thibault, V., Imbert-Bismut, F. et al. (2007). Diagnostic value of serum protein profiling by SELDI-TOF ProteinChip compared with a biochemical marker, FibroTest, for the diagnosis of advanced fibrosis in patients with chronic hepatitis C. *Aliment Pharmacol Ther* 26, 847–858.

Nishida, N., Nagasaka, T., Nishimura, T., Ikai, I., Boland, C.R. and Goel, A. (2008). Aberrant methylation of multiple tumor suppressor genes in aging liver, chronic hepatitis, and hepatocellular carcinoma. *Hepatology* 47, 908–918.

White, I.R., Patel, K., Symonds, W.T., Dev, A., Griffin, P., Tsokanas, N., Skehel, M., Liu, C., Zekry, A., Cutler, P. et al. (2007). Serum proteomic analysis focused on fibrosis in patients with hepatitis C virus infection. *J Transl Med* 5, 33.

Zinkin, N.T., Grall, F., Bhaskar, K., Otu, H.H., Spentzos, D., Kalmowitz, B., Wells, M., Guerrero, M., Asara, J.M., Libermann, T.A. et al. (2008). Serum proteomics and biomarkers in hepatocellular carcinoma and chronic liver disease. *Clin Cancer Res* 14, 470–477.

REFERENCES

Anderson, L. and Seilhamer, J. (1997). A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 18(3–4), 533–537.

Ankoma Sey, V. et al. (1998). Coordinated induction of VEGF receptors in mesenchymal cell types during rat hepatic wound healing. *Oncogene* 17(1), 115–121.

Bataller, R. and Brenner, D.A. (2005). Liver fibrosis. *J Clin Invest* 115(2), 209–218.

Bataller, R., North, K.E. and Brenner, D.A. (2003). Genetic polymorphisms and the progression of liver fibrosis: A critical appraisal. *Hepatology* 37(3), 493–503.

Benyon, R.C. and Iredale, J.P. (2000). Is liver fibrosis reversible? *Gut* 46(4), 443–446.

Bigger, C.B., Brasky, K.M. and Lanford, R.E. (2001). DNA microarray analysis of chimpanzee liver during acute resolving hepatitis C virus infection. *J Virol* 75(15), 7059–7066.

Bogdanov, B. and Smith, R.D. (2005). Proteomics by FTICR mass spectrometry: Top down and bottom up. *Mass Spectrom Rev* 24(2), 168–200.

Callahan, M., Cochran, B.H. and Stiles, C.D. (1985). The PDGF-inducible "competence genes": Intracellular mediators of the mitogenic response. *Ciba Foundation Symposium* 116, 87–97.

Cantin, G.T. and Yates, J.R. (2004). 3rd, Strategies for shotgun identification of post-translational modifications by mass spectrometry. *J Chromatogr A* 1053(1–2), 7–14.

Cao, S.X. et al. (2001). Genomic profiling of short- and long-term caloric restriction effects in the liver of aging mice. *Proc Natl Acad Sci USA* 98(19), 10630–10635.

Chen, J. et al. (2004). Proteome analysis of gastric cancer metastasis by two-dimensional gel electrophoresis and matrix assisted laser desorption/ionization-mass spectrometry for identification of metastasis-related proteins. *J Proteome Res* 3(5), 1009–1016.

Cho, Y.M. et al. (2003). Differential expression of the liver proteome in senescence accelerated mice. *Proteomics* 3(10), 1883–1894.

Comunale, M.A. et al. (2004). Comparative proteomic analysis of de-N-glycosylated serum from hepatitis B carriers reveals polypeptides that correlate with disease status. *Proteomics* 4(3), 826–838.

Corbett, C.E., Duarte, M.I. and Bustamante, S.E. (1993). Regression of diffuse intralobular liver fibrosis associated with visceral leishmaniasis. *Am J Trop Med Hyg* 49(5), 616–624.

Coulouarn, C. et al. (2004). Altered gene expression in acute systemic inflammation detected by complete coverage of the human liver transcriptome. *Hepatology* 39(2), 353–364.

Deaciuc, I.V. et al. (2004). Large-scale gene profiling of the liver in a mouse model of chronic, intragastric ethanol infusion. *J Hepatol* 40(2), 219–227.

Donaldson, P.T. (2004). Genetics of liver disease: Immunogenetics and disease pathogenesis. *Gut* 53(4), 599–608.

Dufour, J.F., DeLellis, R. and Kaplan, M.M. (1998). Regression of hepatic fibrosis in hepatitis C with long-term interferon treatment. *Dig Dis Sci* 43(12), 2573–2576.

Enard, W. et al. (2002). Intra- and interspecific variation in primate gene expression patterns. *Science* 296(5566), 340–343.

Eng, F.J. and Friedman, S.L. (2000). Fibrogenesis I. New insights into hepatic stellate cell activation: The simple becomes complex. *Am J Physiol Gastrointest Liver Physiol* 279(1), G7–G11.

Feezor, R.J. et al. (2005). Functional genomics and gene expression profiling in sepsis: Beyond class prediction. *Clin Infect Dis* 41(Suppl 7), S427–S435.

Fella, K. et al. (2005). Use of two-dimensional gel electrophoresis in predictive toxicology: Identification of potential early protein biomarkers in chemically induced hepatocarcinogenesis. *Proteomics* 5(7), 1914–1927.

Fountoulakis, M. and Suter, L. (2002). Proteomic analysis of the rat liver. *J Chromatogr B Analyt Technol Biomed Life Sci* 782(1–2), 197–218.

Friedman, S.L. (2000). Molecular regulation of hepatic fibrosis, an integrated cellular response to tissue injury. *J Biol Chem* 275(4), 2247–2250.

Galle, P.R. (1997). Apoptosis in liver disease. *J Hepatol* 27(2), 405–412.

Gao, J. et al. (2004). Identification of *in vitro* protein biomarkers of idiosyncratic liver toxicity. *Toxicol Vitro* 18(4), 533–541.

Garry, R.F. and Dash, S. (2003). Proteomics computational analyses suggest that hepatitis C virus E1 and pestivirus E2 envelope glycoproteins are truncated class II fusion proteins. *Virology* 307(2), 255–265.

Greenbaum, D. et al. (2002). Chemical approaches for functionally probing the proteome. *Mol Cell Proteomics* 1(1), 60–68.

Gressner, A.M. (1998). The cell biology of liver fibrogenesis – an imbalance of proliferation, growth arrest and apoptosis of myofibroblasts. *Cell Tissue Res* 292(3), 447–452.

Guzey, C. and Spigset, O. (2002). Genotyping of drug targets: A method to predict adverse drug reactions? *Drug Saf* 25(8), 553–560.

He, Q.Y. et al. (2003). Serum biomarkers of hepatitis B virus infected liver inflammation: A proteomic study. *Proteomics* 3(5), 666–674.

Henkel, C. et al. (2005). Identification of fibrosis-relevant proteins using DIGE (difference in gel electrophoresis) in different models of hepatic fibrosis. *Z Gastroenterol* 43(1), 23–29.

Honda, M. et al. (2001). Differential gene expression between chronic hepatitis B and C hepatic lesion. *Gastroenterology* 120(4), 955–966.

Honda, M. et al. (2006). Different signaling pathways in the livers of patients with chronic hepatitis B or chronic hepatitis C. *Hepatology* 44(5), 1122–1138.

Huang, H. et al. (2006). Identification of two gene variants associated with risk of advanced fibrosis in patients with chronic hepatitis C. *Gastroenterology* 130(6), 1679–1687.

Iizuka, N. et al. (2002). Comparison of gene expression profiles between hepatitis B virus- and hepatitis C virus-infected hepatocellular carcinoma by oligonucleotide microarray data on the basis of a supervised learning method. *Cancer Res* 62(14), 3939–3944.

Iizuka, N. et al. (2004). Molecular signature in three types of hepatocellular carcinoma with different viral origin by oligonucleotide microarray. *Int J Oncol* 24(3), 565–574.

Jansen, R. and Gerstein, M. (2000). Analysis of the yeast transcriptome with structural and functional categories: Characterizing highly expressed proteins. *Nucleic Acids Res* 28(6), 1481–1488.

Joyce, J.A. et al. (2004). Cathepsin cysteine proteases are effectors of invasive growth and angiogenesis during multistage tumorigenesis. *Cancer Cell* 5(5), 443–453.

Kanzler, S. and Galle, P.R. (2000). Apoptosis and the liver. *Semin Canc Biol* 10(3), 173–184.

Kaplowitz, N. (2004). Drug-induced liver injury. *Clin Infect Dis* 38(Suppl 2), S44–S48.

Kawada, N. et al. (2001). Characterization of a stellate cell activation-associated protein (STAP) with peroxidase activity found in rat hepatic stellate cells. *J Biol Chem* 276(27), 25318–25323.

Kim, W. et al. (2003). Comparison of proteome between hepatitis B virus- and hepatitis C virus-associated hepatocellular carcinoma. *Clin Cancer Res* 9(15), 5493–5500.

Kim, B.Y. et al. (2004). Feature genes of hepatitis B virus-positive hepatocellular carcinoma, established by its molecular discrimination approach using prediction analysis of microarray. *Biochim Biophys Acta* 1739(1), 50–61.

Kim, J.W. et al. (2004). Cancer-associated molecular signature in the tissue samples of patients with cirrhosis. *Hepatology* 39(2), 518–527.

Koopmann, J. et al. (2004). Mac-2-binding protein is a diagnostic marker for biliary tract carcinoma. *Cancer* 101(7), 1609–1615.

Kristensen, D.B. et al. (2000). Proteome analysis of rat hepatic stellate cells. *Hepatology* 32(2), 268–277.

Kuramitsu, Y. and Nakamura, K. (2005). Current progress in proteomic study of hepatitis C virus-related human hepatocellular carcinoma. *Expert Rev Proteomics* 2(4), 589–601.

Lee, S.H. et al. (2004). Effects and regulation of osteopontin in rat hepatic stellate cells. *Biochem Pharmacol* 68(12), 2367–2378.

Lefkovits, I., Kettman, J.R. and Frey, J.R. (2000). Global analysis of gene expression in cells of the immune system I. Analytical limitations in obtaining sequence information on polypeptides in two-dimensional gel spots. *Electrophoresis* 21(13), 2688–2693.

Lefkovits, I., Kettman, J.R. and Frey, J.R. (2001). Proteomic analysis of rare molecular species of translated polypeptides from a mouse fetal thymus cDNA library. *Proteomics* 1(4), 560–573.

Levy, M.T. et al. (1999). Fibroblast activation protein: A cell surface dipeptidyl peptidase and gelatinase expressed by stellate cells at the tissue remodelling interface in human cirrhosis. *Hepatology* 29(6), 1768–1778.

Li, C. et al. (2004). Accurate qualitative and quantitative proteomic analysis of clinical hepatocellular carcinoma using laser capture microdissection coupled with isotope-coded affinity tag and two-dimensional liquid chromatography mass spectrometry. *Mol Cell Proteomics* 3(4), 399–409.

Liu, C. et al. (2007). Novel resequencing chip customized to diagnose mutations in patients with inherited syndromes of intrahepatic cholestasis. *Gastroenterology* 132(1), 119–126.

Liu, A.Y. et al. (2005). Analysis of prostate cancer by proteomics using tissue specimens. *J Urol* 173(1), 73–78.

Liu, X.J. et al. (2004). Association of differentially expressed genes with activation of mouse hepatic stellate cells by high-density cDNA microarray. *World J Gastroenterol* 10(11), 1600–1607.

Low, T.Y. et al. (2004). A proteomic analysis of thioacetamide-induced hepatotoxicity and cirrhosis in rat livers. *Proteomics* 4(12), 3960–3974.

Lu, H. et al. (2004). Riboproteomics of the hepatitis C virus internal ribosomal entry site. *J Proteome Res* 3(5), 949–957.

Marra, F. et al. (1994). Regulation of platelet-derived growth factor secretion and gene expression in human liver fat-storing cells. *Gastroenterology* 107(4), 1110–1117.

Mehal, W.Z., Azzaroli, F. and Crispe, I.N. (2001). Immunology of the healthy liver: Old questions and new insights. *Gastroenterology* 120(1), 250–260.

Meneses-Lorente, G. et al. (2004). A proteomic investigation of drug-induced steatosis in rat liver. *Chem Res Toxicol* 17(5), 605–612.

Merrick, B.A. and Bruno, M.E. (2004). Genomic and proteomic profiling for biomarkers and signature profiles of toxicity. *Curr Opin Mol Ther* 6(6), 600–607.

Miklos, G.L. and Maleszka, R. (2001). Integrating molecular medicine with functional proteomics: Realities and expectations. *Proteomics* 1(1), 30–41.

Miklos, G.L. and Maleszka, R. (2001). Protein functions and biological contexts. *Proteomics* 1(2), 169–178.

Moroda, T. et al. (1997). Autologous killing by a population of intermediate T-cell receptor cells and its NK1.1+ and NK1.1- subsets, using Fas ligand/Fas molecules. *Immunology* 91(2), 219–226.

Nakatani, K. et al. (1996). Expression of neural cell adhesion molecule (N-CAM) in perisinusoidal stellate cells of the human liver. *Cell Tissue Res* 283(1), 159–165.

Neverova, I. and Van Eyk, J.E. (2005). Role of chromatographic techniques in proteomic analysis. *J Chromatogr B Analyt Technol Biomed Life Sci* 815(1–2), 51–63.

Niki, T. et al. (1999). Class VI intermediate filament protein nestin is induced during activation of rat hepatic stellate cells. *Hepatology* 29(2), 520–527.

Nordvarg, H. et al. (2004). A proteomics approach to the study of absorption, distribution, metabolism, excretion, and toxicity. *J Biomol Tech* 15(4), 265–275.

Ockner, R.K. (2001). Apoptosis and liver diseases: Recent concepts of mechanism and significance. *J Gastroenterol Hepatol* 16(3), 248–260.

Pan, H.W. et al. (2003). Overexpression of osteopontin is associated with intrahepatic metastasis, early recurrence, and poorer prognosis of surgically resected hepatocellular carcinoma. *Cancer* 98(1), 119–127.

Paradis, V. et al. (2005). Identification of a new marker of hepatocellular carcinoma by serum protein profiling of patients with chronic liver diseases. *Hepatology* 41(1), 40–47.

Pasa-Tolic, L. et al. (2004). Proteomic analyses using an accurate mass and time tag strategy. *Biotechniques* 37(4), 621–624. 626–33, 636 passim

Pinzani, M. et al. (1996). Expression of platelet-derived growth factor and its receptors in normal human liver and during active hepatic fibrogenesis. *Am J Pathol* 148(3), 785–800.

Poon, T.C. et al. (2005). Prediction of liver fibrosis and cirrhosis in chronic hepatitis B infection by serum proteomic fingerprinting: A pilot study. *Clin Chem* 51(2), 328–335.

Poynard, T. et al. (2005). Age and gender will survive to competing risks as fibrosis factors. *Gastroenterology* 128(2), 519–520. author reply 520-521

Prosser, C.C., Yen, R.D. and Wu, J. (2006). Molecular therapy for hepatic injury and fibrosis: Where are we? *World J Gastroenterol* 12(4), 509–515.

Qiang, H. et al. (2006). Differential expression genes analyzed by cDNA array in the regulation of rat hepatic fibrogenesis. *Liver Int* 26(9), 1126–1137.

Qin, L.X. and Tang, Z.Y. (2004). Recent progress in predictive biomarkers for metastatic recurrence of human hepatocellular carcinoma: A review of the literature. *J Cancer Res Clin Oncol* 130(9), 497–513.

Ratziu, V. et al. (2005). Sampling variability of liver biopsy in nonalcoholic fatty liver disease. *Gastroenterology* 128(7), 1898–1906.

Regev, A. et al. (2002). Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. *Am J Gastroenterol* 97(10), 2614–2618.

Righetti, P.G. et al. (2005). Prefractionation techniques in proteome analysis: The mining tools of the third millennium. *Electrophoresis* 26(2), 297–319.

Rockey, D. (1997). The cellular pathogenesis of portal hypertension: Stellate cell contractility, endothelin, and nitric oxide. *Hepatology* 25(1), 2–5.

Roelofsen, H., Balgobind, R. and Vonk, R.J. (2004). Proteomic analyzes of copper metabolism in an *in vitro* model of Wilson disease using surface enhanced laser desorption/ionization-time of flight-mass spectrometry. *J Cell Biochem* 93(4), 732–740.

Rosenberg, S. (2001). Recent advances in the molecular biology of hepatitis C virus. *J Mol Biol* 313(3), 451–464.

Schnabl, B. et al. (2002). Immortal activated human hepatic stellate cells generated by ectopic telomerase expression. *Lab Invest* 82(3), 323–333.

Schnabl, B. et al. (2003). Replicative senescence of activated human hepatic stellate cells is accompanied by a pronounced inflammatory but less fibrogenic phenotype. *Hepatology* 37(3), 653–664.

Scholle, F. et al. (2004). Virus-host cell interactions during hepatitis C virus RNA replication: Impact of polyprotein expression on the cellular transcriptome and cell cycle association with viral RNA synthesis. *J Virol* 78(3), 1513–1524.

Schwegler, E.E. et al. (2005). SELDI-TOF MS profiling of serum for detection of the progression of chronic hepatitis C to hepatocellular carcinoma. *Hepatology* 41(3), 634–642.

Scriver, C.R. (2004). After the genome – the phenome? *J Inherit Metab Dis* 27(3), 305–317.

Sell, S. (2001). Heterogeneity and plasticity of hepatocyte lineage cells. *Hepatology* 33(3), 738–750.

Seow, T.K. et al. (2001). Hepatocellular carcinoma: From bedside to proteomics. *Proteomics* 1(10), 1249–1263.

Seth, D. et al. (2003). Gene expression profiling of alcoholic liver disease in the baboon (*Papio hamadryas*) and human liver. *Am J Pathol* 163(6), 2303–2317.

Shackel, N.A. et al. (2001). Identification of novel molecules and pathogenic pathways in primary biliary cirrhosis: cDNA array analysis of intrahepatic differential gene expression. *Gut* 49(4), 565–576.

Shackel, N.A. et al. (2002). Insights into the pathobiology of hepatitis C virus-associated cirrhosis: Analysis of intrahepatic differential gene expression. *Am J Pathol* 160(2), 641–654.

She, Y.M. et al. (2003). Identification of metal-binding proteins in human hepatoma lines by immobilized metal affinity chromatography and mass spectrometry. *Mol Cell Proteomics* 2(12), 1306–1318.

Shi, Z., Wakil, A.E. and Rockey, D.C. (1997). Strain-specific differences in mouse hepatic wound healing are mediated by divergent T helper cytokine responses. *Proc Natl Acad Sci USA* 94(20), 10663–10668.

Shirota, Y. et al. (2001). Identification of differentially expressed genes in hepatocellular carcinoma with cDNA microarrays. *Hepatology* 33(4), 832–840.

Siegel, C.A. et al. (2005). Liver biopsy 2005: When and how? *Cleve Clin J Med* 72(3), 199–201, 206, 208 passim.

Simakajornboon, N. et al. (2001). *In vivo* PDGF beta receptor activation in the dorsocaudal brainstem of the rat prevents hypoxia-induced apoptosis via activation of Akt and BAD. *Brain Research* 895(1–2), 111–118.

Smith, M.W. et al. (2003). Hepatitis C virus and liver disease: Global transcriptional profiling and identification of potential markers. *Hepatology* 38(6), 1458–1467.

Sreekumar, R. et al. (2003). Hepatic gene expression in histologically progressive nonalcoholic steatohepatitis. *Hepatology* 38(1), 244–251.

Staiger, H. and Loffler, G. (1998). The role of PDGF-dependent suppression of apoptosis in differentiating 3T3-L1 preadipocytes. *Eur J Cell Biol* 77(3), 220–227.

Steel, L.F. et al. (2003). A strategy for the comparative analysis of serum proteomes for the discovery of biomarkers for hepatocellular carcinoma. *Proteomics* 3(5), 601–609.

Tadic, S.D. et al. (2002). Sex differences in hepatic gene expression in a rat model of ethanol-induced liver injury. *J Appl Physiol* 93(3), 1057–1068.

Takashima, M. et al. (2003). Proteomic profiling of heat shock protein 70 family members as biomarkers for hepatitis C virus-related hepatocellular carcinoma. *Proteomics* 3(12), 2487–2493.

Takashima, M. et al. (2005). Overexpression of alpha enolase in hepatitis C virus-related hepatocellular carcinoma: Association with tumor progression as determined by proteomic analysis. *Proteomics* 5(6), 1686–1692.

Tang, Z.Y. et al. (2004). A decade's studies on metastasis of hepatocellular carcinoma. *J Cancer Res Clin Oncol* 130(4), 187–196.

ter Kuile, B.H. and Westerhoff, H.V. (2001). Transcriptome meets metabolome: Hierarchical and metabolic regulation of the glycolytic pathway. *FEBS Lett* 500(3), 169–171.

Tsushima, H. et al. (1999). Reduced plasma transforming growth factor-beta1 levels in patients with chronic hepatitis C after interferon-alpha therapy: Association with regression of hepatic fibrosis. *J Hepatol* 30(1), 1–7.

Tsutsui, H. et al. (1996). IFN-gamma-inducing factor up-regulates Fas ligand-mediated cytotoxic activity of murine natural killer cell clones. *J Immunol* 157(9), 3967–3973.

Venkatraman, A. et al. (2004). Oxidative modification of hepatic mitochondria protein thiols: Effect of chronic alcohol consumption. *Am J Physiol Gastrointest Liver Physiol* 286(4), G521–G527.

Vogel, S. et al. (2000). An immortalized rat liver stellate cell line (HSC-T6): A new cell model for the study of retinoid metabolism *in vitro*. *J Lipid Res* 41(6), 882–893.

Wang, W. et al. (2004). Genomic analysis reveals RhoC as a potential marker in hepatocellular carcinoma with poor prognosis. *Br J Cancer* 90(12), 2349–2355.

Waring, J.F. et al. (2003). Development of a DNA microarray for toxicology based on hepatotoxin-regulated sequences. *EHP Toxicogenomics* 111(1T), 53–60.

Watchko, J.F., Daood, M.J. and Biniwale, M. (2002). Understanding neonatal hyperbilirubinaemia in the era of genomics. *Semin Neonatol* 7(2), 143–152.

Wieland, S. et al. (2004). Genomic analysis of the host response to hepatitis B virus infection. *Proc Natl Acad Sci USA* 101(17), 6669–6674.

Wieland, S.F. and Chisari, F.V. (2005). Stealth and cunning: Hepatitis B and hepatitis C viruses. *J Virol* 79(15), 9369–9380.

Wiesner, A. (2004). Detection of tumor markers with ProteinChip technology. *Curr Pharm Biotechnol* 5(1), 45–67.

Wong, Y.F. et al. (2004). Protein profiling of cervical cancer by protein-biochips: Proteomic scoring to discriminate cervical cancer from normal cervix. *Cancer Lett* 211(2), 227–234.

Xu, C. et al. (2004). Identification and characterization of 177 unreported genes associated with liver regeneration. *Genomics Proteomics Bioinformatics* 2(2), 109–118.

Xu, L. et al. (2005). Human hepatic stellate cell lines, LX-1 and LX-2: New tools for analysis of hepatic fibrosis. *Gut* 54(1), 142–151.

Yamashita, T. et al. (2000). Comprehensive gene expression profile of a normal human liver. *Biochem Biophy Res Comm* 269(1), 110–116.

Yamashita, T. et al. (2004). Genome-wide transcriptome mapping analysis identifies organ-specific gene expression patterns along human chromosomes. *Genomics* 84(5), 867–875.

Yamashita, Y. et al. (2004). cDNA microarray analysis in hepatocyte differentiation in Huh 7 cells. *Cell Transplant* 13(7–8), 793–799.

Yang, X. et al. (2006). Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Res* 16(8), 995–1004.

Yano, N. et al. (2001). Profiling the adult human liver transcriptome: Analysis by cDNA array hybridization. *J Hepatol* 35(2), 178–186.

Ye, Q.H. et al. (2003). Predicting hepatitis B virus-positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. *Nat Med* 9(4), 416–423.

Younossi, Z.M. et al. (2005). A genomic and proteomic study of the spectrum of nonalcoholic fatty liver disease. *Hepatology* 42(3), 665–674.

Younossi, Z.M. et al. (2005). Hepatic gene expression in patients with obesity-related non-alcoholic steatohepatitis. *Liver Int* 25(4), 760–771.

Yuan, J. et al. (2001). Genome analysis with gene-indexing databases. *Pharmacol Ther* 91(2), 115–132.

Zhang, D.Y. et al. (2004). Coordinate expression of regulatory genes differentiates embryonic and perinatal forms of biliary atresia. *Hepatology* 39(4), 954–962.

CHAPTER

Genomic Medicine and Obesity

J. Alfredo Martínez

INTRODUCTION

Genetics of obesity have only achieved quantitative standards in the last decades (Marti et al., 2008). However, the first attempts to link the genetic inheritance to body composition were reported in the early 1920s by Davenport, in "Body built and its heritance", which were pursued by Vague and coworkers in the 1940s when they hypothesized about the occurrence of different obesity subtypes associated to sexual dimorphisms. Also, the "thrifty gene" theory raised in the paper by Neel (1962), outlining the fact that genes that predispose to diabetes and obesity would have relative advantage in populations that often experienced starvations, contributed to establish a relationships between genetics and body weight control (Bell et al., 2005). However, the classic books published in the 1970s about *Energy Balance and Obesity in Man* by Garrow (1974) and *The Obese Patient* by Bray (1976) only made minor references concerning monogenic forms of obesity. In 1977, the Nutritional Heart, Lung and Blood Twins Study first indicated the possibility that the observed aggregation for obesity was due to genetic factors rather than environment (Feinleib et al., 1977). Indeed, only in the International Congress of Obesity held at Jerusalem in 1986, the pioneering works by Bouchard et al. concerning the role of genetic predisposition on obesity onset were debated, which was continued with the recognized study concerning weight gain in overfed twins (Bouchard and Tremblay, 1997). Later, other classic textbooks about obesity by Bjorntorp (2001), as well as by Bouchard, Bray, and James (1998 and 2004), reinforced the fundamentals to characterize the role of inheritance, contributions and gene–lifestyle interactions on obesity. More recently the impact of Biomics, particularly genomics and transcriptomics, on obesity research is producing new and challenging breakthroughs by using specific tools and methodologies such a microarrays (Ommen, 2004). Up to 2006, 12 editions of the human obesity map have been published showing the continuous advances in the field (Rankinen et al., 2006).

OBESITY: CAUSES AND GENETIC PREDISPOSITION

Obesity is a multifactorial disorder characterized by disproportionately high adipose tissue content in the body, accompanying a disequilibrium in the energy balance equation: energy intake $>$ energy expenditure (Bray, 2003). A fat content higher than 25% of the body weight in men and 33% in women are often arbitrarily considered as cut-off points to identify a subject as obese (Seidell and Flegal, 1998), which is commonly associated with excessive weight-for-height. However, very muscular individuals or pregnant women may have a greater weight than expected for height according to the standards, without showing an increased adiposity (Kasper et al., 2005). Another aspect to be ascertained is the regional distribution of the weight as fat, since the abdominal fat deposition or "android obesity" is an increased risk factor for disease as compared to "gynoid obesity", in which fat is more evenly and peripherally distributed around the body and upper regions of the legs (Lean, 1998).

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

A positive energy balance between the amounts of energy consumed over the energy spent in everyday life underlies weight gain. Indeed, the excessive fat accumulation in adipose tissue leading to obesity is the result of a chronic overconsumption of foods and drinks over the energy expenditure requirements, in which dietary and lifestyle habits, sociological factors, metabolic and neuroendocrine alterations as well as hereditary components are involved (Martínez, 2000). Actually, environmental and lifestyle influences promoting excessive caloric intake and sedentary patterns are known to induce a positive energy balance leading to weight gain (Marti et al., 2004). Indeed, the availability of energy-dense meals, and sedentary patterns facilitated by motorized transport and other common physically inactive pursuits (TV viewing, computer work, etc.) have markedly risen in the last decades (Martínez et al., 1999). On the other hand, cross-sectional data show a strong association between unhealthy eating habits and physical inactivity, which have contributed to explain the observed obesity pandemic (OMS Report, 2000). In this context, a high intake of non-starch polysaccharides/fiber has been considered as a protective factor against obesity, while a high consumption of fast food and sweetened sugar drinks or fruit juices is viewed as a obesity risk factor (Bes-Rastrollo et al., 2006). Furthermore, prospective studies provide additional evidence suggesting that an increase in physical activity and small changes in dietary behavior may help to prevent a disproportionate weight gain (Hill et al., 2003).

Body weight stability and the associated regulatory processes depend upon nutrient intake and physical activity patterns, but are also influenced by compensatory genetic-dependent metabolic and neuroendocrine mechanisms (Martinez and Frühbeck, 1996). Thus, despite the daily fluctuations in both components of the energy balance equation (energy intake versus energy expenditure), body weight and adiposity remain in a dynamic steady state for long periods of time (often within ± 1% variation over many years), which has been attributed to the participation of a powerful control system to fine tune the dietary macronutrient supply to body fuel oxidation demands (Jequier and Tappy, 1999). The control of the maintenance of body composition has been the subject of a number of theories or hypotheses such as the occurrence of a physiological set point for body weight, glucostatic or glycogen drives for feeding, metabolic/nutrient partitioning approaches, the participation of the nervous system, an adipostat mediated by signals from the adipose tissue, all of which might be under genetic control and explain individual variability (Martínez, 2000). In this context, it has been hypothesized that the stability of body weight and composition depends upon an axis with three inter-related and self-controlled components: (1) food intake, (2) nutrient turnover and thermogenesis and (3) body fat stores (Figure 51.1). All three elements underlie complex inter-related feedback mechanisms, which are affected by the individual's genetic background. Indeed, obesity is caused by perturbations of the balance between food intake and energy expenditure, which is regulated by a complex physiological system that requires the integration of several peripheral signals and controls coordination in the brain, (Solomon and Martínez, 2006) basically through hypothalamic nuclei and different orexigenic or anorexigenic neuropeptides (NPY, AGRP, orexins, leptin, POMC, etc).

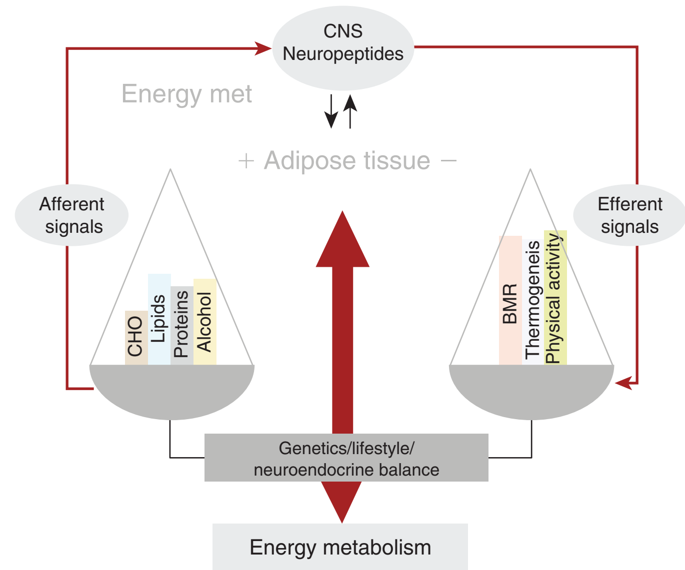


Figure 51.1 Body weight regulation.

Genes may determine afferent and efferent signals, as well as central mechanisms involved in body weight regulation (Bell et al., 2005). Thus, the transferable genetic information involved in short- and long-term stable body weight regulation and composition maintenance is acting via: (1) different peptides and monoamines involved in the regulation of the appetite, (2) variations in energy and nutrient utilization, resting metabolic rate or response to physical activity and (3) individual differences in adipocyte metabolism (Palou et al., 2003). The possible mechanisms through which the genetic susceptibility (Figure 51.2) could be acting include reduced rates of basal metabolism and macronutrients oxidation, alterations of adipogenesis and quantitative and qualitative deviations of food intake (Moreno et al., 2005).

Obesity, similar to other chronic diseases with a recognized monogenic or polygenic origin, is associated with a number of pathological dysfunctions and disturbances with important implications for the individual's and community health (hyperinsulinaemia, diabetes, hypertension, immunological alterations, certain cancer types, etc.), which is further worsened by the growing rates of obesity prevalence (20–40% of the EU population are overweight or obese and more than 50% of US subjects have problems of excessive weight gain) (OMS Report, 2000). Indeed, the health consequences of obesity are serious and diverse, ranging from premature death to different morbidities affecting quality of life (Kasper et al., 2005). Despite that relative risk estimations of the health problems associated with obesity are founded on a limited number of countries, some studies have reflected that the risk of suffering from NIDDM, gallbladder disease, dyslipemia, insulin resistance and sleep apnea are increased at least three times in the obese, while the risk of cardiovascular diseases and hypertension is moderately increased in the obese, and the risks of certain cancer and hormonal

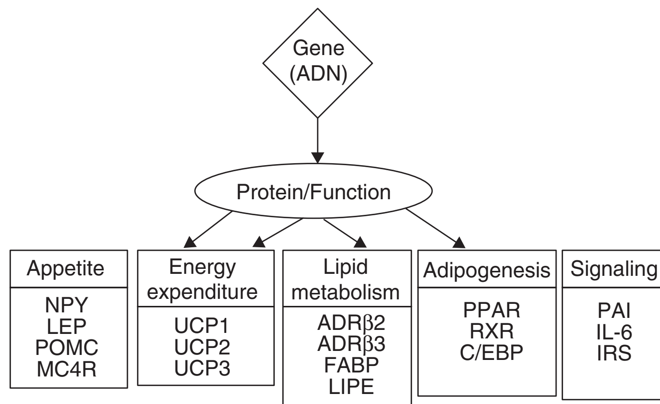


Figure 51.2 Some genes involved in body weight homeostasis categorized by processes.

disturbances as well as back pain associated with obesity show relative risks ranging between 1 and 2 (OMS Report, 2000). Also, severe obesity has been found to produce a 12-fold increase in mortality in 25–35 years old obese individuals when compared to lean subjects.

The rapid increase in obesity incidence over recent years suggests that environmental and lifestyle influences in addition to other psychopathological or genetic determinants are independently affecting the energy balance equation adjustment (Bray and Bouchard, 2004). Actually, the growing prevalence of obesity around the world is mainly attributed to changes in lifestyle (increase of the consumption of high-energy-yielding foods enriched with carbohydrates and fats or reductions of the physical activity, etc.) that specifically may impact genetic susceptibility. Also, from an evolutionary point of view, the individuals with thrifty genes are more resistant to malnutrition, explaining why large proportions of diverse populations are susceptible to become obese (Bell et al., 2005). The mutual interactions between the genetic make-up and the environment undoubtedly complicate the understanding of the specific roles of genes and external influences in obesity (Schuldiner and Munir, 2003).

SEARCH FOR GENES INVOLVED IN OBESITY

The interaction of genetic, environmental, physiological and psychosocial factors is involved in body weight homeostasis (Tremblay et al., 2004). Thus, the specific distribution of energy expenditure requirements and individual substrate partitioning are known to influence the energy balance equation depending on the genetic make-up (Moreno-Aliaga et al., 2005). In this context, the risk of excessive weight gain in children of some families with obese parents is increased two- to three-fold for moderate obesity and up to eight times for severe obesity (Marti et al., 2004). Moreover, twin and adoption studies substantiate a role for genetics in obesity (Bray and Bouchard, 2004). The discovery of populations such as Pima Indians with shared alterations in basal metabolism rates or in fat oxidation after food intake corroborates such hypothesis, as well as the fact that genetic factors can modulate the effects of physical activity and diet on weight and body composition (Bray et al., 1998). Furthermore, several studies carried out in large numbers of families encompassing members of different degrees of relatedness have allowed the quantification of the statistical association regarding objective indicators of obesity (body mass index (BMI) or percentage of fat) depending on the degree of the familial relationship (Bell et al., 2005; Marti et al., 2004). The heritability of BMI is lower between husband and wife (0.10–0.19) and between uncle/aunt and nephews (0.08–0.14), and increases between parents and children (0.15–0.23) and among siblings (0.24–0.34). The concordance for BMI is higher in monozygotic (0.70–0.88) than in dizygotic twins (0.15–0.42). Also, studies of dietary intervention, based on positive and negative energy balances in identical twins, convincingly pointed out that the differences in the susceptibility to overfeeding or periods of semi-starvation seem to be partially explained by genetic factors (Bouchard and Tremblay, 1997). Family studies have largely revealed that the heritability of the Quetelet index is about 25–50%, while twin studies have mostly estimated the contribution of genetic factors at 70–80% (Bell et al., 2005; Comuzzie and Allison, 1998). However, this information is not enough to prove unequivocally the genetic origin of obesity, since the families share, besides genes, other factors implied in obesity such as lifestyle, dietary habits and environmental factors. In this context, genetic association studies search for statistical relationships among a gene polymorphism with a given phenotype, generally among non-related individuals (Rosmond, 2003). This research strategy can consider the comparison between cases and controls, analysis of the variability for specific loci or the discrimination between mutation carriers and non-carriers regarding a given character. On the other hand, genetic linkage studies imply the persistence or co-segregation of a genetic marker or locus and a phenotypical character within a family (Bouchard, 1997). Furthermore, quantitative genetics allow assessing the influence of environmental factors on the hereditary variability, which can be based on diverse genes (polygenic interaction). The studies of gene expression as affected by dietary and activity patterns constitute another approach to relate the specific participation, both quantitative and qualitative, of different genes in energy homeostasis processes (Viguerie et al., 2005a). A salient feature of a study concerning the role of the lipid/carbohydrate content or the calorie restriction in affecting the mRNA levels is that the energy intake is more important than the macronutrient distribution (Viguerie et al., 2005b). Additionally, animal models have been able to characterize the impact of fat intake on gene expression (Lopez et al., 2005), which have evidenced that a number of genes controlling many different metabolic pathways such as lipogenesis, thermogenesis, lipolysis, adipogenesis, differentiation, oxidative stress, inflammation, etc. are affected by lipid consumption, which may help in the future to define molecular phenotypes or subsets of obesity.

TABLE 51.1 Approaches to study the genetics of obesity
• Studies within families, adoptions and twins
• Mendelian syndromes with obesity manifestations
Autosomal recessive
Autosomal dominant
X-linked
• Animal models
Spontaneous genetically obese animals
Transgenics and KO animals
Quantitative trait loci (QTL) studies
• Association and linkage studies
Candidate genes
Genome-wide scans

Obesity as a complex syndrome with a multiple etiology may be explained in some circumstances by monogenic mutations, but in most cases appears as a polygenic condition, which may be additionally affected by a myriad of environmental influences (Guy-Grand, 2003). A widely accepted hypothesis assumes that complex diseases such as obesity are likely to be based on a limited number of predisposing alleles, each conferring a small increase in the risk to the individual. Heterogeneity in complex phenotypes implies that the genetic predisposition may also result from any one of several rare variants in a number of genes (Marti et al., 2004).

The role of a genetic predisposition in obesity has long been assumed to affect both sides (intake/expenditure) of the energy equation (Figure 51.2). In this context, evidence from human single-gene mutations (leptin: LEP; leptin receptor: LEPR; pro-opiomelanocortin: POMC; melanocortin 4 receptor: MCR4; protein convertase 1: PC1; etc) has solely implicated energy intake (Rankinen et al., 2006). However, Mendelian syndromes with obesity as a clinical feature (e.g., Prader–Willi syndrome) have also revealed reductions in energy expenditure as a contributing factor to obesity (Lyon and Hirschhorn 2005).

In this context, evidence from single-gene mutation obesity cases, Mendelian disorders exhibiting obesity as a clinical feature, transgenic and knockout murine models relevant to obesity, quantitative trait loci (QTL) screening from animal cross-breeding experiments, association studies with candidate genes, and linkage studies from genome scans have been reported (Table 51.1). So far, at least, 176 human obesity cases due to single-gene mutations in 11 different genes have been reported, 50 loci related to Mendelian syndromes relevant to human obesity have been mapped to a genomic region, and causal genes or strong candidates have been identified for most of these syndromes. There are about 250 genes that, when mutated or expressed as transgenes in the mouse, result in phenotypes that affect body weight and adiposity. The number of QTLs reported from animal models currently reaches more than 450 markers. The description of human obesity QTLs derived from genome scans continues to grow, and at least 250 QTLs for obesity-related phenotypes from genome-wide scans have been reported. More than 50 genomic regions harbor QTLs supported by two or more investigations. The number of studies reporting associations between DNA sequence variation in specific genes and obesity phenotypes has also increased considerably, with more than 400 findings of positive associations with 127 candidate genes. A promising observation is that 22 genes are each supported by at least five positive studies (Tables 51.2 and 51.3). The obesity gene map shows putative loci on all chromosomes except Y. The latest electronic version of the map with links to useful publications and relevant sites can be found at http://obesitygene.pbrc.edu. Overall more than 600 genes, markers or chromosomal regions have been associated with human obesity phenotypes (Rankinen et al., 2006).

In this context, genetic epidemiology of obesity aims to screen and assess the phenotype traits associated to the genetic make-up and their relationships with lifestyle factors (Hill et al., 2003). The research strategies to identify the genetic determinants of obesity are multiple (Table 51.1). Indeed, the process of ascribing a gene to a phenotype is complex due to the low density of coded DNA (<5%), the potential interactions of genes with environmental factors and the diversity of methods and tools to assess body fat and energy metabolism phenotypes, which have hampered the advances in this area (Marti et al., 2004). The identification of genes may follow ascending strategies from the genotype to the phenotype, in which a gene or genes are investigated to find out a relation with a given obesity marker, but also descending strategies in which from a clinical feature of excessive fat accumulation, the genetic determinant is searched concerning the heritability of BMI, skinfolds and waist circumference (Bray et al., 1998). In Caucasian families, an estimate of 0.45–0.60 for measures of fat men such as BMI and 0.29–0.48 for measures of fat distribution has been calculated, which fit in well with the estimate from twin studies, where a minimum 0.40 for fatness or obesity was reported (Bell et al., 2005). The co-existence of obesity manifestations on various members of a family corroborates that the genetic make-up can play a role in obesity, which is reinforced by the fact that impairments in the thermogenesis, basal metabolic rate and sympathetic activity as well as specific effects on hyperphagia or physical activity performances are genetically controlled (Palou et al., 2003). However, these investigations do not discard a role for the shared environment conditioning dietary habits and physical activity patterns. Another aspect is that genetic predisposition may influence the android/gynecoid fat distribution (central or gluteo–femoral phenotypes), as has been proven in familial obesity risks studies (Katzmarzyk et al., 1999), which may be mediated in some cases by sexual gene-dependent differentiation of the adipose tissue muscle ratio (Vague et al., 1989).

DIAGNOSIS AND CHARACTERIZATION OF GENES ASSOCIATED WITH OBESITY

Genetic diseases are usually categorized as chromosomal alteration, monogenic or mendelian and multifactorial syndromes

TABLE 51.2 Selected list of genes that have been associated with obesity phenotypes

Gene	Name/Function	Associated phenotype
ACE	Angiotensin 1-converting enzyme	BMI, waist circumference, body fat
ADIPOQ	Adiponectin, C1Q and collagen containing domain	BMI, waist circumference
ADRB2	Adrenergic, beta 2 receptor	BMI, waist-to-hip ratio, body fat, skinfolds, lipolysis
ADBR3	Adrenergic, beta 3 receptor	BMI, abdominal visceral fat, waist-to-hip ratio
DRD2	Dopamine receptor D2	Body weight, reduced energy expenditure, skinfolds
GNB3	Guanine nucleotide-binding protein 3(G-protein)	BMI, reduced lypolysis, body fat
HTR2C	5-Hydroxtryptamine (serotonin) receptor 2C	BMI, weight change
IL-6	Interleukin 6	BMI (in men) waist circumference (in men) reduced, fasting energy expenditure
INS	Insulin	BMI, body weight, leptin
LDLR	Low-density lipoprotein receptor	BMI
LEP	Leptin	BMI, body weight
LEPR	Leptin receptor	BMI, fat-free mass, fat-free mass, body fat (%), leptin, abdominal total fat
LIPE	Lipase	BMI, body fat, skinfolds, waist-to-hip ratio, reduced lipolysis
MC4R	Melanocortin 4 receptor	BMI, body fat (%), resting energy expenditure
NR3CI	Nuclear receptor 3C 1 (glucocorticoid receptor)	BMI, waist-to-hip ratio (in men), leptin, skinfolds
PPARG	Peroxisome proliferative-activated receptor gamma	BMI, leptin, waist circumference, body fat (%), leptin, lipid oxidation
RETN	Resistin	BMI, waist circumference
TNFA	Tumor necrosis factor alfa	BMI, body fat (%), waist circumference
UCP1	Uncoupling protein 1	BMI, waist-to-hip ratio, body fat (%)
UCP2	Uncoupling protein 2	BMI, energy expenditure, macronutrient oxidation, body fat, skinfolds
UCP3	Uncoupling protein 3	BMI, waist-to-hip ratio, body fat (%), resting energy expenditure, leptin, skinfolds
VDR	Vitamin D (1,2,5-dihydroxyvitamin D3) receptor	Fat mass

(Kasper et al., 2005). The chromosome alterations involve absence, duplication or abnormal distribution of one of several chromosomes related to excessive or faulty material, which may affect body fat content as occurs in Down Syndrome (trisomy 21) or Turner syndrome (X monosomy), which are often accompanied by obesity and other abnormal anatomical manifestations.

Obesity-related Mendelian disorders can be classified as autosomal dominant, autosomal recessive or X-linked (Table 51.4). Thus, a search (July 2006) in the obesity heading of the OMIM database produced 75 entries concerning autosomal dominant transmission, 74 entries concerning autosomal recessive transmission and 40 X-linked. Most of these syndromes have been already assigned to a specific locus. The clinical features of some of these metabolic disorders include, in addition to obesity, mental retardation and some craniofacial/ anatomical deformations as well as insulin resistance, hypertension, etc. Interestingly, 11 single-gene mutations with a specific obesity phenotype have been identified so far (CRH1, CRH2, GPR24, LEP, LEPR, MC3R, MC4R, NRTK2, POMC, PCSK1 and SIM1). The majority of the monogenic obesity cases reported in the literature remain those with a genetic defect in the MC4R gene (Hebebrand et al., 2003).

The fundamental requirements for the candidate gene approach are the identification of a gene that is involved in

TABLE 51.3 Chromosomal location of selected candidate genes with more than three positive association studies and two positive linkage studies

Chromosome	Association studies	Linkage studies
1	*AGT, ATP1A2, LEPR, LMNA, NROB2*	*LEPR, ATP1A2, LEPR*
2	*IRS1, POMC*	*ACP1*
3	*ADPQ, GHRL, PPARG*	
4	*FABP2, PPAGC1A, UPC1*	
5	*ADRAB2, NR3C1*	
6	*ENPP1, ESR1, TNF-α*	*SE30*
7	*IL-6, LEP, NPY, SERPINE1*	*LEP*
8	*ADRAB3, LPL*	
9		
10	*ADRA2A; GAD2*	
11	*AD0A4; DRD2, IGF2, INS, UCP2,UCP3*	*ATA34 E8*
12	*GNB3, VDR*	*IGF-1*
13		
14		
15	*CYP19A1; PLIN*	
16	*AGRP, FOXC2*	
17	*ACE*	
18	*MC4R*	
19	*APOE, LDLR, LIPE, RETN*	
20		
21		
22	*PPARA*	
X	*AR, HTR2C, SLC6A14*	
Y		

the disease phenotype, a polymorphic maker within that gene and a suitable set of subjects to genotype for that marker (Bray et al., 1998). Association studies seek to establish statistical relations between a genetic polymorphism and a given phenotype. The statistical approach may search the comparison between cases and controls, the analysis of variations between cases and controls, the analysis of variations between specific loci and discrimination study between carriers and non-carriers for a trait (Rosmond, 2003). The association studies provide information in order to identify genes with low involvement, but they require establishing high statistical cut-offs (usually $p < 0.001$) to avoid random or biased findings. Evidence for association between markers of candidate genes with obesity-related phenotypes (BMI, skinfolds, waist circumference, LEP, etc.) has been reported for at least 125 genes (Perusse et al., 2005). Thus, statistical relationships with $p < 0.0001$ have been found for the following genes: ADA, ADBR2 and ADBR3, ATP1A2, ENPP1, GNB3, MRT2C, IL6R, INS, LEPR, MC3R, NR3C1, PPARG, PYY, RETN, UCP1, UCP2 and UCP3. (Tables 51.2 and 51.3).

Linkage studies are designed to investigate the persistence or co-segregation during generations of a genetic marker or locus within a family (Bray et al., 1998). This approach can be applied with candidate genes or genetic markers concerning specific polymorphisms. This protocol is commonly used on nuclear families or pedigrees in order to investigate the role of new genes through genome-wide scans (Bosse et al., 2004; Rankinen and Tiwari, 2004). Linkage studies with obesity phenotypes have been able to identify more than 200 QTLs related to excessive weight gain (Rankinen et al., 2006). Thus, linkage with obesity-related phenotypes (LOD > 2.0) has been reported for different genes or markers including the following: LEPR, AMPD1, ACP1, ISL1, NR3C1, BF, NPY, LEP, KEL, IGF1, MC5R, LDLR y ADA (Table 51.3). Genome-wide linkage scans involve the typing of family using polymorphic markers that are positioned across the whole genome followed by calculating the degree of linkage of the marker to a disease trait (Rankinen and Tiwari, 2004). In obesity, the sample sets are families representative of the general population and families that include at least an obese proband.

The 2005 update of the human obesity gene map revealed that the number of genes or markers that have been directly and indirectly linked with human obesity are increasing rapidly and now are higher than 625 (Figure 51.3). Some of these genes or chromosomal regions, such as the uncoupling proteins (UCPs), LEP, LEPR, adrenergic receptors (ADRB2, ADRB3), peroxisome proliferator-activated receptors (PPARs), fatty acid binding protein (FABP), etc. have been related to energy metabolism control and may be specifically affected by dietary intake and composition, and also by physical activity. Other genes are specifically involved in control of food intake (NPY, POMC, CCK, MCH, etc.), while some others influence different metabolic and signaling pathways, adipogenesis, etc. (PPAR, FABP, PKA, c/EBP, etc.), affecting the energy equation and, consequently, body weights of those individuals who are carriers of specific defective gene mutations/polymorphisms with an influence on fat deposition (Rankinen et al., 2006).

The studies of interactions of the genotype with environmental factors constitute a new challenge to establish the role of diet and physical activity in the genetics of obesity as well as those investigations devoted to evaluate the impact of lifestyle on gene expression (Cheung and Spielman, 2002). Indeed, genes predisposing to obesity potentially have an impact on dietary intake as well as on physical activity performance, while gender and age have been also reported as effect modifiers on

TABLE 51.4 Some Mendelian syndromes with obesity manifestations

	Prader Willi	Bardet-Biedel	Alstrom	Borjesson Forssman Lenhman	Simpson Golabi Behmel
Inheritance	Dominant 15q11.2	Recessive (> 8 loci)	Recessive 2p13.1	X-linked Xq26.3	X-linked (2 loci)
Craniofacies abnormalities	Frequent	No distinctive	No distinctive	Frequent	Frequent
Limbs/hands abnormalities	Swell hands and limbs short stature	Polydactily Brachidactily Syndactyly	No distinctive	Short stature Short toes	Big hands and short finger Very tall syndachtyly
Obesity	Generalized onset (1–3 years)	Mild obesity onset (1–2 years)	Truncal onset (1–2 years)	Truncal obesity	Slight obesity
Mental retardation	Mild to moderate	Not always	No	Severe	Psychosocial impairment
Other common features	Hypogonadism Muscular hypotonia Depigmentation	Hypertension diabetes mellitus Hypogonadism Renal abnormalities Retina dystrophy	Diabetes blindness deafness	Hypometabolism epilepsy	Cardiac arrhythmias

obesity risk in subjects carrying different gene polymorphisms, respectively (Marti et al., 2004). Several association studies have described interactions between macronutrient intake and different polymorphisms in candidate genes affecting the obesity trait in predisposed subjects (Martínez et al., 2003; Moreno-Aliaga et al., 2005; Nieters et al., 2002).

Genes Involved in Appetite Regulation

A number of genes and chromosomal regions have involved in the regulation of the food intake through association or linkage studies such as leptin, neuropeptide Y, ghrelin, AGRP, orexins, CART, etc. (Riccardi et al., 2004). In this context, a monogenic form of obesity in humans occurs as a consequence of very rare mutations of the leptin gene, which lead to undetectable levels of serum leptin (Clement, 2005). Also, members of a consanguineous family with a rare LEPR deficiency were identified, exhibiting severe early-onset obesity with several metabolic disturbances. Affected individuals were homozygous for a mutation that truncates the receptor before the transmembrane domain, and the mutated receptor circulates bound to leptin. Additionally, several more common mutations at the LEPR gene have been described, but only some of the studies have been able to show an association with overweight and fat mass excess (Perusse et al., 2005).

Much attention has been focused on the role of the hypothalamic LEP–melanocortin system in body weight regulation and obesity (Hebebrand et al., 2003). The importance of the melanocortin signaling pathway in humans has been suggested by a number of monogenic polymorphisms identified in genes involved in the synthesis or processing of the glycoprotein POMC or in mutations leading to defects in POMC signaling via melanocortin receptors. All mutations result in profound obesity; however, specific gene mutations in the LEP/POMC pathway account for 5% of cases of obesity. In this context, dominant inheritance of obesity conferred by different missense, nonsense and frame-shift mutations in the MC4R gene has been extensively reported in many populations including European and American individuals (Hebebrand et al., 2003). Genetic variation in serotonin receptors, NPY, orexins and other neuropeptides, has also been widely studied in obesity given its potential involvement on food intake and body weight gain (Solomon and Martínez, 2006).

Genes Involved in Energy Expenditure Regulation

There are a number of studies that have evaluated the role of specific polymorphisms in relation to the control of the energy expenditure (Bastarrachea et al., 2004). Thus, adrenergic receptor genes mediate the rate of lipolysis in response to endogenous and exogenous catecholamines. The Trp64Arg variant of the β_3-adrenergic receptor have been associated in some studies to lipid accumulation in the adipose tissue and also there are studies concerning the β_2 and α_2 adrenoceptors in the same direction concerning different polymorphisms (Bray and Bouchard, 2004). Also genes controlling the expression of the inner mitochondrial uncoupling proteins (UCP1, UCP2 and UCP3) have been related to obesity phenotypes (Moreno et al., 2005). A number of human's studies have found relationships between UCP polymorphisms and exercise efficiency resting energy expenditure, substrate oxidation, energy metabolism, BMI, fat accumulation and body weight changes (Macho et al., 2000).

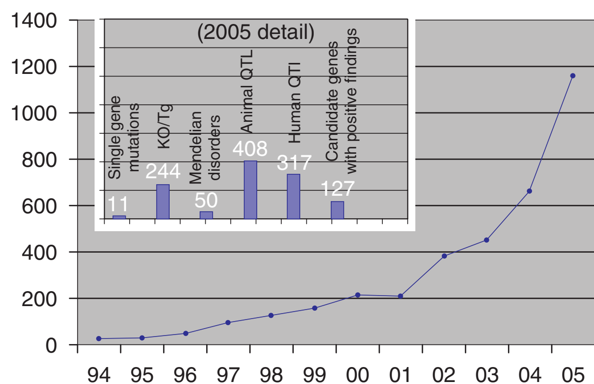


Figure 51.3 Evolution of the genes and chromosomal regions related to obesity between 1994–2005.

Genes Involved in Adipocyte Metabolism

Some genes involved in the regulation of adipocyte growth and differentiation have been associated with regulation of metabolism and body weight control (Bray and Bouchard, 2004). Thus, PPARγ particularly the adipose specific isoform (PPARγ2) is a key transcription factor implicated in adipogenesis as well as glucose and lipid homeostasis. A higher BMI for Ala carriers was also observed compared to non-carriers in 3-year and 10-year follow-up studies (Moreno-Aliaga et al., 2005). Another example is adiponectin, which is an adipocyte expressed protein that participates in the homeostatic control of glucose, lipid, and energy metabolism. Adiponectin gene polymorphisms have been associated with obesity, insulin sensitivity and type 2 diabetes in some cross-sectional studies (Rubio, 2006). Also, PGC1 is a cofactor regulating the expression of UCPs, the mitochondrial biogenesis, and other processes related to adipocyte energy homeostasis, whose variants are being related to excessive weight gain (Bray and Bouchard, 2004). Furthermore, the renin-angiotensin system is involved in adipocyte growth and differentiation.

Other Genes Related to Obesity

Common allelic variants of genes related to the insulin signaling pathway have been evaluated in relation to energy metabolism control such as the insulin gene, insulin-like growth factor 1 receptor (IGF-1R), plasma cell membrane glycoprotein 1, insulin receptor substrates 1 and 2 (IRS-1 and IRS-2), and the phosphatidylinositol 3-kinase p85 gene. Also, resistin, PAI, and different genes involved in lipid metabolism have been shown to potentially play a role in energy homeostasis and obesity such those codifying hepatic lipase (HL), lipoprotein lipase (LPL) and several lipoproteins (Rubio, 2006). Polymorphism of angiotensin converting enzyme (ACE) gene has been associated with overweight and abdominal adiposity, insulin resistance and hypertension in humans. Moreover, heterotrimeric guanine nucleotide-binding proteins (G proteins) mediate many pathways including the β-adrenergic signaling pathway. The C825T polymorphism in the gene coding for the β_3 subunit of G proteins (GNB3) has been shown to be associated with several phenotypes such as hypertension, obesity, and diabetes mellitus comprising the metabolic syndrome, although the conclusions have been inconsistent (Moreno et al., 2005). The fat mass and obesity gene (FTO) has also shown to contribute to common forms of human obesity (Loos and Bouchard, 2008).

SCREENING AND DIAGNOSIS

Epidemiologic, genetic and molecular studies suggest that there are people who are most susceptible than other to becoming overweight or obese (Bray and Bouchard, 1998). The degree of excess of fat, its distribution within the body and the associated health consequences depends on the genetic background (Moreno et al., 2005). Therefore, it is important to adequately assess the obesity phenotype not only for diagnosis, but also for therapeutic purposes. Thus, the graded classification of overweight and obesity aims for meaningful comparison of weight status among populations and identification of individuals/groups at risk, which helps to set up the intervention. Body weight for height and BMI (Kg/m^2) are common measures to define the obesity situation, which are suitable for general population samples, despite being relatively crude (OMS Report, 2000). In this context, other anthropometric or body composition determinations can be used such as the waist–hip ratio, percentage body fat, circulating leptin levels, etc. Other tools to characterize the obesity phenotype are based on the application of methods to assess food intake such as 72-h recall, food frequency questionnaires, dietary records, etc. Additional screening/diagnostic protocols aiming to identify the etiology of excessive weight gain are focused to estimate the energy expenditure (macronutrient oxidation and BMR) on the energy devoted to physical activity, through pedometers or calorimetry techniques, heart rate monitoring, sedentarism questionnaires, etc. (Table 51.5). Indeed, both appetite and physical activity patterns are potentially involved in obesity onset and complications, which must be adequately assessed.

Genomic medicine, aiming to improve the quality of medical care, is benefitting from the use of genotypic testing (DNA analysis) to identify genetic predisposition markers of obesity and to design individualized medical therapy based on the genotype (Challis and Yeo, 2002). Also, clinically oriented transcriptomics, focusing on the analysis of RNA by measuring the level of all or selected subset of genes based on the amount of mRNA in a given sample, are helping to understand and screen the causes of obesity (Moreno-Aliaga et al., 2001).

DNA sequence analysis is increasingly used as a diagnostic tool for determining carrier status and prenatal testing in monogenic disorders (Kasper et al., 2005). A number of techniques described in this textbook are available for the detection of mutations in monogenic disorders, which include cytogenetics, fluorescence in situ hybridization (FISH), Southern blotting, polymerase chain reaction based methods (PCR and RT-PCR),

Table 51.5 Methods to phenotypically characterize obesity

- *Body composition*: BMI, waist/hip circumferences, isotope dilution, bioelectrical impendence, skinfolds, etc.
- *Anatomical distribution of fat*: Imaging techniques (CT, NMR, DEXA, etc.)
- *Energy expenditure*: Stable isotopes, indirect/direct calorimetry, respiratory quotient, physical activity questionnaires, pedometers, heart rate, etc.
- *Energy intake*: Dietary records and recall, diaries, food frequency questionnaires, etc.
- *Molecular phenotypes*: Hormone levels, DNA, mRNA, and protein measurements

DNA sequencing and restriction fragment length polymorphisms (RFLP). Other approaches that have been used to identify mutations related to obesity are simple-strand conformational polymorphism (SSCP), gradient gel electrophoresis (DGGE) RNAse cleavage and protein truncated test (PTT). Novel technologies for the analysis of mutations, genetic mapping and mRNA expression profiles are rapidly developing (Ross, 1998; Ommen, 2004). Thus, microarray-based methods allowing the hybridization of thousands of genes simultaneous are being clinically used for mutational analysis and, characterizations of gene expression patterns in both obesity screens/diagnosis and research in humans and animals models (Kasper et al., 2005; Moreno-Aliaga et al., 2001).

The application of complex analytical techniques to encompass all the types of studies has led to the conclusion that the heritability estimate for BMI in large sample sizes was likely to be from 25 to 40%, while the amount of abdominal fat was influenced by a genetic component attaining about 50% of the individual differences (Bell et al., 2005; Comuzzie and Allison, 1998). Also, it should be stressed that the current epidemic of obesity is not really a consequence of the sudden appearance of mutations in genes regulating weight homeostasis, but the lack of adaptation in gene expression to environmental factors, such as the consumption of high-energy-yielding foods and sedentary habits.

Possible mechanisms through which a genetic susceptibility may be operating is the low body metabolism rate, low levels of macronutrient oxidation rates, low fat-free mass, low appetite control or low RQ post-prandial adjustments to intake or low oxidative process associated to physical activity (OMS Report, 2000). Other factors involved in the development of obesity thought to be genetically modulated are associated to impairments on adipose tissue, lipid metabolism, adipogenesis, on nutrient partitioning as well as endocrine/neural disturbances related to energy homeostasis such as insulin sensitivity, GH status, leptin action, hypothalamic neuropeptides, etc. (Martinez et al., 2002). Indeed, it is envisaged that genetic analyses, specifically in high-risk individuals, may be used for gene-based diagnosis and more interestingly for gene-founded therapies concerning genotypes affecting food intake, thermogenesis, adipogenesis, etc. In this context, gene polymorphism analysis could orient clinical recommendations to high-risk patients, since physical activity programs may be more suitable for some genotypes (Corbalan et al., 2002) or changes in carbohydrate or lipid intake may be more convenient for some specific gene variants (Luan et al., 2001; Martínez et al., 2003).

PROGNOSIS AND GENE-BASED TREATMENTS

The health consequences of obesity are diverse and serious, ranging from an increased risk of premature death to higher associated risks of suffering for diabetes, hypertension, inflammatory disturbances, cancer, etc. (Bray and Bouchard, 2004). The obesity prognosis depends upon a number of anatomical-linked phenotypical characteristics, such as the number of fat cells and fat distribution as well the causal factors, such a neuroendocrine disorders, drug-induced, weight gain sedentary lifestyles, psychological misbehaviors, which may be genetically dependent (Bray, 2003). Indeed, the prognosis and the treatments are affected by the genetic make-up, since genetic predisposition may influence the associated clinical manifestations and complications commonly found in obese subjects (Lean, 1998). This situation envisages that genetically based diagnosis and therapeutical approaches will be routinely applied in the obesity area as is occurring in some others diseases such as inborn errors of metabolism or diseases related to lipid metabolism (Basse et al., 2004; Kasper et al., 2005). Actually, it is of current interest to identify the involvement of some genetic polymorphism in relation to the ability of losing weight induced by different types of interventions commonly prescribed to obese patients such as diet therapy, pharmacotherapy and surgical approaches. Indeed, on the horizon is the possibility to individually treat obesity subjects depending on the genotype (Table 51.6) as reviewed by Moreno-Aliaga et al. (2005); a role for pharmacogenomics is not immediately clear since not many specific drugs to fight against obesity are currently in the market (Moreno et al., 2005).

In this context, congenital leptin deficiency is an extreme case of genetic obesity, whose deleterious effects can be restored by the administration of exogenous recombinant human leptin. In this special situation, the success of the pharmacological treatment with leptin would completely depend on the genetic make-up of the patient (O'Rahilly et al., 2003). By contrast, initial trials in obese patients with no leptin deficiency showed no significant benefits in terms of weight reduction unless substantially high doses were given. Using patients with polygenic (common) obesity, the response to low-calorie diet in relation with genetic polymorphisms in the promoter region of the LEP gene was studied (Mammes et al., 1998). At position C-2549A of the LEP gene, a weaker BMI reduction was found for −2549A allele carriers than the observed for homozygotes for the −2549C allele.

Some attention has been paid to the potential effects of Trp64Arg polymorphism in the ADBR3 gene in weight reduction programs (Table 51.6). Initially, resistance to lose weight

TABLE 51.6 Candidate genes and prognosis of weight loss response to dietary programs

Gene	Polymorphism*	Comments
Genes related to appetite control		
Leptin	C-2549 A (5-region)	Potential influence of BMI reduction
POMC	R236G	Weight loss is apparently unaffected
HTR2C	759C/T	Lower resistance to weight loss
Genes related to energy expenditure		
ADB3R	Trp64Arg	Potential influence in weight and fat distribution losses
ADB2R	Gln27Glu	Participation in rebound weight gain
UCP1	A-3826G	Potential impact on the magnitude of weight loss
UCP3	−55C⟶T (promoter)	Potential role on body fat distribution loss
Adipogenesis		
PPARG2	Pro12Ala	Potential involvement in maintenance after weight loss
Genes related to insulin resistance insulin signaling pathway		
IRS-1	Gly972Arg	No apparent role
IRS-2	G1057D	No apparent role
IGF-1R	GAA1013GAA	No apparent role
Genes related to lipid metabolism		
LIPC	G-250G (Promoter)	Weight loss dependent on the genotype
LPL	Hind III polymorphism	Weight loss dependent on the genotype
Apolipoproteins	ApoE e4	Possible relation with responsiveness to diet
	ApoB/VNTR	Possible relation with responsiveness to diet
	ApoA-IV-1/2	Potential involvement in food intake
Other genes potentially related to obesity		
PAI-1	4G/5G (promoter)	Genotype dependent reduction in BMI
IL6	−174 G > C	Potential involvement in maintenance after weight loss

*Modified from Moreno-Aliaga et al. (2005) and Martinez et al. (2008). Polymorphism nomenclature is shown as reported by the authors.

in homozygotes for the Arg64 allele compared with higher reductions either in heterozygotes or Trp64 homozygotes have been reported. On the contrary, other investigations did not find differences in weight loss by Trp64Arg genotypes in women submitted to a low-calorie diet. It is worth mentioning the differences on design in these studies according to the selection of participants (obese versus non-obese, diabetic versus non-diabetic) or the characteristics of the weight loss protocol (Moreno-Aliaga et al., 2005). Additionally, a lower ratio of visceral to subcutaneous fat has been reported in Arg64 carriers compared to non-carriers, especially in pre-menopausal women after a 3-month reduction period. Interestingly, a loss of visceral tissue has been found in Arg64 carriers compared with non-carriers after the intervention, but no differences in either body composition or energy expenditure by genotype groups were found. The response to exercise as a weight-lowering strategy appears to be affected by polymorphisms in the β_2-adrenergic receptor gene (Macho-Azcarate et al., 2003). A high initial degree of body fat mass determined by the Gly16 allele for the β_2-adrenoceptor polymorphisms predict those individuals who will have a rebound weight gain after their initial successful weight loss (Masuo et al., 2005).

The effect of uncoupling proteins on weight loss has been assessed in different studies (Riccardi et al., 2004). Thus, a relevant

role for the A-3826G polymorphism of the UCP1 gene in the magnitude of weight loss has been suggested (Table 51.6). In opposition to these results, no differences by genotype in weight loss of obese patients submitted to a gastroplasty were found. On the other hand, comparing proton leak and UCP2 and UCP3 expression in skeletal muscle of diet-resistant and diet-responsive overweight women, it was found that weight loss, mitochondrial proton leak-dependent respiration, and expression of UCP3 mRNA was greater in diet-responsive than in diet-resistant subjects, while no changes were found in UCP2 mRNA levels. Also, some UCP3 polymorphisms were associated with the changes in mRNA levels and body weight loss (Cha et al., 2006).

Other studies have reported a possible synergistic effect of UCP1 and ADBR3 gene polymorphisms on obesity and body weight changes (Sivenius et al., 2000). Thus, subjects with mutations in ADBR3 and UCP1 genes lost less weight than either those with the Trp64Arg variant alone (ADRB3) or the GG genotype of the UCP1 gene. Other researchers found a lower weight reduction in subjects with both mutations compared with subjects with no mutations, and a faster weight gain after the ending of the very-low-calorie diet intervention, while similar weight reductions after a clinical intervention for the four defined genetic categories concerning the Trp64Arg polymorphism (ADRB3 gene) and a promoter polymorphism (−55C→T) of the UCP3 gene in overweight-obese subjects with coronary artery disease or metabolic syndrome. However, a beneficial effect of the wild-type genotype for both variants was suggested for body fat distribution and glycemic control (Table 51.6).

Also, some genotype-dependent food intake control related cases have been studied. A recent trial to compensate for the genetic lack of hypothalamic melanocortin function was conducted by the administration of adrenocorticotropic hormone 4–10 (ACTH4-10), a melanocortin fragment with anorexic effect. However, after 3 months of treatment with increasing doses of ACTH4-10, no change of body weight or metabolic rate was observed. On the other hand, it has been reported that the individual ability to lose weight was not hampered in three children heterozygous for the R236G variant of the POMC gene that underwent a weight reduction program during 11 months (Santoro et al., 2004). Also, the serotonin receptor HTR2C genotype (−759C/T polymorphism) on weight loss program, was involved in the resistance to weight loss in the heterozygous genotype compared with either CC or TT homozygous genotypes and attributed such observation to a heterosis effect.

Moreover, several intervention studies were aimed to explore the relationship between PPARγ2 genotypes and weight loss (Nieters et al., 2002). Thus a 3-year study showed that subjects with the Ala12 allele lost more weight than subjects with the other genotypes, although a significant higher weight regain was achieved during subsequent follow-up (12 months) in carriers of the Ala12 allele in comparison with the Pro12 homozygotes (Lindi et al., 2002).

These observations suggest that a general distinction should be made between genetic factors affecting weight loss *per se* and weight loss maintenance. Thus, also, it has been shown that a IL-6 polymorphism (−174 G > C) may influence weight regain after a weight-lowering program based on a energy-restricted diet, which may interact the Pro12Ala polymorphism of the PPARγ gene (Goyenechea et al., 2006). On the other hand, no significant differences in weight reductions linked to the genotype were detected, in the subjects carrying risk genotypes in the IGF-1R, IRS-1 and IRS-2 genes after lifestyle interventions (Moreno-Aliaga et al., 2005). However, a reinforced effect on weight loss with the presence of both the Gly972Arg polymorphism of the IRS-1 gene and Trp64Arg polymorphism in the ADRB3 gene in women who underwent a formal weight-loss program was described (Benecke et al., 2000).

In the Finnish DPS, it was found that subjects with the G-250G promoter polymorphism of the HL gene (LIPC) tended to loss less weight than subjects with the −250A allele both in the control and in the intervention group (Todorova et al., 2004). In addition, having the G-250G genotype predicted the conversion from impaired glucose tolerance to Type 2 diabetes independently of weight at baseline and weight change. Furthermore, numerous polymorphisms of the apolipoprotein B gene have been described, which have been accompanied by alterations in lipid metabolism (Bosse et al., 2004). A meta-analysis indicated that the apoB EcoRI and MspI polymorphisms are associated with responsiveness to diet. A study investigating the effects of apoA-IV genotype during weight loss found that subjects with apoA-IV-1/2 lost more weight than apoA-IV-1/1 subjects. On the contrary, it was found by others no significant differences in weight loss between both genotypes, although subjects with the apoA-IV-1/1 genotype showed a higher increase in HDL-C after in response to weight loss compared to the apoA-IV-1/2 subjects. Also the potential relationship between the PAI-1 promoter 4G/5G genotype and weight loss on the fibrinolytic system and lipid parameters in obese children was assessed and after a 3-month period of treatment (nutritional counseling and physical activity). A decrease in PAI-1 levels was observed in obese children, who have reduced their BMI in comparison with those obese children who did not decrease their BMI (Estelles et al., 2001). On the other hand, a behavioral interventional activity in obese carriers of the Trp64Arg polymorphism of the ADBR3 gene demonstrated difficulties in losing weight (Shiwaku et al., 2003).

Additionally, a large scale European intervention trial (NUGENOB) concerning the comparison of the impact of more than 40 genetic polymorphisms on weight loss revealed that much work is required to be performed in this area (Sorensen et al., 2006). Other studies have been reported, by using microarrays, that mRNA levels (gene expression) could be affected by the energy restriction, but not apparently by the composition of the hypocaloric diet (low versus moderate fat content), suggesting that more research is needed in this field (Viguerie et al., 2005b).

2009 UPDATE

Obesity is nowadays considered as the result of a dysregulation of energy homestasis inducing excesive weight gain and fat deposition, which is explained in most cases by the interactions of the genetic background with inadequate dietary habits, physical inactivity, especific diseases, drug-side effects, etc (Ichihara and Yamada, 2008; Marti et al., 2008).

Current ongoing genome-wide association scans are yielding new loci and chromosomical regions related with measures of obesity (Thorleifsson et al., 2009). Such studies are also facilitating the detection and interpretation of the role of single nucleotide polymorphisms and common copy number variations, as well as epigenomic phenomena such as DNA methylation, histone modifications leading to chromatin remodelling and regulation of gene expression patterns that underlie the developmental programming of obesity and metabolic syndrome features (Campion et al., 2009b; Herbert, 2008; Junien and Nathanielsz, 2007). All such variations have a potential impact on genomic medicine.

In this context, new candidate genes and novel variants involved in obesity physiopathology such as FTO (Qi et al., 2008), MC4R (Krakoff et al., 2008), interleukin-15 (Nielsen et al., 2008), adiponectin (Goyenechea et al., 2009a), etc., have been described.

Further, the developments and advances in high-throughput genomics tools and –omics technologies such as proteomics, metabolomics, epigenomics, transcriptomics and integrative system biology approaches are having an impact on predicting obesity related outcomes (Mutch et al., 2008) and providing new views concerning the role of the adipose tissue and adipokines (Dahlman and Arner, 2009; Herbert, 2008), inflamation and oxidative stress factors (Crujeiras et al., 2008a; Goyenechea et al., 2008), mitocondrial functions (Højlund et al., 2008) and metabolic efficiency (Sorensen., 2008) among others, which will contribute to "personalize" medicine and nutrition (Goyenechea et al., 2009b).

Upcoming new molecular procedures devised to transfer specific genetic material (Cordero et al., 2008) as well as the development of powerful mRNA profiling protocols (Capel et al., 2008; Crujeiras et al., 2008b; Goyenechea et al., 2009b), microRNA analyses (Lovis et al., 2008) and methylated DNA measurements (Campion et al., 2009a) are enabling us to assess and characterize different cell functions and mechanisms under genetic or epigenetic regulation concerning genes such as PPAR, sirtuins, TNF-alfa, etc., which are contributing to identify new pharmacogenetic and nutrigenomic targets (Crujeiras et al., 2008b; Goyenechea et al., 2009b) based on the genotype and, therefore, to a more "individualized" and genomic style of medicine.

2009 UPDATE REFERENCES

Campión, J., Milagro, F., Goyenechea, E. and Martínez, J.A. (2009a). TNF-alpha promoter methylation as a predictive biomarker for weight-loss response. *Obesity*, in press.

Campión, J., Milagro, F. and Martínez, J.A. (2009b). Epigenetic patterns and individuality in obesity: Epiobesigenes. *Obesity Rev*, Apr 1 [Epub ahead of print].

Capel, F., Viguerie, N., Vega, N., Dejean, S., Arner, P., Klimcakova, E., Martinez, J.A., Saris, W.H., Holst, C., Taylor, M. et al. (2008). Contribution of energy restriction and macronutrient composition to changes in adipose tissue gene expression during dietary weight-loss programs in obese women. *J Clin Endocrinol Metab* 93, 4315–4322.

Cordero, P., Campion, J., Milagro, F.I., Marzo, F. and Martinez, J.A. (2008). Fat-to-glucose interconversion by hydrodynamic transfer of two glyoxylate cycle enzime genes. *Lipids Health Dis* 7, 49.

Crujeiras, A.B., Parra, D., Milagro, F.I., Goyenechea, E., Larrarte, E., Margareto, J. and Martínez, J.A. (2008a). Differential expression of oxidative stress and inflammation related genes in peripheral blood mononuclear cells in response to a low-calorie diet: A nutrigenomics study. *OMICS* 12, 251–261

Crujeiras, A.B., Parra, D., Goyenechea, E. and Martínez, J.A. (2008b). Sirtuin gene expression in human mononuclear cells is modulated by caloric restriction. *Eur J Clin Invest* 38, 672–678.

Dahlman, I. and Arner, P. (2009). Obesity and polymorphisms in genes regulating human adipose tissue. *Int J Obes* 31, 1629–1641.

Goyenechea, E., Crujeiras, A.B., Abete, I., Parra, D. and Martínez, J.A. (2008). Enhanced short-term improvement of insulin response to a low-caloric diet in obese carriers the Gly482Ser variant of the PGC-1alpha gene. *Diabetes Res Clin Pract* 82, 190–196.

Goyenechea, E., Collins, L.J., Parra, D., Abete, I., Crujeiras, A.B., O'Dell, S.D. and Martínez, J.A. (2009a). The –11391 G/A polymorphism of the adiponectin gene promoter is associated with metabolic syndrome traits and the outcome of an energy-restricted diet in obese subjects. *Horm Metab Res* 41, 55–61.

Goyenechea, E., Parra, D., Crujeiras, A.B., Abete, I. and Martínez, J.A. (2009b). An inflammation-related approach to predict the weight-loss regain in obese subjects. *Ann Nutr Metab*, 54(1), 43–51.

Herbert, A. (2008). The fat tail of obesity as told by the genome. *Curr Opin Clin Nutr Metab Care* 11, 366–370.

Højlund, K., Mogensen, M., Sahlin, K. and Beck-Nielsen, H. (2008). Mitochondrial dysfunction in type 2 diabetes and obesity. *Endocrinol Metab Clin North Am* 37, 713–731.

Ichihara, S. and Yamada, Y. (2008). Genetic factors for human obesity. *Cell Mol Life Sci* 65, 1086–1098.

Junien, C. and Nathanielsz, P. (2007). Early and lifelong environmental epigenomic programming of metabolic syndrome, obesity and type II diabetes. *Obes Rev* 8, 487–502.

Krakoff, J., Ma, L., Kobes, S., Knowler, W.C., Hanson, R.L., Bogardus, C. and Baier, L.J. (2008). Lower metabolic rate in individuals

heterozygous for either a frameshift or a functional missense MC4R variant. *Diabetes* 57, 3267–3272.

Lovis, P., Roggli, E., Laybutt, D.R., Gattesco, S., Yang, J.Y., Widmann, C., Abderrahmani, A. and Regazzi, R. (2008). Alterations in microRNA expression contribute to fatty acid-induced pancreatic beta-cell dysfunction. *Diabetes* 57, 2728–2736.

Marti, A., Martinez-González, M.A. and Martínez, J.A. (2008). Interaction between genes and lifestyle factors on obesity. *Proc Nutr Soc* 67, 1–8.

Mutch, D.M., Temanni, M.R., Henegar, C., Combes, F., Pelloux, V., Holst, C., Sørensen, T.I., Astrup, A., Martinez, J.A., Saris, W.H. et al. (2007). Adipose gene expression prior to weight loss can differentiate and weakly predict dietary responders. *PLoS ONE* 2, e1344.

Nielsen, A.R., Hojman, P., Erikstrup, C., Fischer, C.P., Plomgaard, P., Mounier, R., Mortensen, O.H., Broholm, C., Taudorf, S., Krogh-Madsen, R. et al. (2008). Association between interleukin-15 and obesity: Interleukin-15 as a potential regulator of fat mass. *J Clin Endocrinol Metab* 93, 4486–4493.

Qi, L., Kang, K., Zhang, C., van Dam, R.M., Kraft, P., Hunter, D., Lee, C.H. and Hu, F.B. (2008). Fat mass-and obesity-associated (FTO) gene variant is associated with obesity: Longitudinal analyses in two cohort studies and functional test. *Diabetes* 57, 3145–3151.

Sørensen, T.I. (2008). Challenges in the study of causation of obesity. *Proc Nutr Soc* 15, 1–12.

Thorleifsson, G., Walters, G.B., Gudbjartsson, D.F., Steinthorsdottir, V., Sulem, P., Helgadottir, A., Styrkarsdottir ,U., Gretarsdottir, S., Thorlacius, S., Jonsdottir, I. et al. (2009). Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet* 41, 18–24.

ACKNOWLEDGEMENTS

The author of the chapter wishes to express his appreciation and recognition to the following coworkers and colleagues by generous and valuable collaboration in this field: J. Bresan, J. Campion, B. de Fanti, P. González-Muniesa, A. Marti, M.A. Martínez-Gonzalez, F. Milagro, M.J. Moreno-Aliaga, M.D. Parra, P.C. Pérez-Matute, M.P. Portillo, J.L. Santos, A. Solomon and M.A. Zulet as well as to members of the NUGENOB and DIOGENES Consortia.

REFERENCES

Barsh, G.S., Farooqui, I.S. and O'Rahilly, S. (2000). Genetics of body-weight regulation. *Nature* 404, 644–651.

Bastarrachea, R., Cole, S. and Comuzzie, A. (2004). Genómica de la regulación del peso corporal: mecanismos moleculares que predisponen a la obesidad. *Medicina Clinica* 3, 104–117.

Bell, C.G., Walley, A.J. and Froguel, P. (2005). The genetics of human obesity. *Nat Rev* 6, 221–233.

Benecke, H., Topak, H., Von Zur Muhlen, A. and Schuppert, F. (2000). A study on the genetics of obesity: Influence of polimorphyms of the beta-3-adrenergic receptor and insulin receptor substrate 1 in relation to weight loss, waist to hip ratio and frequencies of common cardiovascular risk factors. *Exp Clin Endocrinol Diabetes* 108, 86–92.

Bes-Rastrollo, M., Sanchez-Villegas, A., Gomez-Gracia, E., Martinez, J.A., Pajares, R.M. and Martinez-Gonzalez, M.A. (2006). Predictors of weight gain in a Mediterranean cohort: The Seguimiento Universidad de Navarra Study. *Am J Clin Nutr* 83, 189–190.

Bjorntorp, P. (2001). *International Textbook of Obesity*. Wiley and Sons, New York.

Bosse, Y., Chagnon, Y.C., Despres, J.P., Rice, T., Rao, D.C., Bouchard, C., Perusse, L. and Vohl, M.C. (2004). Compendium of genome-wide scans of lipid-related phenotypes: Adding a new genome-wide search of apolipoprotein levels. *J Lipid Res* 45, 2174–2184.

Bouchard, C. (1997). Genetics of human obesity: Recent results from linkage studies. *J Nutr* 127, 1887–1890.

Bouchard, C. and Tremblay, A. (1997). Genetic influences on the response of body fat distribution to positive and negative energy balances in human identical twins. *J Nutr* 127, S943–S947.

Bray, G. (1976). *The Obese Patient*. Saunders Co, Philadelphia.

Bray, G.A. (2003). *Contemporary Diagnosis and Management of Obesity*, Handbooks in Health Care, Newton.

Bray, G.A., Bouchard, C. and James, W.P.T. (1998). *Handbook of Obesity*. Marcel Dekker, New York.

Bray, G.A. and Bouchard, C. (2004). *Handbook of Obesity: Etiology and Pathophysiology*. Wiley and sons, New York.

Cha, M.A., Shin, H.D., Kim, K.S., Lee, B.H. and Yoon, Y. (2006). The effects of uncoupling protein 3 haplotypes on obesity phenotypes and very low-energy diet induced changes among overweight Korean female subjects. *Metabolism* 55, 578–586.

Challis, B.G. and Yeo, G.S. (2002). Past, present and future strategies to study the genetics of body weight regulation. *Brief Funct Genomic Proteomic* 3, 290–304.

Cheung, V.G. and Spielman, R.S. (2002). The genetics of variation in gene expression. *Nat Genet* 32, 522–525.

Clement, K. (2005). Genetics of human obesity. *Proc Nutr Soc* 64, 133–142.

Comuzzie, A.G. and Allison, D.B. (1998). The search for human obesity genes. *Science* 280, 1374–1377.

Corbalan, M.J., Marti, A., Forga, C., Martínez-Gonzalez, M.A. and Martínez, J.A. (2002). The 27Glu polymorphism of the beta2-adrenergic receptor gene interacts with physical activity influencing obesity risk among female subjects. *Clin Genet* 61, 305–307.

Davenport, C.B. (1925). *Body Built and its Inheritance*, Carnegie Institute of Washington.

Estelles, A., Dalmau, J., Falco, C., Berbel, O., Castello, R., España, F. and Aznar, J. (2001). Plasma PAI-1 levels in obese children-effect of weight loss and influence of PAI-1 promoter 4G/5G genotype. *Thromb Haemost* 86, 647–652.

Feinleib, M., Garrison, R.J., Fabsitz, R., Christian, J.C., Hrubec, Z., Borhani, N.O., Kannel, W.B., Rosenman, R., Schwartz, J.T. and Wagner, J.O. (1977). The NHLBI twin study of cardiovascular

disease risk factors: Methodology and summary of results. *Am J Epidemiol* 106, 284–285.

Garrow, J. (1974). *Energy Balance and Obesity in Man*. North Holland Publishing Co, Amsterdam.

Goyenechea, E., Parra, M.D. and Martínez, J.A. (2006). Weight regain after slimming induced by an energy-restricted diet depends on interleukin-6 and peroxisome-proliferator-activated-receptor-gamma2 gene polymorphisms. *Br J Nutr* 96, 1–9.

Guy-Grand, B. (2003). From obesity to obesities: From concepts to practices. *Ann Endocrinol* 64, 7–15.

Hebebrand, J., Friedel, S., Schauble, N., Geller, F. and Hinney, A. (2003). Perspectives: Molecular genetic research in human obesity. *Obes Rev* 4, 139–146.

Hill, J.O., Wyatt, H.R., Reed, G.W. and Peters, J.C. (2003). Obesity and the environment: Where to we go from here?. *Science* 299, 853–855.

Jacobson, P., Rankinen, T., Tremblay, A., Pérusse, L., Chagnon, Y.C. and Bouchard, C. (2006). Resting metabolic rate and respiratory quotient: Results from a genome-wide scan in the Quebec Family Study. *Am J Clin Nutr* 84, 1527–1533.

Jequier, E. and Tappy, I. (1999). Regulation of body weight in humans. *Physiol Rev* 79, 451–480.

Kaput, J., Ordovas, J.M., Ferguson, L., van Ommen, B., Rodriguez, R.L., Allen, L., Ames, B.N., Dawson, K., German, B., Krauss, R. et al. (2005). The case for strategic international alliances to harness nutritional genomics for public and personal health. *Br J Nutr* 94, 623–632.

Kasper, D.L., Braunwald, E., Fauci, A., Hauser, S., Longo, D. and Jameson, J.L. (2005). *Harrison's Principles of Internal Medicine*. McGraw-Hill, New York.

Katzmarzyk, P.T., Perusse, L., Rao, D.C. and Bouchard, C. (1999). Familial risk of obesity and central adipose tissue distribution in the general Canadian population. *Am J Epidemiol* 15, 933–942.

Lean, M. (1998). *Clinical Handbook of Weight Management*. Dunitz, London.

Loss, R.J. and Bouchard, C. (2008). </pubmed/18373508?ordinalpos=1&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_RVDocSum> FTO: the first gene contributing to common forms of human obesity. *Obes Rev* 9, 246–250.

Lopez, I.P., Milagro, F.I., Marti, A., Moreno-Aliaga, M.J., Martínez, J.A. and De Miguel, C. (2005). High-fat feeding period affects gene expression in rat white adipose tissue. *Mol Cell Biochem* 275, 109–115.

Luan, J., Browne, P.O., Harding, A.H., Halsall, D.J., O'Rahilly, S., Chatterjee, V.K. and Wareham, N.J. (2001). Evidence for gene-nutrient interaction at the PPARgamma locus. *Diabetes* 50, 686–689.

Lyon, H.N. and Hirschhorn, J.N. (2005). Genetics of common forms of obesity: A brief overview. *Am J Clin Nutr* 82, 2152–2157.

Macho, T., Marti, A. and Martinez, J.A. (2000). Estudios genéticos de la obesidad en humanos. *Medicina Clinica* 115, 103–110.

Macho-Azcarate, T., Marti, A., González, A., Martínez, J.A. and Ibañez, J. (2002). Gln27Glu polymorphism in the beta 2 adrenergic receptor gene and lipid metabolism during exercise in obese women. *J Obes Relat Metab Disord* 26, 1434–1441.

Mammes, O., Betoulle, D., Aubert, R., Giraud, V., Tuzet, S., Petiet, A., Colas-Linhart, N. and Fumeron, F. (1998). Novel polymorphisms in the leptin gene and response to low calorie diet. *Diabetes* 47, 487–489.

Marti, A., Moreno-Aliaga, M.J., Ochoa, M.C., Marrades, P., Santos, J.L. and Martínez, J.A. (2003). Genética de la obesidad en humanos: Rumbos de investigación. *Rev Esp Obes* 2, 351–362.

Marti, A., Moreno-Aliaga, M.J., Hebebrand, J. and Martinez, J.A. (2004). Genes, lifestyles and obesity. *Int J Obes* 28, 29–36.

Marti, A., Martinez-González, M.A. and Martinez, J.A. (2008). Interaction between genes and lifestyle factors on obesity. *Proc Nutr Soc* 67, 1–8.

Martínez, J.A. (2000). Body-weight regulation: Causes of obesity. *Proc Nutr Soc* 59, 337–445.

Martínez, J.A. and Frühbeck, G. (1996). Regulation of energy balance and adiposity: A model with new approaches. *J Physiol Biochem* 52, 255–258.

Martínez, J.A., Kearney, J.M., Kafatos, A., Paquer, S. and Martinez-Gonzalez, M.A. (1999). Variables independently associated with self-reported obesity in the European Union. *Pub Health Nutr* 2, 125–135.

Martínez, J.A., Moreno-Aliaga, M.J., Marques-Lopes, I. and Marti, A. (2002). Causes of obesity. *Anal Sis San Nav* 25, 17–27.

Martinez, J.A., Parra, M.D., Santos, J.L., Moreno-Aliaga, M.J., Marti, A. and Martinez-Gonzalez, M.A. (2008). </pubmed/18296317?ordinalpos=1&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_RVDocSum > Genotype-dependent response to energy-restricted diets in obese subjects: towards personalized nutrition. *Asia Pac J Clin Nutr* 17, 119–122.

Martínez, J.A., Corbalan, M.S., Sánchez-Villegas, A., Forga, L., Marti, A. and Martínez-Gonzalez, M.A. (2003). Obesity risk is associated with carbohydrate intake in women carrying the Gln27Glu beta2-adrenoceptor polymorphism. *J Nutr* 133, 2549–2554.

Masuo, K., Katsuya, T., Kawaguchi, H., Fu, Y., Rakugi, H., Ogihara, T. and Tuck, M. (2006). Rebound weight gain as associated with high plasma norepinephrine levels that are mediated through polymorphisms in the beta 2-adrenoceptor. *AJH* 18, 1508–1516.

Moreno, B., Monereo, S. and Álvarez, J. (2005). *La obesidad en el Tercer Milenio. 3ª edición*, Editorial Médica Panamericana, Buenos Aires.

Moreno-Aliaga, M.J., Marti, A., García Foncillas, J. and Martínez, J.A. (2001). DNA hybridization arrays: A powerful technology for nutritional and obesity research. *Br J Nutr* 2, 119–122.

Moreno-Aliaga, M.J., Santos, J.L., Marti, A. and Martínez, J.A. (2005). Does weight loss prognosis depend on genetic make-up? *Obes Rev* 6, 155–168.

Neel, J.V. (1962). Diabetes mellitus: A "thrifty" genotype rendered detrimental by "progress"?. *Am J Hum Genet* 14, 353–362.

Nieters, A., Becker, N. and Linseisen, J. (2002). Polymorphisms in candidate obesity genes and their interaction with dietary intake of *n*-6 polyunsaturated fatty acids affect obesity risk in a sub-sample of the EPIC-Heidelberg cohort. *Eur J Nutr* 41, 210–221.

O'Rahilly, S., Farooqi, I.S., Yeo, G.S. and Challis, B.G. (2003). Minireview: Human obesity-lessons from monogenic disorders. *Endocrinology* 144, 3757–3764.

Ommen, B.V. (2004). Nutrigenomics: Exploiting systems biology in the nutrition and health arenas. *Nutrition* 20, 4–8.

OMS Report (2000). *Obesity: Preventing and Managing the Global Epidemic*. Technical report 894, Geneva.

Palou, A., Pico, C. and Bonet, M.L. (2003). Molecular basis of body weight control. *Forum Nutr* 56, 164–168.

Perusse, L. and Bouchard, C. (1999). Genotype-environment interactions in human obesity. *Nutr Rev* 57, S31–S37.

Perusse, L., Rankinen, R., Zuberi, A., Chagnon, Y.C., Weisnagel, S.J., Argyropoulos, G., Walts, B., Snyder, E.E. and Bouchard, C. (2005). The human obesity gene map: The 2004 update. *Obes Res* 13, 381–490.

Rankinen, T. and Tiwari, H. (2004). Genome scans for human nutritional traits: What have we learned?. *Nutrition* 20, 9–13.

Rankinen, T., Zuberi, A., Chagnon, Y.C., Weisnagel, S.J., Argyropoulos, G., Walts, B., Perusse, L. and Bouchard, C. (2006). The human obesity gene map: The 2005 update. *Obesity* 14, 529–644.

Riccardi, G., Aggett, P., Brighenti, F., Delzenne, N., Frayn, K., Nieuwenhuizen, A., Pannemans, D. et al. (2004). PASSCLAIM – Body weight regulation, insulin sensitivity and diabetes risk. *Eur J Nutr* 43, S7–S46.

Rosmond, R. (2003). Association studies of genetic polymorphisms in central obesity: A critical review. *Int J Obes Relat Metab Disord* 27, 1141–1151.

(2006). . Editorial Médica Panamericana, Madrid.

Santoro, N., Miraglia del Giudice, E. and Grazia, C. (2004). Weight loss in obese children with the POMC R236G variant. *Int J Obes Relat Metab Disord* 28, S30.

Seidell, J.C. and Flegal, K.M. (1997). Assessing obesity: Classification and epidemiology. In *Obesity*, Vol. 53, RSM Press, 238–252.

Shiwaku, K., Nogi, A., Anuurad, E., Kitajima, K., Enkhmaa, B., Shimono, K. and Yamane, Y. (2003). Difficulty in losing weight by behavioural intervention for women with Trp64Arg polymorphism of the beta3-adregenic receptor gene. *Int J Obes Relat Metab Disord* 27, 1028–1036.

Shuldiner, A.R. and Munir, K.M. (2003). Genetics of obesity: More complicated than initially thought. *Lipids*, 38, 97–101, New York.

Sivenius, K., Valve, R., Lindi, V., Niskamen, L., Laakso, M. and Uusitupa, M. (2000). Synergistic effect of polymorphism in uncoupling protein 1 and beta 3 adrenergic receptor genes on long-term body weight change in Finnish type 2 diabetic and non-diabetic control subjects. *Int J Obes Relat Metab Disord* 24, 514–519.

Solomon, A. and Martínez, J.A. (2006). Involvement of the nervous and gastrointestinal system on the energy homeostasis. *Rev Med Univ Navarra* 50, 27–37.

Sorensen, T.I.A., Boutin, P., Taylor, M.A., Larsen, L.H., Verdich, C., Petersen, L., Holst, C., Echwald, S.M., Dina, C., et al. (2006). Genetic polymorphisms and weight loss in obesity: A randomised trial of hypo-energetic high- versus low-fat diets, *PLOS Clinical Trials*. (in press).

Todorova, B., Kubaszed, A., Pihjlajamaki, J., Lindstrom, J., Eriksson, J., Valle, T.T., Hamalainen, H. et al. (2002). Diabetes prevention Study. The G-250A promoter polymorphism of the hepatic lipase gene predicts the conversion from impaired glucose tolerance to type 2 diabetes mellitus: The Finnish diabetes prevention study. *J Clin Endocrinol Metab* 89, 2019–2023.

Tremblay, A., Perusse, L. and Bouchard, C. (2004). Energy balance and body-weight stability: Impact of gene–environment interactions. *Br J Nutr* 92, S63–S66.

Vague, J. (1947). The différentiation sexuelle, facteur determinant des formes d'obesité. *Presse Med* 55, 339.

Vague, J., Vague, P., Jubelin, J. and Barre, A.M. (1989). Sexual differentiation of the adipose tissue-muscle ratio. Its metabolic impact. *Bull Acad Natl Med* 173, 309–318.

Viguerie, N., Poitou, C., Cancello, R., Stich, V., Clement, K. and Langin, D. (2005a). Transcriptomics applied to obesity and caloric restriction. *Biochimie* 87, 117–123.

Viguerie, N., Vidal, H., Arner, P., Holst, C., Verdich, C., Avizou, S., Astrup, A., Saris, W.H., Macdonald, I.A., Klimcakova, E., Clement, K., Martinez, A., Hoffstedt, J., Sorensen, T.I. and Langin, D. (2005b). Adipose tissue gene expression in obese subjects during low-fat and high-fat hypocaloric diets. *Diabetologia* 48, 123–131.

CHAPTER

Diabetes

Maggie Ng and Nancy J. Cox

INTRODUCTION

Diabetes mellitus is a group of metabolic disorders characterized by high blood glucose levels resulting from defects in insulin secretion, insulin action or both. Chronic high blood glucose levels are associated with long-term tissue damage, dysfunction and failure of various organs especially the eyes, kidneys, nerves, heart and blood vessels. The most common forms of diabetes are type 1 diabetes, in which autoimmune destruction of the insulin-secreting pancreatic β-cells leads to an absolute insulin deficiency (or complete absence of endogenous insulin), and type 2 diabetes, which is characterized by a relative deficiency of insulin attributable to defects leading to insulin resistance, defects in insulin secretion, or both. Type 1 diabetes commonly arises in childhood or adolescence, and the lifetime risk of type 1 diabetes in US populations of European descent is about one in 250 (0.004), but lower in US populations of recent Asian, African or Native American ancestry. Type 2 diabetes is more frequently diagnosed in late adulthood and is far more common, with 8–10% of Americans of European descent diagnosed with type 2 diabetes by age 60. Age-specific rates of type 2 diabetes are even higher in some US populations, including African Americans, Mexican Americans and Native Americans, and there are alarming increases in the incidence of type 2 diabetes in younger Americans, due at least in part to increasing levels of obesity in the United States coupled with increasingly sedentary lifestyles (Narayan et al., 2003).

Some 1.6 million Americans have been diagnosed with type 1 diabetes and more than 16 million Americans with type 2 diabetes. And for many of these individuals, their family members are at a substantially increased risk relative to the population average; for example, sibs of a patient with type 1 diabetes have a 15-fold increased risk of disease (i.e., $\lambda_s = 15$), while sibs of a patient with type 2 diabetes have a 1.8–2.8-fold increased risk of disease (i.e., $\lambda_s = 1.8–2.8$) (Rich, 1990). Thus, because it is easy and inexpensive to diagnose, common and familial, diabetes has long been a model for exploring the genetic basis of complex disorders – those that cluster in families but do not have simple patterns of transmission (Florez et al., 2003). In addition, given especially the incidence of type 2 diabetes in the general adult population and the expectation that one or more genetic variants could alter individual risk of disease, it has become an attractive candidate for early thinking about the introduction of genetic and genomic concepts into medical practice.

Despite the efforts of many investigators over many years, though, association studies at candidate genes and genome-wide linkage studies have, until very recently, yielded only a small number of loci having genetic variation reproducibly shown to affect risk of either type 1 or type 2 diabetes. Prior to recent GWAS (see Chapter 4), susceptibility loci for type 1 diabetes with broad support over multiple studies included the HLA region, with primary determinants thought to be variation at the *HLA-DRB1*, *-DQA1* and *-DQB1* loci (Davies et al., 1994), the insulin gene (*INS*) (Bell et al., 1984), the gene for cytotoxic T-lymphocyte associated protein 4 (*CTLA4*) (Ueda et al., 2003) and the protein tyrosine phosphatase-22 gene (*PTPN22*) (Bottini et al., 2004). All of these were originally identified through candidate gene studies. With respect to linkage, in a large type 1 diabetes

Copyright © 2010, Elsevier Inc.
All rights reserved.

consortium dataset with 1435 multiplex families (Concannon et al., 2005), the significance of the evidence for linkage at the HLA region was quite dramatic, at $p < 2 \times 10^{-52}$, with the next best region located near *CTLA4*, with $p < 9 \times 10^{-5}$. Two other regions with notable evidence for linkage were 10p14-q11 with $p < 1.2 \times 10^{-4}$, and 16q22-q24 with $p < 4.9 \times 10^{-4}$. In contrast, there was only modest evidence for linkage at *INS*, and loci with relatively modest sibling risk ratios could be excluded from much of the rest of the genome ($\lambda_s > 1.3$ excluded from 82% of the genome; $\lambda_s > 1.5$ excluded from 95% of the genome).

Results of early candidate gene association studies as well as linkage and positional cloning studies in type 2 diabetes had yielded a similar paucity of genes and regions with widely reproducible linkage or association (for a recent review, see Permutt et al., 2005). Although a number of candidate genes, including candidates derived on the basis of effective treatments for type 2 diabetes such as *KCNJ11* and *PPARG*, had been validated through large-scale meta-analysis (Altshuler et al., 2000; Gloyn et al., 2003), the vast majority of candidate gene studies in type 2 diabetes had not yielded reproducible associations. Similarly, while a few linkage studies yielded positionally cloned genes with at least some support from other studies, including *CAPN10* (Horikawa et al., 2000) and *HNF4A* (Love-Gregory et al., 2004; Silander et al., 2004), only *TCF7L2* (identified by investigators at deCode in Iceland in positional cloning studies based on one of their best conditional linkage signals [Grant et al., 2006]) has been widely replicated in multiple populations (Cauchi et al., 2006; Groves et al., 2006).

As predicted by Risch and Merikangas (1997), GWAS have allowed identification of genetic risk factors with relatively modest effects while preserving the utility of the genome-wide context. The first of the GWAS in diabetes were published in 2007, and to date, some 10 such studies have been published for type 2 diabetes (Diabetes Genetics Initiative, 2007; Florez et al., 2007; Hanson et al., 2007; Hayes et al., 2007; Rampersaud et al., 2007; Salonen et al., 2007; Scott et al., 2007; Sladek et al., 2007; Steinthorsdottir et al., 2007; Zeggini et al., 2007), while three GWAS (Hakonarson et al., 2007; Smyth et al., 2007; WTCCC, 2007) and a follow-up study (Todd et al., 2007) have been published for type 1 diabetes.

These studies have enabled the identification of new genetic risk factors for both disorders and have provided valuable additional context for understanding the relative importance of some of the previously identified risk factors. For example, *TCF7L2*, identified originally through positional cloning studies in an Icelandic sample (Grant et al., 2006), is clearly the largest individual genetic risk factor for type 2 diabetes, at least in type 2 diabetes cases from northern Europe. Similarly, HLA remains the undisputed major genetic risk factor for type 1 diabetes (WTCCC, 2007); as with the linkage studies, no other association comes close to the magnitude of effect detected at HLA. For both disorders, outside the locus with the largest individual effect, the remaining genetic risk factors have quite modest odds ratios (OR), with a few for type 1 diabetes being near 2.0, but most in the range of 1.1–1.2. Moreover, there are clearly many additional genetic risk factors to be identified, as the established genetic risk factors account for only a small proportion (2.3%) of the inter-individual variation in genetic risk for type 2 diabetes and just over half of the variation in genetic risk for type 1 diabetes (due mostly to the large effects in the HLA region). Given that the genetic risk factors yet to be identified are unlikely to have much larger effects than those already identified, but rather are likely to be of similar magnitude or smaller, it is clear that more studies, larger studies, or more sophisticated or novel approaches will be needed to identify additional loci and allow us to achieve a more comprehensive understanding of the genetic risk factors for diabetes.

Indeed, it could be argued that the loci identified to date have added little to our fundamental understanding of the genetic component to either type 1 or type 2 diabetes. That is, type 1 diabetes was already known to be an autoimmune disease, and genetic risk factors identified to date merely confirm the importance of genes involved in T-lymphocyte development, regulation of the immune system and tolerance. Similarly, genetic risk factors identified to date for type 2 diabetes have provided no more insight than we have ever had toward a comprehensive understanding of its fundamental biological basis. Indeed, the more pessimistic viewer might reasonably argue that the modest effect sizes observed to date are the "big" ones – that whatever additional genetic risk factors are identified will have such modest effects as to be nearly indistinguishable from background loci and thus of limited scientific or clinical value.

So, is the glass already half full and filling fast with diabetes susceptibility genes that will provide us that elusive and comprehensive understanding of the genetic component to diabetes? Or is the glass, at best, half empty, with little likelihood of filling no matter how much money we spend at the tap? As with almost all of the earlier genetic research in complex disorders, including candidate gene and linkage mapping studies, the GWAS in diabetes have been among the first, and are certainly among the most numerous with large sample sizes, to be conducted on complex traits; thus, the answers to these questions have implications not only for type 1 and type 2 diabetes, but for other complex disorders as well. In order to address these questions, it is important to examine critically the studies that have been conducted to date and to review what we have learned from them.

GWAS IN TYPE 2 DIABETES

The GWAS for type 2 diabetes have been conducted in a variety of populations, using a variety of study designs and high-throughput platforms. Studies on type 2 diabetes using the first generation GWAS platforms (e.g., the Affymetrix 100K single nucleotide polymorphism [SNP] platform) were quite expensive and consequently included relatively modest numbers of samples. The low sample sizes coupled with the low density and high redundancy of the 100K platform precluded any of the studies on the 100K platform from identifying loci with genome-wide significant associations (Florez et al., 2007; Hanson et al., 2007; Hayes et al., 2007; Rampersaud et al., 2007).

TABLE 52.1 Samples, study designs and SNP genotyping platforms used in dense GWAS for type 2 diabetes published in 2007

Study	Samples	Design	Platform	Follow-up
Sladek et al. (2007)	French 690 cases 660 controls	Case/control Lean cases	Illumina 300K + 100K	57 SNPs 2617 cases 2894 controls
Steinthorsdottir et al. (2007)	Icelandic 1399 cases 5275 controls	Case/control	Illumina 300K	47 SNPs 1110 cases 2272 controls
Scott et al. (2007)	Finns 1161 cases 1174 controls	Case/control	Illumina 300K	80 SNPs 1215 cases 1258 controls
Diabetes Genetics Initiative (2007)	Finns 1464 cases 1467 controls	Case/control and family BMI matched	Affymetrix 500K	107 SNPs 5065 cases 5785 controls
WTCCC (2007)	English 1924 cases 2934 controls	Case/control	Affymetrix 500K	21 + 56 SNPs 2022 cases 2037 controls
Salonen et al. (2007)	European 500 cases 497 controls	Case/control	Illumina 300K	10 SNPs 2573 cases 2776 controls

This situation changed dramatically with the availability of higher density SNP platforms, and Table 52.1 summarizes population, sample size and study design information for the studies published in 2007 and conducted using the denser SNP genotyping platforms on type 2 diabetes. The scans conducted using sufficiently dense platforms to be considered truly genome-wide have been focused exclusively on populations of recent European descent, primarily northern European populations, and it is only in these dense scans on very large samples that loci meeting genome-wide criteria for significance have been identified and confirmed (Table 52.2). Some studies utilized only lean cases (Sladek et al., 2007), matched cases to controls on the basis of body mass index (BMI) (Diabetes Genetics Initiative, 2007), or conducted analyses separately in lean and obese cases (Steinthorsdottir et al., 2007). Many of the studies enhanced power by utilizing cases with a family history of disease in their primary screen, their follow-up studies, or both.

The varying designs have led to intriguing differences in results that help to extend the observed associations. For example, although most studies have reported the OR for *TCF7L2* to be in the range of 1.35–1.4 even within studies in which cases were chosen to have a positive family history of type 2 diabetes, Sladek et al. (2007) estimated the OR for *TCFL2* to be 1.65. This higher estimate may be attributable to the use of lean cases, as the allele at *TCF7L2* increasing risk of type 2 diabetes is associated with lower BMI in both cases and controls; thus lean cases are likely to be particularly enriched for individuals carrying the high-risk allele for type 2 diabetes at *TCF7L2*. Similarly, while the *FTO* locus was easily detected in the WTCCC study and was also apparent in the FUSION study, it had little signal in the DGI study, largely because the DGI study matched cases and controls on BMI. As *FTO* appears to affect risk of type 2 diabetes primarily through its effects on obesity and BMI, the failure to detect the effects of *FTO* in the DGI sample is easily understood; conditioning on BMI similarly reduced the *FTO* effects in the WTCCC samples (Frayling et al., 2007).

The confirmed loci for type 2 diabetes listed in Table 52.2 have little in common with respect to known function, although the risk allele at several of the susceptibility loci (*TCF7L2*, *CDKAL1*, *SLC30A8*) has been shown to be associated with reduced insulin secretion. Clearly, however, considerable additional research into the physiological basis of the risk imparted by the genetic variation at these loci is necessary. Moreover, it is clear we have barely begun to identify genetic risk factors for type 2 diabetes. The WTCCC group noted that the combined effects of all of the risk factors identified in the combined samples of the DGI, FUSION and WTCCC studies together would yield an estimate of λ_s of 1.07 (WTCCC, 2007), which contrasts substantially with the overall λ_s value for type 2 diabetes estimated to be in the range of 1.8–2.8. Investigators from the DGI study noted that the eight loci identified in the combined DGI, FUSION and WTCCC studies account for only about 2.3% of the overall variance in type 2 diabetes risk among individuals (Diabetes Genetics Initiative, 2007). In addition, OR are modest overall and can be quite low within any individual study. As indicated in Table 52.2, the maximum OR for most loci is less than 1.2 and most of these loci have sufficiently low OR in at least one study that the locus did not meet the study's threshold for reporting the result.

TABLE 52.2 Confirmed susceptibility loci/regions for type 2 diabetes

Locus	Chromosome	Odds ratio (range across studies)	Risk allele frequency (controls)
TCF7L2	10	1.34–1.71	0.30
SLC30A8	8	1.01[a]–1.18	0.70
CDKAL1	6	1.16[a]–1.20	0.32
HHEX	10	1.15[a]–1.13	0.58
PPARG	3	1.02[a]–1.20	0.86
CDKN2A/B	9	1.12[a]–1.17	0.83
KCNJ11	11	1.14[a]–1.15	0.45
IGF2BP2	3	1.12[a]–1.18	0.32
FTO	16	1.03[a]–1.23	0.40

[a]Lowest value found, but the lowest OR could not be determined because some studies did not report all results, and the locus indicated did not have OR greater than the threshold for reporting in at least one study.

As has been found in GWAS for other complex traits, most of the alleles associated with increased risk of type 2 diabetes are located in non-coding sequence without obvious function (sometimes between genes, sometimes within introns), a notable exception being an amino acid polymorphism in *SLC30A8*. Even when the associated allele is within an intron of a gene, however, we should be cautious about ascribing risk to the gene in which the SNP resides, and it is clearly premature to ascribe the function of a SNP between genes to the closest gene based only on physical proximity. It is possible that at least some of the risk alleles affect expression of genes outside the one in which they are located or near and may well exert their effects through a gene or genes on other chromosomes. Thus, it may be prudent to continue to refer to these as susceptibility loci rather than as susceptibility genes.

It is also notable that many of the loci with confirmed associations have very common risk alleles (i.e., the frequency of the risk allele is greater than 0.5). While no one has claimed that the polymorphisms identified to date are the actual causal alleles driving the observed associations, it is unlikely that the true susceptibility alleles will be very different in frequency from those already identified. Even with D' of 1.0 between the tag SNP and the causal variant, differences in allele frequencies of the causal and associated alleles would reduce the r^2 value between them sufficiently to degrade the power to detect the association.

The observation that the risk allele is the more common allele for several of the confirmed associations raises the possibility that it is the ancestral allele (i.e., the allele fixed in other primate species) at the causal site that increases risk of disease, with a more recent mutation at that site (i.e., the derived allele) reducing risk relative to the ancestral state. As noted by Di Rienzo and Hudson (2005), quite a number of the polymorphisms confirmed to affect risk to common disorders have ancestral susceptibility alleles. For any biallelic polymorphism there is always one allele that increases and one allele that decreases risk relative to the other, and thus characterization of a susceptibility locus as a "risk" locus or a "protective" locus is often made in arbitrary ways. Considering risk versus protection in the context of ancestral versus derived alleles can substantially improve our understanding of associations. Patterns of variation in the vicinity of susceptibility loci with ancestral susceptibility alleles can be quite different from expectations shaped by the more familiar Mendelian paradigm in which alleles increasing risk of disease are the rare, recent mutations. Accordingly, geneticists may need to revamp their thinking to correctly interpret results of association studies involving such loci.

While it is undeniably exciting to have identified so many susceptibility loci for type 2 diabetes, what we have is still far from the comprehensive understanding of the genetic component to type 2 diabetes that we seek. Given the sample sizes (on the order of 40,000 cases and 40,000 controls for all dense-platform studies combined) and total resources necessary to make the discoveries summarized in Table 52.2, it is daunting to consider what it might take to identify additional susceptibility loci.

FUTURE RESEARCH IN TYPE 2 DIABETES GENETICS

There would clearly be value in combining results across all of the dense GWAS carried out to date. Fortunately, the long-standing collegiality of the diabetes genetics community continues to flourish, with the early collaboration of the DGI, FUSION and WTCCC groups providing an outstanding example of the value of collaborative science in the GWAS era, and it is likely that combined results from all available GWAS in type 2 diabetes will be completed in the near future. In addition to those that have been published, dense GWAS have been conducted in the Framingham cohort (>9000 samples), and in smaller Chinese and Korean samples, and are underway in Pima Indian, Mexican American and African American samples, as well as a variety of additional populations of European descent. If combined analyses can be conducted with even the GWAS that will be available in 2008, it is likely that additional susceptibility loci for type 2 diabetes can be identified.

Notably, studies completed by mid-2008 will provide additional context for the susceptibility loci that have been identified largely in northern European populations, as GWAS in many other populations will be available. Among these additional populations are some in which cases are likely to be leaner than those in northern European populations (e.g., Chinese, Korean), as well as some in which cases are likely to be more obese and insulin resistant than those in northern European populations (e.g., Mexican Americans, African Americans, Pima Indians). The similarities and differences in susceptibility loci detected in these populations may provide more insight into the genetic component of diabetes by highlighting different pathways or parts of pathways in different populations. For example, it is unlikely that *TCF7L2* will be the most significantly associated susceptibility locus in the Chinese, Korean, Pima or Mexican Americans populations, because the allele increasing risk at *TCF7L2* has much

lower frequency in Asian and Native American populations. In addition, as summarized in Table 52.2, only very modest numbers of SNPs have been included in follow-up studies conducted in GWAS on type 2 diabetes to date. Deeper follow-up may well lead to the identification of additional susceptibility loci as well.

But it should not be necessary to rely solely on more studies in more samples to identify additional susceptibility loci for type 2 diabetes. It may be possible to use various types of "enrichment" analyses to determine whether the top signals are enriched for SNPs within genes in certain pathways or within genes associated with particular biological processes. Such analyses are challenging for a number of reasons. While our signals come as SNPs, these bioinformatics analyses require gene lists. As noted above, it may be perilous to annotate SNPs to genes without knowing what the causal polymorphisms actually are. Moreover, even if causal SNPs have been identified, those SNPs may lie outside genes and/or may affect the expression of genes outside the gene in which they are located. Additional studies on the association of genetic variation to expression phenotypes in tissues relevant to type 2 diabetes may help to annotate SNPs to genes for downstream bioinformatics studies. Regardless, enrichment analyses should be conducted in a way that explicitly allows for the uneven interrogation of genes that inevitably occurs even with dense SNP genotyping platforms. For example, genes might be weighted for bioinformatics enrichment studies according to the proportion of known variation that is interrogated on the platform used for analysis.

Potential Clinical Utility

Ideally, the accumulation of additional dense GWAS and the thoughtful application of bioinformatics approaches to these data will enable the identification of a sufficient number of type 2 diabetes susceptibility loci to obtain a better understanding of the nature of the genetic component to type 2 diabetes in all populations. What will the clinical utility of this information be? Type 2 diabetes is one of the few disorders in which there are confirmed associations in genes relevant to the mechanism of drug action for some of the drug therapies used to treat diabetes. This certainly raises the possibility that the identification of additional susceptibility loci will lead to discovery of new drug therapies for this disorder. It has also been argued that if we identify enough of the genetic risk factors, we may be able to develop sufficiently accurate predictive models to permit risk reduction (through behavior modification or even drug therapies) in those with the highest risk of disease.

The problem with that idea is that we already have good models for predicting the risk of type 2 diabetes, and it is unclear that the inclusion of genetic risk factors will improve risk prediction any time soon, if at all (Wilson et al., 2007). These models use a variety of inexpensively measured complex traits (lipid levels, BMI, family history) that are likely to share some of both their genetic and their non-genetic determinants with type 2 diabetes. Thus, knowing the *overall* genetic risk is not really necessary to develop accurate predictions of risk for type 2 diabetes. Nonetheless, it is certainly possible that knowing *specific* genetic risk factors that an individual carries may help determine the most appropriate therapy for that individual. Thus, clinical trials of diabetes therapies should be incorporating information on genetic risk factors as quickly as associations become replicated. It is also possible that knowledge of the specific genetic risk factors will allow us to conduct more targeted epidemiological studies to identify non-genetic risk factors more specific than "Western diet and lifestyle" that increase risk of type 2 diabetes. These more specific non-genetic risk factors may be cost-effective targets for interventions that could prevent, or at least delay onset of disease.

GWAS IN TYPE 1 DIABETES

The first GWAS to be published for type 1 diabetes was a survey of 6500 non-synonymous SNPs conducted in 2029 cases and 1755 controls (Smyth et al., 2006). In addition to confirming the previously known association at *PTPN22*, a SNP at the gene *interferon induced with helicase C domain 1*, *IFIH1,* was found to be strongly associated with type 1 diabetes in both the original and the replication samples of 2471 cases, 4593 controls and 2134 parent–child trios. Subsequent GWAS used dense SNP genotyping platforms to assess the entire genome for type 1 diabetes susceptibility loci. Hakonarson et al. (2007) reported results of a GWAS for type 1 diabetes conducted using the Illumina 550K SNP set on 563 cases, 1163 controls and 483 complete trios for primary analysis and 939 families including at least a trio for follow-up studies. Their studies confirmed associations at HLA, *INS* and *PTPN22*, and identified variation at KIAA0350 as contributing risk to type 1 diabetes.

The initial genome-wide screen in the WTCCC (2007) included ~2000 cases with type 1 diabetes and ~3000 controls unphenotyped with respect to type 1 diabetes. These studies also confirmed the previous associations at HLA and *PTPN22*, and provided independent support for recent reports implicating variation at *CD25*. Subsequent studies reported on both follow-up to the findings in the WTCCC studies and follow-up to a study of 13,378 non-synonymous SNPs in 3400 cases and 3300 controls, with the follow-up being conducted in 4000 cases, 5000 controls and 2997 trios (Todd et al., 2007). Based on the initial WTCCC genome-wide studies, 11 SNPs from regions with $p < 1.64 \times 10^{-5}$ and not previously implicated in type 1 diabetes were genotyped in all of the follow-up samples. Four of these regions yielded compelling replication of association: a region on 18p11 in which *PTPN2* was the only gene, a region on 12q24 in which a single non-synonymous SNP in *SH2B3* was sufficient to model the observed association, the 16p13 region containing KIAA0350, and the 12q13 region, in which a SNP near *ERBB3* was the most strongly associated SNP. They also followed up 14 SNPs from the non-synonymous screen, which provided replication for a SNP in the T-lymphocyte co-stimulation gene *CD226* as well as for SNPs in *CAPSL, C20orf168*, *IL7R* and *CFTR*.

The results of genetic studies in type 1 diabetes provide a more cohesive picture of the genetic risk factors for this autoimmune disease. Genes involved in T-cell development, immune regulation and recognition of pancreatic autoantigens all appear to play a role in the development of type 1 diabetes. As with type 2 diabetes, very large samples were required to identify and confirm these

susceptibility loci. OR for HLA approach 7, and for *INS* and *PTPN22* are around 2.0, but are between 1.2 and 1.3 for many of the other loci that have been identified. Thus, the initial screens were only modestly powered to detect effects at these loci.

FUTURE STUDIES IN TYPE 1 DIABETES

The clinical utility of genetic risk factors identified for type 1 diabetes rises considerably if it becomes possible to reduce the likelihood that individuals with high risk of developing type 1 diabetes actually develop the disease. Good predictive models for type 1 diabetes risk remain elusive, however, and we still lack understanding of the non-genetic triggers for type 1 diabetes. Thus, one may also harness genetic risk factors to improve epidemiological studies targeting identification of specific non-genetic risk factors for type 1 diabetes. While gene–environment interaction complicates the identification of genetic risk factors, it can be an ally for clinical studies, as only one element of the interaction need be disrupted to reduce the entire risk due to the interaction. Thus, prevention strategies might also focus on non-genetic risk factors that interact with genetic risk factors.

CLINICAL UTILITY OF GENETIC RESEARCH IN DIABETES

The most immediate clinical utility of the genetic research in diabetes relates to the ability to use genetic information to improve the quality of diagnosis with respect to the particular subtype of diabetes, which in turn can have profound implications for how patients should be treated. The most dramatic examples of altered diagnosis and treatment regimens relate to the subset of individuals who have been diagnosed with type 1 diabetes at a very early age (usually before 6 months of age), and therefore treated with insulin. Individuals diagnosed with diabetes at this early age are more likely to have a rare Mendelian form of diabetes called permanent neonatal diabetes (PND). The most common cause for PND is mutation at *KCNJ11* within Kir6.2, the inwardly rectifying ATP-sensitive potassium channel. These patients can be safely and more effectively treated with sulfonylureas (taken orally) than with insulin (Pearson et al., 2006), and this change in treatment regimen dramatically alters the lifestyle of patients for the better.

As with the rare monogenic forms of diabetes, we expect that better understanding of the genetic risk factors for diabetes will ultimately lead to new drugs targeting the new genes and pathways that are identified, ultimately leading to more individualized therapies based on primary etiology of disease. Pharmacogenetic studies are providing insights into other ways of individualizing therapies. For example, recent studies (Shu et al., 2006) indicate that genetic variation at *OCT1* may contribute to interindividual variation in response to metformin, a commonly used agent for treatment of diabetes. *OCT1* (organic cation transporter 1) has a role in hepatic uptake of metformin and thus was a candidate gene for generating such effects. The coming years are likely to see more direct investigation of pharmacogenetic effects for drugs used to treat diabetes, and translation of these findings will also personalize diabetes treatment.

2009 UPDATE

The past year has seen continuing advances in efforts to identify and better understand the genes and variants that contribute to the development of diabetes.

Genetics of Type I Diabetes

On the basis of linkage studies and candidate gene studies, at least two dozen genes have now been implicated in type I diabetes (Concannon et al., 2009). Overwhelming evidence supports a clear role for genes in the HLA region on chromosome 6 (notably *HLA-DR* and *–DQ*, but also the gene for insulin, *INS*), with only modest effects on individual risk reported for genes elsewhere in the genome. While the odds ratio reported for HLA is nearly 7 and variants at *INS* and *PNPN22* have odds ratios near 2, the rest of the variants implicated in type I diabetes have reported only modest odds ratios in the range of 1.1–1.3. Nonetheless, most of the implicated genes have reported functions in immunity, suggesting potential pathways for further functional analysis (Concannon et al., 2009).

A large genome-wide association study involving over 7500 cases and 9000 controls raised the estimate to some 41 genetic loci with evidence for association (Barrett et al., 2009). In addition to confirming known susceptibility loci, 18 regions were newly replicated in a further study of 4200 cases and 4500 controls. Among the new candidate genes implicated by this study were *IL10, IL19, LGIS3, CD69* and *IL27*. As in previous studies, odds ratios were reported to be modest.

In a landmark study for identifying causal variants, Todd and colleagues took a candidate gene resequencing approach to search for sequence variants that are rare in the general population, but that reside close to SNP variants implicated by association studies (Nejentsev et al., 2009). They resequenced exons and splice sites from 10 candidate genes in DNA from nearly 500 patients with type 1 diabetes. They reported four rare variants in the *IFIH1* gene, a gene located in a genomic region previously implicated by genome-wide association studies.

Genetics of Type 2 Diabetes

As compiled by Florez (2008), at least 17 genetic loci have now been implicated in type 2 diabetes on the basis of genome-wide association studies involving more than 10,000 individuals. While the associated SNPs are relatively common in those of European ancestry, the reported contributions to risk are, as in type 1 diabetes, quite modest, in the range of ~1.1–1.2 (Florez, 2008; Zeggini et al., 2008). The one confirmed exception to this is variation in the *TCF7L2* gene, with a reported odds ratio of ~1.5. Several other useful reviews of genome-wide association studies have also appeared in the last year (Prokopenko et al., 2008; Ridderstrale and Groop, 2009).

Several large cohort studies have explored the ability of such variants to predict diabetes in a clinical setting. To examine the combined effect of multiple variants, Cornelis et al. (2009) carried out a nested case-control study of over 2800 patients and 3500 healthy controls, all of European ancestry, studying a total of 10 polymorphisms in 9 loci known to be associated with type 2 diabetes. The combined "genetic risk score" resulted in odds ratios ranging from ~0.5 to ~2.0 for different groups of patients, but such scores improved individual risk prediction only slightly beyond conventional risk factors such as BMI, family history, smoking and physical (in)activity. Nonetheless, for the 20% of the population with the highest genetic risk score, the authors conclude this score may be useful for identifying those at highest risk (Cornelis et al., 2009).

In addition, Lyssenko et al. (2008) studied 16 SNPs associated with type 2 diabetes risk, together with clinical factors, in a prospective cohort of over 16,000 Swedish and 2779 Finnish subjects. Over 2200 subjects developed diabetes over a follow-up period of about two decades. Variants in 11 genes were associated with risk of progression to diabetes, with 8 of the genes known to be involved in beta-cell function. Again, however, the addition of specific SNP information only modestly improved the prediction of future diabetes over conventional clinical risk factors (Lyssenko et al., 2008).

All of these variants are likely to contribute to our understanding of the biological underpinnings of this complex disease and may even help to stratify populations into different subtypes beyond simply "type 1" and "type 2". However, it still remains unclear how these variants will contribute to the clinical evaluation and treatment of patients who are at risk. The clinical utility of genetic variation that predisposes to (or protects from) diabetes remains to be fully established. (*Update prepared by editors*)

2009 UPDATE REFERENCES

Barrett, J.C., Clayton, D., Concannon, P., Akolkar, B., Cooper, J.D., Erlich, H., Julier, C., Morahan, G., Nerup, J., Nierras, C. et al. (2009). Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat genet* 41, 703–707.

Concannon, P., Rich, S.S. and Neporn, G.T. (2009). Genetics of type IA diabetes. *N Engl J Med* 360, 1646–1654.

Cornelis, M., Qi, L., Zhang, C., Kraft, P., Manson, J., Cai, T., Hunter, D. and Hu, F.B. (2009). Joint effects of common genetic variants on the risk for type 2 diabetes in U.S. men and women of European ancestry. *Ann Internal Med* 150, 541–550.

Florez, J.C. (2008). The genetics of type 2 diabetes: A realistic appraisal in 2008. *J Clin Endocrinol Metab* 93, 4633–4642.

Lyssenko, V., Jonsson, A., Almgren, P., Pulizzi, N., Isomaa, B., Tuomi, T., Berglund, G., Alshuler, D., Nilsson, P. and Groop, L. (2008). Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N Engl J Med* 359, 2220–2232.

Nejentsev, S., Walker, N., Riches, D., Egholm, M. and Todd, J.A. (2009). Rare variants of *IFIH1*, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* 324, 387–389.

Prokopenko, I., McCarthy, M. and Lindgren, C.M. (2008). Type 2 diabetes: New genes, new understanding. *Trends Genet* 24, 613–621.

Ridderstrale, M. and Groop, L. (2009). Genetic dissection of type 2 diabetes. *Mol Cell Endocrin* 297, 10–17.

Zeggini, E., Scott, L., Saxena, R., Voight, B., Marchini, J., Hu, T., de Bakker, P., Abecasis, G., Almgren, P., Andersen, G. et al. (2008). Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nature genet* 40, 638–645.

REFERENCES

Altshuler, D., Hirschhorn, J.N., Klannemark, M., Lindgren, C.M., Vohl, M.-C., Nemesh, J., Lane, C.R., Schaffner, S.F., Bolk, S., Brewer, C. et al. (2000). The common PPARg Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet* 26, 76–80.

Bell, G.I., Horita, S. and Karam, J.H. (1984). A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. *Diabetes* 33(2), 176–183.

Bottini, N., Musumeci, L., Alonso, A., Rahmouni, S., Nika, K., Rostamkhani, M., MacMurray, J., Meloni, G.F., Lucarelli, P., Pellecchia, M. et al. (2004). A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. *Nat Genet* 36, 337–338.

Cauchi, S., Meyre, D., Dina, C., Choquet, H., Samson, C., Gallina, S., Balkau, B., Charpentier, G., Pattou, F., Stetsyuk, V. et al. (2006). Transcription factor TCF7L2 genetic study in the French population: Expression

in human beta-cells and adipose tissue and strong association with type 2 diabetes. *Diabetes* 55(10), 2903–2908.

Davies, J.L., Kawaguchi, Y., Bennett, S.T., Copeman, J.B., Cordell, H.J., Pritchard, L.E., Reed, P.W., Gough, S.C.L., Jenkins, S.C., Palmer, S.M. et al. (1994). A genome-wide search for human type 1 diabetes susceptibility genes. *Nature* 371(8), 130–136.

Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research, Saxena, R., Voight, B.F., Lyssenko, V., Burtt, N.P., de Bakker, P.I., Chen, H., Roix, J.J. et al. (2007). Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 316(5829), 1331–1336.

Di Rienzo, A. and Hudson, R.R. (2005). An evolutionary framework for common diseases: The ancestral-susceptibility model. *Trends Genet* 21(11), 596–601.

Florez, J.C., Hirschhorn, J. and Altshuler, D. (2003). The inherited basis of diabetes mellitus: Implications for the genetic analysis of complex traits. *Ann Rev Genomics Hum Genet* 4, 257–291.

Florez, J.C., Manning, A.K., Dupuis, J., McAteer, J., Irenze, K., Gianniny, L., Mirel, D.B., Fox, C.S., Cupples, L.A. and Meigs, J.B. (2007). A 100K genome-wide association scan for diabetes and related traits in the Framingham Heart Study: Replication and integration with other genome-wide datasets. *Diabetes* 56(12), 3063–3074.

Frayling, T.M., Timpson, N.J., Weedon, M.N., Zeggini, E., Freathy, R.M., Lindgren, C.M., Perry, J.R., Elliott, K.S., Lango, H., Rayner, N.W. et al. (2007). A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 316(5826), 889–894.

Gloyn, A.L., Weedon, M.N., Owen, K.R., Turner, M.J., Knight, B.A., Hitman, G., Walker, M., Levy, J.C., Sampson, M., Halford, S. et al. (2003). Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. *Diabetes* 52(2), 568–572.

Grant, S.F., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Manolescu, A., Sainz, J., Helgason, A., Stefansson, H., Emilsson, V., Helgadottir, A. et al. (2006). Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet* 38(3), 320–323.

Groves, C.J., Zeggini, E., Minton, J., Frayling, T.M., Weedon, M.N., Rayner, N.W., Hitman, G.A., Walker, M., Wiltshire, S., Hattersley, A.T. et al. (2006). Association analysis of 6,736 U.K. subjects provides replication and confirms TCF7L2 as a type 2 diabetes susceptibility gene with a substantial effect on individual risk. *Diabetes* 55(9), 2640–2644.

Hakonarson, H., Grant, S.F., Bradfield, J.P., Marchand, L., Kim, C.E., Glessner, J.T., Grabs, R., Casalunovo, T., Taback, S.P., Frackelton, E.C. et al. (2007). A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature* 448(7153), 591–594.

Hayes, M.G., Pluzhnikov, A., Miyake, K., Sun, Y., Ng, M.C., Roe, C.A., Below, J.E., Nicolae, R.I., Konkashbaev, A., Bell, G.I. et al. (2007). Identification of type 2 diabetes genes in Mexican Americans through genome-wide association studies. *Diabetes* 56(12), 3033–3044.

Hanson, R.L., Bogardus, C., Duggan, D., Kobes, S., Knowlton, M., Infante, A.M., Marovich, L., Benitez, D., Baier, L.J. and Knowler, W.C. (2007). A search for variants associated with young-onset type 2 diabetes in American Indians in a 100K genotyping array. *Diabetes* 56(12), 3045–3052.

Love-Gregory, L.D., Wasson, J., Ma, J., Jin, C.H., Glaser, B., Suarez, B.K. and Permutt, M.A. (2004). A common polymorphism in the upstream promoter region of the hepatocyte nuclear factor-4 alpha gene on chromosome 20q is associated with type 2 diabetes and appears to contribute to the evidence for linkage in an ashkenazi jewish population. *Diabetes* 53(4), 1134–1140.

Narayan, K.M., Boyle, J.P., Thompson, T.J., Sorensen, S.W. and Williamson, D.F. (2003). Lifetime risk for diabetes mellitus in the United States. *JAMA* 290, 1884–1890.

Pearson, E.R., Flechtner, I., Njølstad, P.R., Malecki, M.T., Flanagan, S.E., Larkin, B., Ashcroft, F.M., Klimes, I., Codner, E., Iotova, V. et al. (2006). Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations. *N Engl J Med* 355(5), 467–477.

Permutt, M.A., Wasson, J. and Cox, N. (2005). Genetic epidemiology of diabetes. *J Clin Invest* 115(6), 1431–1439.

Rampersaud, E., Damcott, C.M., Fu, M., Shen, H., McArdle, P., Shi, X., Shelton, J., Yin, J., Chang, Y.P., Ott, S.H. et al. (2007). Identification of novel candidate genes for type 2 diabetes from a genome-wide association scan in the Old Order Amish: Evidence for replication from diabetes-related quantitative traits and from independent populations. *Diabetes* 56(12), 3053–3062.

Rich, S.S. (1990). Mapping genes in diabetes. Genetic epidemiological perspective. *Diabetes* 39(11), 1315–1319.

Salonen, J.T., Uimari, P., Aalto, J.M., Pirskanen, M., Kaikkonen, J., Todorova, B., Hyppönen, J., Korhonen, V.P., Asikainen, J., Devine, C. et al. (2007). Type 2 diabetes whole-genome association study in four populations: The DiaGen consortium. *Am J Hum Genet* 81(2), 338–345.

Scott, L.J., Mohlke, K.L., Bonnycastle, L.L., Willer, C.J., Li, Y., Duren, W.L., Erdos, M.R., Stringham, H.M., Chines, P.S., Jackson, A.U. et al. (2007). A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316(5829), 1341–1345.

Shu, Y., Sheardown, S.A., Brown, C., Owen, R.P., Zhang, S., Castro, R.A., Ianculescu, A.G., Yue, L., Lo, J.C., Burchard, E.G. et al. (2006). Effect of genetic variation in the organic cation transporter 1 (OCT1) on metformin action. *J Clin Invest* 117(5), 1422–1431.

Silander, K., Mohlke, K.L., Scott, L.J., Peck, E.C., Hollstein, P., Skol, A.D., Jackson, A.U., Deloukas, P., Hunt, S., Stavrides, G. et al. (2004). Genetic variation near the hepatocyte nuclear factor-4 alpha gene predicts susceptibility to type 2 diabetes. *Diabetes* 53(4), 1141–1149.

Sladek, R., Rocheleau, G., Rung, J., Dina, C., Shen, L., Serre, D., Boutin, P., Vincent, D., Belisle, A., Hadjadj, S. et al. (2007). A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445(7130), 881–885.

Smyth, D.J., Cooper, J.D., Bailey, R., Field, S., Burren, O., Smink, L.J., Guja, C., Ionescu-Tirgoviste, C., Widmer, B., Dunger, D.B. et al. (2006). A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. *Nat Genet* 38, 617–619.

Steinthorsdottir, V., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Jonsdottir, T., Walters, G.B., Styrkarsdottir, U., Gretarsdottir, S., Emilsson, V., Ghosh, S. et al. (2007). A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet* 39(6), 770–775.

Todd, J.A., Walker, N.M., Cooper, J.D., Smyth, D.J., Downes, K., Plagnol, V., Bailey, R., Nejentsev, S., Field, S.F., Payne, F., et al. (2007) Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet* 39(7), 857–864.

Ueda, H., Howson, J.M., Esposito, L., Heward, J., Snook, H., Chamberlain, G., Rainbow, D.B., Hunter, K.M., Smith, A.N., Di Genova, G. et al. (2003). Association of the T-cell regulatory

gene CTLA4 with susceptibility to autoimmune disease. *Nature* 423(6939), 506–511.

The Welcome Trust Case Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661–678.

Wilson, P.W., Meigs, J.B., Sullivan, L., Fox, C.S., Nathan, D.M. and D'Agostino, R.B., Sr. (2007). Prediction of incident diabetes mellitus in middle-aged adults: The Framingham Offspring Study. *Arch Intern Med* 167(10), 1068–1074.

Zeggini, E., Weedon, M.N., Lindgren, C.M., Frayling, T.M., Elliott, K.S., Lango, H., Timpson, N.J., Perry, J.R., Rayner, N.W., Freathy, R.M. et al. (2007). Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 316, 1336–1341.

Clinical (Neurological Disease)

Section 8

53. Genetics and Genomics of Dementia
54. Genetics and Genomics of Parkinson's Disease
55. Genomic Considerations in Ophthalmology
56. Genomics in the Diagnosis and Management of Depression

This page intentionally left blank

CHAPTER

Genetics and Genomics of Dementia

Robert L. Nussbaum

INTRODUCTION

Dementia is a serious and growing problem as the average age of the population increases. As one example, the over-65 population of Canada has more than doubled over the past 70 years, from 5% to 12% of the population, while the fraction of the population over age 85 has quadrupled, from 0.2% to 0.8% (Hogan and Hogan, 2003); a similarly aging population is found throughout the developed world. The number of dementia patients is rising proportionately to the aging of the population, placing a tremendous economic and emotional burden on their families and society. Dementia is a serious problem in younger individuals as well, with an estimated half-million individuals under age 65 afflicted in the United States alone (Maslow, 2006). There is a clear need to understand the causes of dementia, to determine their pathogenetic pathways, and to develop methods to intervene to prevent or delay the onset.

Dementia is a clinical syndrome, defined as a progressive impairment of cognitive function, particularly in the areas of memory, judgment, decision-making, attentiveness when communicating with others, orientation to familiar surroundings, and language (Dugue et al., 2003). Many different diseases with different pathological processes cause dementia. In 30% of cases, the dementia is secondary to systemic disease such as hypertension or atherosclerosis, leading to vascular-occlusive dementia with multiple small infarcts (Barker et al., 2002). Other secondary causes include paraneoplastic syndromes, alcoholism, and AIDS (Dugue et al., 2003). The majority of cases, however, reflect primary dementia; among these, the different forms can be roughly divided into two general categories based on pathological appearance of the postmortem brain: those with neurodegeneration and aggregates of protein, and those with vascular-occlusive disease and infarction. Many elderly individuals who die with dementia, however, have both vascular pathology and protein aggregation findings at autopsy.

More than half of all primary dementia is due to Alzheimer disease (AD). Another 10% of dementia is due to diffuse Lewy body disease (DLBD), which presents as a dementia clinically similar to AD but is associated with the characteristic intracellular protein aggregates referred to as Lewy bodies (Lippa et al., 2007). Finally, there are a number of early-onset, autosomal dominant hereditary dementias, ranging from the relatively common, such as frontotemporal dementia (FTD), to the rare, such as CADASIL and the prion diseases (Figure 53.1).

Minimal cognitive impairment (MCI) is a term used to describe individuals with mild difficulty with one component of cognition, such as short-term memory ("amnestic MCI") (Dugue et al., 2003). Between 10% and 15% of amnestic MCI patients progress each year to true dementia, usually of the AD-type, which is five- to seven-times greater than the rate at which dementia occurs in a group of age-matched elderly without MCI (Petersen et al., 1995).

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

In this chapter, we focus on the various forms of primary dementia. We will describe the clinical features, incidence, and what is known about the underlying genetic bases for these diseases. The underlying pathophysiology leading to neuronal cell death is not completely understood for any of the primary dementias, but protein aggregation in or around neurons is a striking feature that is in common among nearly all the primary dementias (CADASIL being the exception, where protein aggregates occur in blood vessels, leading to micro-infarcts).

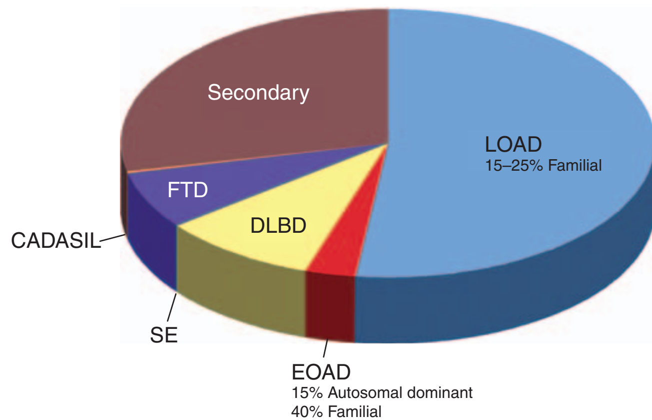


Figure 53.1 Relative contributions of various causes of dementia to the overall population burden of disease. LOAD: late-onset Alzheimer disease; EOAD: early-onset Alzheimer disease; FTD: frontotemporal dementia; DLBD: diffuse Lewy body disease; SE: spongy encephalopathy (prion disease); CADASIL: cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy. Figures are approximate.

INCIDENCE OF DEMENTIA

Over 95–98% of all dementias occur after age 65, an age that is often cited as the arbitrary dividing line between early- and late-onset diseases. Early-onset dementias (i.e., those that occur before the age of 65) are rare and have an incidence of approximately 2–5 per 1000 person-years in the age group 55–64 (Knopman et al., 2006; Maslow, 2006), whereas the incidence of dementia over the age of 65 is approximately 20 per 1000 person-years. The incidence of dementia increases steadily with increasing age, from nearly 5 per 1000 in the 65–69-year-old group to over 84 per 1000 person-years in individuals over 90 (Kukull and Ganguli, 2000; Kukull et al., 2002) (Figure 53.2). Over age 65, well over 95% of dementia is primary, caused by one of the neurodegenerations (principally AD) described in this chapter, or is secondary to vascular-occlusive disease of the brain (McMurtray et al., 2006). In contrast, below age 65, only 70% of dementia is either neurodegenerative or vascular in origin.

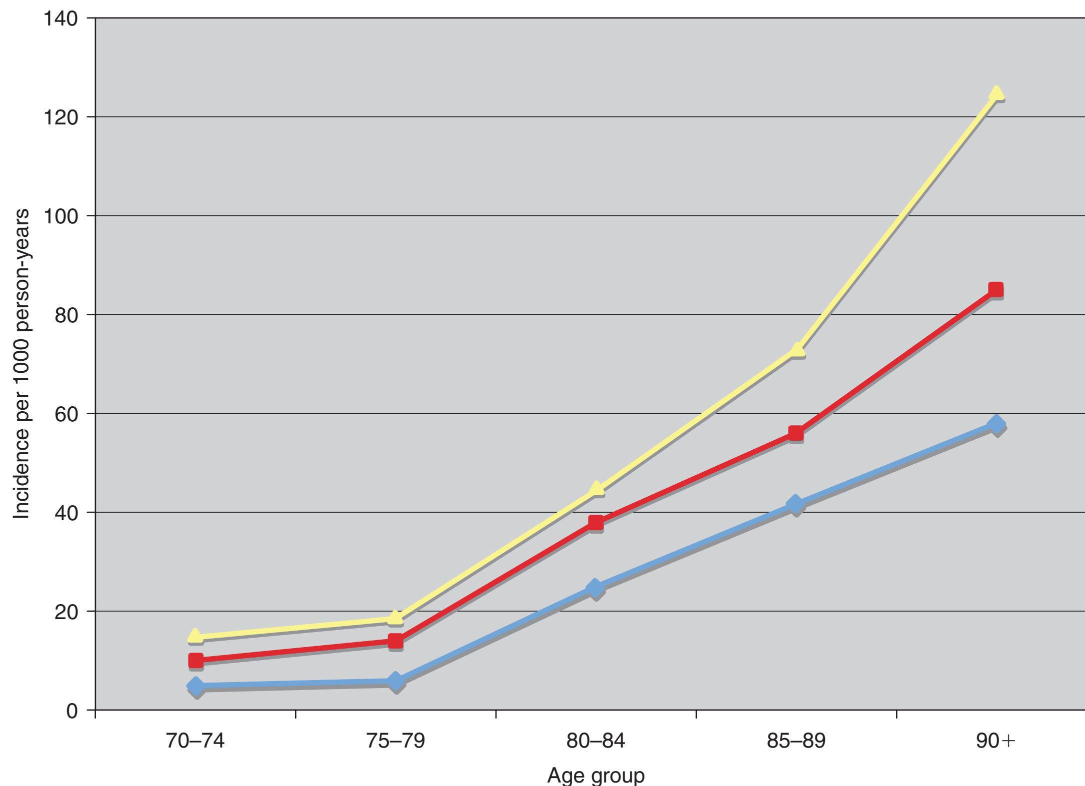


Figure 53.2 Incidence per 1000 person-years of dementia of all types as a function of age (red line). Yellow and blue lines are the 95% confidence limits. Data from Dugue et al. (2003).

PRIMARY DEMENTIAS

Alzheimer Disease

AD is a disorder of slowly progressive dementia associated with extracellular aggregates of beta-amyloid ("amyloid plaques") and intracellular aggregates of the cytoskeletal protein tau ("neurofibrillary tangles") (Figure 53.3). At autopsy, AD patients have plaques and tangles throughout the cortex, but most particularly in the entorhinal cortex and hippocampal CA1 layer. Approximately 5–7% of all AD has an early age of onset; among these early-onset cases, approximately one-sixth are inherited in an autosomal dominant manner while another 40% show some familial recurrence without a clear mendelian inheritance pattern (Ertekin-Taner, 2007; Rao et al., 1994). The majority of late-onset AD (LOAD) is sporadic (i.e., a single individual is affected in the family) although 15–25% do have other affected individuals in their families. Even when there are multiple AD patients in a family, however, the disease usually lacks a clear-cut mendelian inheritance pattern (Nussbaum et al., 2007).

Early-Onset AD

The majority of all of the autosomal dominant, early-onset forms of AD are caused by mutations in one of three genes: the amyloid precursor protein gene (*APP*) located on human chromosome 21, the presenilin 1 gene (*PSEN1*), and the presenilin 2 gene (*PSEN2*) (Rogaeva et al., 2006). Mutations in *APP* and *PSEN1* cause a highly penetrant, predominantly early-onset disease. In contrast, *PSEN2* mutations have more variable penetrance and an older age of onset (Rogaeva et al., 2006).

APP, *PSEN1*, and *PSEN2* mutations cause AD through their effect on the processing of the protein encoded by *APP*, the amyloid precursor protein. This protein is a single-pass transmembrane protein located at the cell surface. It can be cleaved by one of two pathways: either via an a-secretase or a b-secretase (Figure 53.4) (Nussbaum and Ellis, 2003). Further cleavage by c-secretase following b-secretase cleavage creates peptide fragments containing 40 or 42 amino acid residues, referred to as the Ab40 and Ab42 peptides. A large body of evidence implicates the Ab42 fragment, a highly amyloidogenic peptide, as a major toxic agent in the pathogenesis of AD (Eckman and Eckman, 2007). First, of the nearly two dozen missense mutations in *APP* that cause autosomal dominant, early-onset AD (EOAD), all result in amino acid substitutions near the a-, b- or c-secretase cleavage sites. Through mechanisms that are not completely understood, these mutations promote cleavage via the b- and c-secretase pathways and generate relatively or absolutely increased amounts of Ab42. Second, copy number mutations of the *APP* gene, ranging in size from a few hundreds of kilobases to the entire chromosome 21 (as in Down syndrome) lead to EOAD and also cause an increased production of the Ab42 peptide (Cabrejo et al., 2006; Lai and Williams, 1989; Rovelet-Lecrux et al., 2006, 2007; Schupf et al., 2007). Finally, elevated production of Ab42 also occurs in association with mutations in the *PSEN1* and *PSEN2* genes in families with autosomal dominant EOAD (Citron et al., 1997). Hundreds of different missense mutations have been found throughout the *PSEN1* gene; in contrast, only a few mutations have been described in the much rarer families with *PSEN2* mutations. How these mutations affect the activity of the various secretases and cause increased Ab42 production is not known.

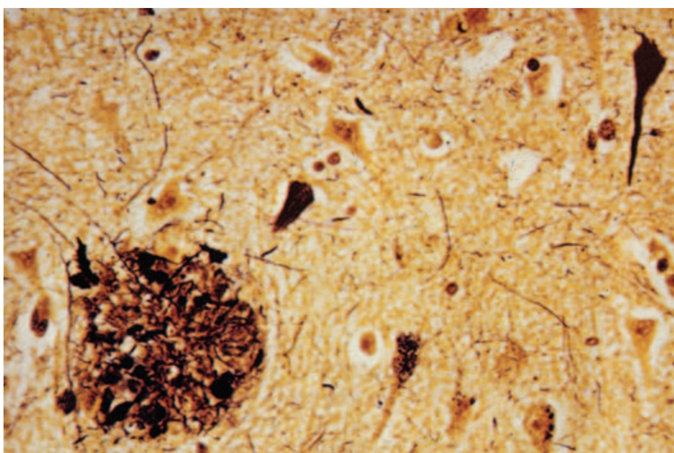


Figure 53.3 Pathology in AD. The large circular plaque in the lower left-hand corner is an extracellular amyloid plaque. The dark triangular shaped bodies are neurons filled with intracellular neurofibrillary tangles of tau protein. (Formalin-fixed brain treated with silver stain.)

Late Age of Onset AD

Genetics also contributes substantially to the risk for late age of onset AD, as documented in twin and family epidemiological studies. For example, in a large Swedish twin registry with long-term follow-up, a typical concordance rate for AD was 59% for monozygotic twins (Gatz et al., 2005). In contrast, concordance for dizygotic twins was typically between 24%, for unlike-sexed, and 32%, for like-sexed twins. Such twin studies suggest that about half of the liability to develop AD is genetic in origin (Pedersen et al., 2004).

Comparing the frequency of AD in the family members of AD patients versus the relatives of unaffected controls is a complementary approach to estimating the genetic contribution to LOAD. Among first-degree relatives of patients affected with AD, the average lifetime risk for AD is 40% up to age 96, with a somewhat higher risk applying to women, as compared to a lifetime risk of 15–20% in individuals without affected relatives (Lautenschlager et al., 1996; Seshadri et al., 1995). Based on family data, it is estimated that about half of the liability to develop late age of onset AD is genetic in origin (Martinez et al., 1998). A similar heritability was noted when MRI scans were used to detect AD endophenotypes in the relatives of AD patients (Lunetta et al., 2007).

Although twin and family studies clearly support a major genetic contribution to LOAD, only one gene variant has so far been incontrovertibly shown to contribute significantly to AD disease risk, the ε4 allele at the apolipoprotein E (*APOE*) gene.

There are three alleles at *APOE*, ε2, ε3, and ε4, and the risk for AD increases substantially and additively with each ε4 allele an individual carries (Corder et al., 1993; Saunders et al., 1993). Among Caucasians, the risk for AD increases approximately threefold in carriers of one ε4 allele and approximately eightfold in ε4/ε4 homozygotes. The impact of the ε4 allele appears to be predominantly through its effect on the average age of onset: each ε4 allele lowers the average age of onset of AD by approximately 10 years.

The effect of the ε4 alleles at the *APOE* locus appears to be the strongest in populations of European and Japanese origin. The data are much more equivocal for African American populations, with some studies suggesting an association between ε4 alleles and AD and others showing no association (Murrell et al., 2006; Tang et al., 1998). The controversies surrounding the association between *APOE* ε4 and AD in different ethnic or racial groups has been recently reviewed (Ertekin-Taner, 2007).

The mechanism by which the *APOE* ε4 allele increases the risk for AD has not been clearly determined. However, it is clear that the apolipoprotein E is synthesized in neurons in response to injury and undergoes proteolytic cleavage (Mahley et al., 2006). The product of the ε4 allele appears to be cytotoxic through an effect on the cytoskeleton and on mitochondrial function. Further work needs to be done to elucidate the role of APOE, and the ε4 allele in particular, in neurodegeneration.

It should be stressed, however, that the *APOE* genotype alone does not explain the entire genetic contribution to LOAD (Martinez et al., 1998). A large number of linkage and association studies have been performed in an attempt to find additional loci at which variants increase the risk for AD. Studies in many populations in the United States and Europe have identified more than 75 different loci with significant linkage (LOD scores) or statistically significant association (Ertekin-Taner, 2007). Unfortunately, some of these studies used overlapping sample sets and cannot be considered independently confirmed. Others have not been replicated, and there is some persistent skepticism as to whether these loci represent real linkage or association findings or false positive signals (Finckh, 2003). For

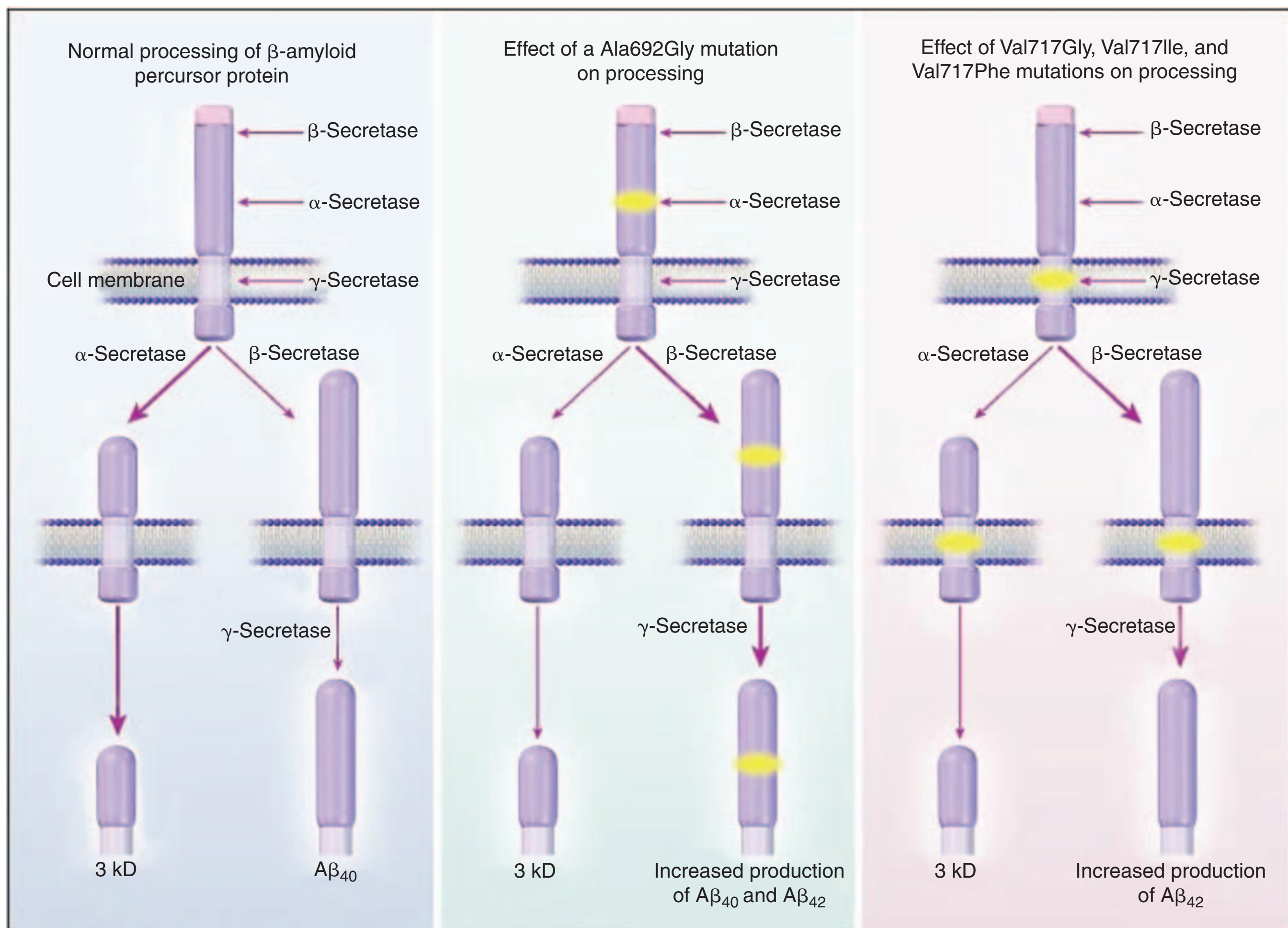


Figure 53.4 Pathways of APP cleavage by the a-, b-, and c-secretases, demonstrating the steps by which the amyloidogenic, toxic Ab42 peptide is generated (Panel A). Panels B and C depict the effect of various missense mutations in the *APP* gene that lead to increased Ab42 production. Reproduced from Nussbaum and Ellis (2003) with permission.

example, one large genome-wide scan using over 500,000 SNP markers in over 1000 autopsy-proven AD patients and controls confirmed that *APOE* itself is contributing to LOAD, but failed to find any other loci in which variants are significantly associated with AD (Coon et al., 2007). Nevertheless, a number of chromosomal regions stand out as possible locations for genes involved in AD risk because they have suggestive LOD scores or demonstrate significant association that appear to have been replicated in multiple, different studies that analyzed independent samples: chromosome 6 (regions 6p21-2 and 6q27), chromosome 10 (region 10q21.3), chromosome 12 (regions 12p11.23 and 12q13.2) and chromosome 9 (regions 9p21.2-p22.2 and 9q22.2) (Ertekin-Taner, 2007). Positional candidate genes within these regions have been analyzed for association to AD, and some genes demonstrate an association with AD that has been replicated in multiple studies. A database of such associations is available at the AlzGene website (Bertram et al., 2007). These genes and regions of the genome are, therefore, attractive candidates for further genetic and functional studies.

Clues from the cell biology of Ab42 production have also been used to identify potential functional candidate genes in which variants might be associated with a risk for AD. Variants in one interesting class of genes, those that encode proteins involved in APP sorting and delivery to the various secretase pathways, have been analyzed for association. The sortilin-related receptor gene *SORL1* has emerged from these studies as a candidate contributor to AD risk (Lee et al., 2007; Rogaeva et al., 2007). In cell culture systems, loss of SORL1 function causes an increase in the amount of APP that is sorted to the b- and c-secretase pathways, thereby increasing the production of Ab42. Variants in the non-coding regions upstream and downstream of the gene constitute two independent haplotypes that show significant association to AD in three independent populations. These same risk haplotypes were associated with a decreased expression of *SORL1* mRNA in white blood cells, thereby providing a functional link to the genetic association.

Despite extensive genetic analysis, much remains to be done to identify the genetic contributors to AD. The field of AD genetics is represented in the current deluge of genome-wide association studies being carried out in many complex diseases (see Chapter 4). One hopes that the pathogenic variants in most of the genes, even those associated with modest increases or decreases in AD risk, will be discovered and confirmed. With these genes in hand, novel therapeutic targets will become evident for the potential development of interventions to combat the disease.

Frontotemporal Dementia

FTD is an early-onset form of dementia that is nearly as common a cause of early-onset dementia as is EOAD (Ratnavalli et al., 2002). The disease may, however, occur even after age 65 and may account for up to 5–10% of dementia in the elderly. The disease is often familial and can be inherited as an autosomal dominant trait with very high, age-dependent penetrance. FTD is clinically heterogeneous, and the dementia seen in FTD differs in some respect from that seen in AD (Haugarvoll et al., 2007). For example, early in the course of FTD, the dementia is often associated with difficulties with social interactions that result from behavioral and personality changes, disinhibition, and neglect of personal hygiene. Disorientation to familiar surroundings and short-term memory loss are less prominent features early in FTD compared to AD. A Parkinsonian movement disorder is very common, as is occasional lower motor neuron dysfunction reminiscent of amyotrophic lateral sclerosis.

FTD results from mutations in four different genes (Haugarvoll et al., 2007). The loci most commonly found mutated in FTD are the progranulin gene (*GRN*) and the gene encoding the microtubular-associated protein tau (*MAPT*), each accounting for an approximately equal number of cases of FTD. Together, they are responsible for just over half of all cases of FTD. At autopsy, FTD brains show loss of large cortical neurons in the prefrontal and anterior temporal lobes. In cases of FTD that are due to mutations in *MAPT*, there are aggregates of the microtubular protein tau associated with gliosis. In other FTD patients, those with mutations in *GRN*, neuronal inclusions contain ubiquitin-positive protein aggregates containing the TAR DNA binding protein-43, seen with microvacuolization of the neuropil and scant gliosis.

The types of mutations causing FTD are very different in the cases due to *GRN* and those due to *MAPT* mutations. The mutations in *GRN* are overwhelmingly loss-of-function mutations, implying that haploinsufficiency is the dominant genetic mechanism (Haugarvoll et al., 2007). The precise biological role of progranulin and the mechanism by which haploinsufficiency of this gene causes highly penetrant FTD are unknown. In contrast, mutations in *MAPT* are toxic gain-of-function mutations (Hutton, 2001). The *MAPT* gene has four exons (exons 9 through 12) that contain copies of a microtubular-binding domain. *MAPT* undergoes extensive alternative splicing that generates at least six isoforms. These six *MAPT* isoforms can be grouped into two classes, those that include exon 10 and have four copies of the repeat and those that skip exon 10 and have only three copies of the repeat (Figure 53.5). Many of the mutations in *MAPT* that cause FTD interfere with the skipping of exon 10 and generate an excess of tau protein containing four microtubular binding domain repeats. Tau with four microtubular binding domains is much more prone to undergo fibrillization and aggregation. There are other *MAPT* mutations that do not affect splicing of exon 10 but alter the properties of the tau protein, rendering it more prone to aggregation and, presumably, lead to FTD that way.

In addition to mutations in *GRN* or *MAPT*, rare missense mutations in the gene for the chromatin-modifying 2B protein (*CHMP2B*) can cause FTD associated with amyotrophic lateral sclerosis; nuclear inclusions are seen at autopsy. Another rare form of FTD results from missense mutations in the valosin-containing protein gene (*VCP*), often in association with an inclusion-body myopathy or Paget disease of bone. The pathogenic mechanism by which missense mutations in *CHMP2B* or *VCP* cause FTD is unknown.

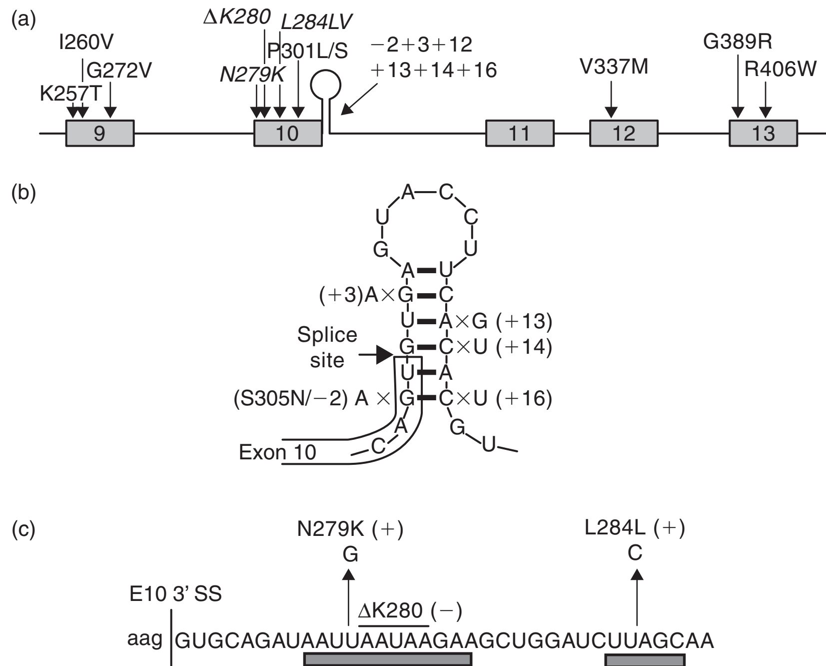


Figure 53.5 *MAPT* mutations in FTD. (a) Schematic diagram of a segment of the gene containing exons 9 through 13. Exonic missense mutations and mutations in intron 10 are shown. (b) Schematic diagram of the stem-loop structure in the 5′ end of intron 10. Mutations that disrupt the stem-loop reduce the inhibitory effect of the stem-loop on splicing that would exclude exon 10, leading to more frequent inclusion of exon 10 and, therefore, increased synthesis of tau protein with four microtubule binding domains. (c) Exonic mutations that strengthen a splice enhancer, lead to more frequent inclusion of exon 10 and, therefore, increased synthesis of tau protein with four microtubule-binding domains. Reproduced from Hutton (2001) with permission.

DLBD and Parkinson Disease with Dementia

DLBD is another form of neurodegeneration, accounting for ~10% of cases of dementia, although some estimates suggest it is nearly as frequent as AD as a cause of dementia in the elderly (Lippa et al., 2007). DLBD is clinically similar to AD but characteristically shows a higher frequency of psychosis with auditory and visual hallucinations and is often accompanied by the symptoms and signs of Parkinsonism (Lippa et al., 2007; McKeith et al., 2005; Walker et al., 2002). At autopsy, brains of DLBD patients have the characteristic a-synuclein aggregates known as Lewy bodies throughout the cerebral cortex as well as in the olfactory bulb/anterior olfactory nucleus and in subcortical regions such as the substantia nigra, locus coeruleus, and dorsal motor nucleus of the vagus (Figure 53.6). The Lewy bodies seen in DLBD are similar to the aggregates found in classic Parkinson disease (PD), although their distribution is more cortical in DLBD, while in PD they are concentrated more in subcortical areas, at least during early stages of the disease (Braak et al., 2003, 2004).

A longstanding controversy exists over the relationship between DLBD and the dementia seen in PD patients (Parkinson disease with dementia or PDD). Dementia develops in 20–40% of patients with PD (Mayeux et al., 1992; Mindham et al., 1993) while clinical or subclinical abnormalities involving the dopaminergic system of the substantia nigra are frequent in patients with DLBD (Walker et al., 2002). An arbitrary distinction used by clinicians to separate PDD from DLBD is the "1-year" rule. If dementia occurs before or up to 1 year following the onset of the Parkinsonian movement disorder, the patient is said to have DLBD. If dementia occurs more than 1 year after the onset of the movement abnormality, the diagnosis of PDD is made instead. Although there are some subtle differences in the dementia seen in PDD and DLBD, both disorders have similar a-synuclein aggregates. Furthermore, the occurrence of PDD and DLBD in different members of the same family carrying a duplication or a triplication of the a-synuclein gene (Chartier-Harlin et al., 2004; Farrer et al., 2004; Singleton et al., 2003) suggest that the underlying pathogenic mechanism is the same in both disorders, even if the clinical features and distribution of the Lewy body pathology is somewhat different.

Mendelian DLBD is rare but does occur in families with autosomal dominant forms of PD. Families with mutations (Kruger et al., 1998; Polymeropoulos et al., 1997; Zarranz et al., 2004) or copy number changes (Chartier-Harlin et al., 2004; Farrer et al., 2004; Singleton et al., 2003) in the a-synuclein gene have early-onset, autosomal dominant disease, presenting clinically as PDD in some family members, but as DLBD in others. Patients with mutations in the *LRRK2* gene, which is responsible

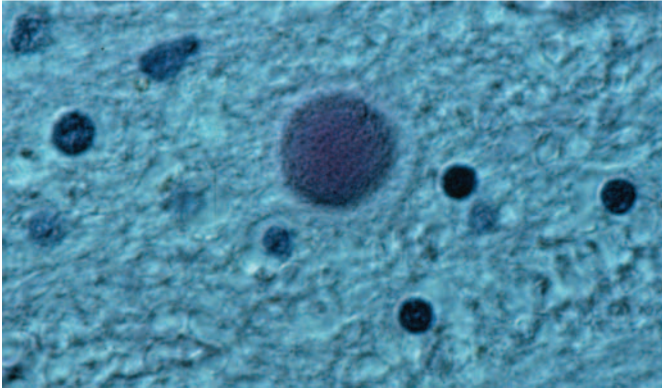


Figure 53.6 Pathology in DLBD. The large circular structure in the center of the field is a Lewy body. (Formalin-fixed human brain section stained with hematoxylin and eosin.)

for a form of autosomal dominant PD with variable penetrance, can also present with DLBD (Ross et al., 2006).

The vast majority of DLBD and PDD cases, however, are sporadic, with a late age at onset. Far less is known about the genetic contributions to Lewy body dementia than is known about AD. There is some suggestive evidence that the *APOE* ε4 allele is increased in individuals with DLBD, as it clearly is in AD, but the evidence is less clear and somewhat contradictory (Jasinska-Myga et al., 2007). One confounding factor is that many patients have a mixed pathological picture at autopsy, with both diffuse cortical Lewy bodies and Alzheimer pathology of amyloid plaques and neurofibrillary tangles. The possible role of the *APOE* ε4 allele in DLBD is difficult to determine when there is mixed pathology and when AD, in which the *APOE* ε4 clearly plays a role, is also present.

Prion Diseases

Prion diseases (spongiform encephalopathies) are a rare cause of dementia, occurring in approximately one in a million individuals (Eggenberger, 2007). About 15% of all spongiform encephalopathy is familial and is referred to as familial Creutzfeldt–Jakob disease (fCJD). In this disorder, patients suffer a general loss of their sense of well being and undergo vague personality changes, confusion, and a form of dementia characterized by difficulties with judgment, memory, and reasoning. Ataxia, myoclonic jerking, and choreo-athetosis appear soon after the initial signs and symptoms. At autopsy, there is widespread neuronal loss with diffuse spongiform change; deposition of amyloid plaques that stain positive with antibodies against the prion protein may also be present. Although classical fCJD has a different clinical presentation and course than early-onset familial AD, there is enough difficulty in making the clinical distinction antemortem that the prion protein gene should be considered for sequencing along with *APPl*, *PSEN1*, and *PSEN2* in all cases of early-onset, familial AD (Finckh et al., 2000).

Other rarer variants of familial prion disease are known that differ in their phenotypic manifestations from fCJD. These disorders, known as fatal familial insomnia (FFI) (Gambetti et al., 1995) and Gerstmann–Sträussler–Scheinker syndrome (GSSs) (Ghetti et al., 1996), occur in patients carrying different missense or frameshift mutations in the prion protein gene. In these disorders, severe autonomic and brainstem dysfunction (in FFI) or cerebellar and basal ganglia dysfunction (in GSSs) are far more prominent than is dementia, although cognition is generally not spared in these familial prion disease variants, especially late in the course.

Prion diseases are caused by aggregation of the prion protein, a naturally occurring neuronal protein that, upon adoption of an abnormal conformation, has a very high propensity to aggregate. This conformational change is induced in normal prion protein molecules by prion molecules that have already adopted the aggregation-prone conformation, thus allowing for propagation of the abnormal conformation and acceleration of the protein aggregation. In familial prion diseases, mutations in the prion gene increase its propensity to adopt an aggregation-prone conformation.

Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy

Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) is a rare form of early-onset dementia inherited in an autosomal dominant manner. The disorder has a prevalence of approximately 1 per 50,000 and most commonly presents between age 20 and 60, with a peak in the fifth decade (Opherk et al., 2004; Razvi et al., 2005). Migraine headache is a frequent early symptom of CADASIL patients as are transient ischemic attacks, neuropsychiatric abnormalities, seizures, and, later, a progressive dementia with prominent defects in executive function, judgment, and language (Dichgans et al., 1998; Opherk et al., 2004). Patients generally survive 5–10 years after diagnosis with progressive dementia and neurological disability leading to death.

As opposed to all the other hereditary forms of primary dementia discussed previously in this chapter, CADASIL is not a primary neurodegenerative disorder with protein aggregation in or around neurons. Instead, CADASIL is characterized by an angiopathy in which there is degeneration of vascular smooth muscle cells in association with electron-dense, extracellular particles, visible with the electron microscope, in and around degenerating smooth muscle cells in the media of arterioles. This characteristic finding can be seen in tissue obtained at autopsy or in biopsy material from skin. The angiopathy in CADASIL leads to a progressive, occlusive disease of small blood vessels, resulting in multiple small infarcts, called lacunae, throughout the brain. The neurological defects seen with lacunar infarcts are usually more circumscribed both clinically and pathologically than those seen with the typical large cerebral artery occlusive stroke (Mohr, 1983; van den Boom et al., 2002). Lacunar infarcts in the cortex may result in a clinical

picture of a pure motor stroke, a pure sensory stroke, or a mixed stroke, while the classic syndrome of dysarthria/clumsy hand is seen with deep pontine lacunae. The accumulation of lacunar infarcts throughout the cortex is the likely cause of the progressive dementia seen in three-quarters of patients with CADASIL (Dichgans et al., 1998; Opherk et al., 2004).

The only molecular lesions known to cause CADASIL are mutations in *NOTCH3*, which encodes the cell surface receptor NOTC3 (Dotti et al., 2005; Federico et al., 2005; Joutel and Tournier-Lasserve, 1998). Most *NOTCH3* mutations in CADASIL are missense or splicing mutations that result in an odd number, instead of the normal even number, of cysteine residues in the extracellular epidermal growth factor-like repeats of NOTC3. The abnormal conformation caused by having an odd number of cysteine residues in the NOTC3 extracellular domain leads to the accumulation of misfolded NOTC3 fragments that form the deposits in the arteriolar walls that are visible on electron microscopy (Joutel and Tournier-Lasserve, 1998; Schroder et al., 1995). How these mutations lead to the aggregation of NOTC3 fragments is not understood.

CADASIL is inherited as an autosomal dominant disorder, nearly always from a carrier parent (i.e., new mutations are rare). CADASIL is one of the few human autosomal dominant diseases that is a true dominant rather than a semidominant condition, in that homozygotes and heterozygotes have a very similar clinical phenotype (Tuominen et al., 2001).

CLINICAL APPROACH TO THE DEMENTIAS

The clinical approach to the dementias relies on careful history and neurological examination, neuropsychological testing, family history, brain imaging, and molecular diagnostics when appropriate. The rapidity of disease progression, the presence of neurological signs besides dementia, such as Parkinsonism as seen with DLBD or FTD, or ataxia, dysarthria and myoclonus as in fCJD, the pattern of cognitive losses (language, short-term memory, judgment or executive function), which differ among the FTDs and AD, the degree of atrophy of frontal versus temporal or hippocampal structures seen on brain MRI, the presence of vascular-occlusive lacunae on brain MRI, may all be helpful in making an antemortem diagnosis (Grossman et al., 2007; Mastrianni and Roos, 2000; van den Boom et al., 2002, 2003). Unfortunately, definitive diagnosis of the cause of dementia is not always clear prior to death and requires postmortem examination at autopsy.

When the clinical presentation, family history, neurological examination, and neuropsychological profile are suggestive of one of the dementias in which the gene is known, testing the affected patient by DNA sequencing of the appropriate genes can provide a definitive diagnosis. Physicians who submit DNA for testing should be aware, however, that the test may show a gene variant that has not been seen before and is of unknown pathogenic significance. It is also possible that the sequencing will be normal in a patient with EOAD or FTD because the disease is due to mutations in a yet to be discovered gene.

It is important to distinguish between diagnostic testing in an affected individual and presymptomatic testing of asymptomatic relatives of a patient shown to have a pathogenic mutation in a known gene. At-risk, asymptomatic relatives may choose to have testing to help with decision-making with regard to having more children or for the purposes of career or financial planning. For some of these disorders, such as CADASIL, FTD, or AD due to *APP* or *PSEN1* mutations, penetrance is very high, although age-dependent. A positive test result in these diseases has high predictive value for development of the disease, but not necessarily for when it will develop. In another familial dementia, such as AD due to *PSEN2* mutations or DLBD/PDD due to *LRRK2* mutations, penetrance is much lower and, as a consequence, a positive test has much lower positive predictive value. Because of the risk for serious psychological and potential financial (life or disability insurance) damage from receiving a positive test, pre-test counseling and post-test psychological and social support are mandatory. Most adults who choose such testing and test positive seem able to handle and use the information (Steinbart et al., 2001). Genetic professionals universally recommend against testing children for these disorders.

The role of testing for the ε4 allele at the *APOE* locus in AD is more controversial. In an affected individual in the 65–75 age range with suspected AD, the presence of one or two ε4 *APOE* alleles has a positive predictive value for AD of ~75% and ~98% respectively and, therefore, may help to confirm a diagnosis of AD (Nussbaum et al., 2007). Most physicians, however, would still make sure to rule out one of the rare but treatable causes of dementia, such as depression or endocrine imbalance, regardless of *APOE* genotype. Predictive testing using *APOE* genotype is currently considered to be of little value because no intervention is currently available to prevent or delay the disease process. In the over 65 age group, when ~1 in 50 individuals would be expected to develop dementia each year, the vast majority of people with one ε4 allele and most of those with two ε4 alleles will still not develop AD. However, it should be noted that the utility of identifying individuals at increased risk for AD through *APOE* genotyping would increase dramatically once a preventive intervention becomes available.

2009 UPDATE

Genetic Association Studies

Whole-genome association studies continued apace in an attempt to confirm, narrow, or disprove the association of various regions of the genome with late-onset AD. The results reinforce once more how difficult it can be to pin down real genetic associations in a complex disease. For example, there were four regions that were noted in the original version of this chapter to have support from more than one association study, that is, regions on chromosome 6, 10, 12, and 9. In the interim, no specific gene and gene variant has been identified and confirmed to be associated with the risk for late-onset AD in any of these regions. In the meantime, additional large genome-wide associations have been reported that suggest there may be other loci that will now need further study and replication (Bertram et al., 2008; Li et al., 2008a; Zuchner et al., 2008).

New association analyses with the sortilin-related receptor gene, *SORL1*, have been published as follow-up to the original report of an association between variants at this gene and AD (Rogaeva et al., 2007). Variants at the *SORL1* locus were initially chosen for association analysis with late-onset AD because of the role of *SORL1* in the biology of intracellular trafficking of APP. This association has undergone further evaluation with conflicting results. Meng et al. (2007) mined the large genome-wide association study from Reiman et al. (2007) for any evidence of association with *SORL1*. They reported a positive replication with *SORL1* without applying a multiple testing correction (Bonferroni), arguing it was unnecessary to do so since they used the dataset specifically to ask about this one locus and not to do a genome-wide association. The very same dataset was interpreted differently by Webster et al. (2008), who reported only nominally significant association with some, but not other, SNPs in the linkage disequilibrium block around *SORL1*. In particular, they did not replicate the association with the specific SNPs that showed the initial association reported by Rogaeva et al. (2007). Other studies have either replicated the *SORL1* association (Lee et al., 2007, 2008), found the association to be marginal (Li et al., 2008b), or failed to replicate it (Minster et al., 2008; Shibata et al., 2008).

Reiman et al. (2007) reported a potential gene–gene interaction in their association study of late-onset AD. They performed a genome-wide association using ~500,000 SNPs in a well-designed case-control study consisting of three Caucasian cohorts, one an autopsy-proven group of AD patients serving as "discovery" cohort, a second autopsy-proven group of AD patients serving as "replication" cohort, and a third group of living patients with clinically-diagnosed AD patients serving as a second replication cohort. They reported a highly significant association between SNPs in a 189 kb linkage disequilibrium block containing *GAB2*, which encodes a GRB-binding partner involved in PI3-kinase signaling and late-onset AD based on Bonferroni-corrected p values to correct for multiple testing. Notably, the association was found only in individuals who carried at least one ε4 allele at the *APOE* locus. Among individuals homozygous for *APOE* ε4, being homozygous as well for the *GAB2* risk haplotype increased the risk of developing AD (odds ratio = 4) as compared to those who did not have the *GAB2* risk haplotype. The *GAB2* haplotype had no effect on disease risk in the absence of an ε4 allele. As is often the case with genome-wide association studies, replicating a finding in one ethnic group failed in another ethnic group. A study in Japanese cohorts failed to find any effect of *GAB2* haplotype on the risk for AD in individuals who were ε4ε4 homozygotes (Miyashita et al., 2008). Thus, additional cohorts will need to be studied.

Proteomics in the Diagnosis of Dementia

Although the rapidity of the disease process, the specific neuropsychiatric signs exhibited, and the findings on brain MRI are all very helpful in distinguishing among the various forms of dementia, it still remains difficult to make a definitive diagnosis prior to death and autopsy study of the brain. Attempts to improve the diagnostic capabilities have recently focused on proteomic analysis of serum or CSF to find distinguishing protein markers for the different disorders discussed in this chapter (Aluise et al., 2008; Roche et al., 2008). Methods that rely on two-dimensional protein electrophoresis or surface-enhanced laser desorption and ionization mass time-of-flight mass spectroscopy (SELDI-TOF MS) to separate proteins and look for differences among the different clinical entities have begun to show success, although their wide-spread clinical use awaits further validation. For example, the presence of certain specific protein biomarkers in serum analysis by SELDI-TOF MS can distinguish diffuse Lewy body disease patients from AD patients, with positive predictive values of 91% and negative predictive value of 92% (Wada-Isoe et al., 2007). In another application, CSF analysis by two-dimensional electrophoresis identifies biomarkers that are far more sensitive and specific for Creutzfeldt–Jacob disease than the analysis of 14-3-3 chaperone protein currently used in clinical practice (Cepek et al., 2007).

Another major clinical challenge in the dementias is identifying those patients with MCI who will go on to develop progressive AD. For example, CSF findings such as reduced levels of Aβ42, increased levels of tau, the ratio of Aβ42 to Aβ40, and the ratio of various phosphorylated species of tau to Abeta42 together provide a sensitivity of 90–95% and specificity of 78–87% for the transition from MCI to AD in patients followed from 4–6 years after CSF testing (Hansson et al., 2006, 2007). A number of additional components of CSF and serum have been studied for their predictive value for MCI patients going on to develop AD. However, to date, there is no evidence that there is substantial improvement over currently used biomarkers (reviewed in Frankfort et al., 2008 and Craig-Schapiro et al., 2008).

2009 UPDATE REFERENCES

Aluise, C.D., Sowell, R.A. and Butterfield, D.A. (2008). Peptides and proteins in plasma and cerebrospinal fluid as biomarkers for the prediction, diagnosis, and monitoring of therapeutic efficacy of Alzheimer's disease. *Biochim Biophys Acta* 1782, 549–558.

Bertram, L., Lange, C., Mullin, K., Parkinson, M., Hsiao, M., Hogan, M.F., Schjeide, B.M., Hooli, B., Divito, J., Ionita, I. et al. (2008). Genome-wide association analysis reveals putative Alzheimer's disease susceptibility loci in addition to APOE. *Am J Hum Genet* 83, 623–632.

Cepek, L., Brechlin, P., Steinacker, P., Mollenhauer, B., Klingebiel, E., Bibl, M., Kretzschmar, H.A., Wiltfang, J. and Otto, M. (2007). Proteomic analysis of the cerebrospinal fluid of patients with Creutzfeldt–Jakob disease. *Dement Geriatr Cogn Disord* 23, 22–28.

Craig-Schapiro, R., Fagan, A.M. and Holtzman, D.M. Biomarkers of Alzheimer's disease. *Neurobiol Dis* 2008 October 28; [Epub ahead of print].

Frankfort, S.V., Tulner, L.R., van Campen, J.P., Verbeek, M.M., Jansen, R.W. and Beijnen, J.H. (2008). Amyloid beta protein and tau in cerebrospinal fluid and plasma as biomarkers for dementia: A review of recent literature. *Curr Clin Pharmacol* 3, 123–131.

Hansson, O., Zetterberg, H., Buchhave, P., Andreasson, U., Londos, E., Minthon, L. and Blennow, K. (2007). Prediction of Alzheimer's disease using the CSF Abeta42/Abeta40 ratio in patients with mild cognitive impairment. *Dement Geriatr Cogn Disord* 23, 316–320.

Hansson, O., Zetterberg, H., Buchhave, P., Londos, E., Blennow, K. and Minthon, L. (2006). Association between CSF biomarkers and incipient Alzheimer's disease in patients with mild cognitive impairment: A follow-up study. *Lancet Neurol* 5, 228–234.

Lee, J.H., Cheng, R., Honig, L.S., Vonsattel, J.P., Clark, L. and Mayeux, R. (2008). Association between genetic variants in SORL1 and autopsy-confirmed Alzheimer disease. *Neurology* 70, 887–889.

Lee, J.H., Cheng, R., Schupf, N., Manly, J., Lantigua, R., Stern, Y., Rogaeva, E., Wakutani, Y., Farrer, L., St George-Hyslop, P. and Mayeux, R. (2007). The association between genetic variants in SORL1 and Alzheimer disease in an urban, multiethnic, community-based cohort. *Arch Neurol* 64, 501–506.

Li, H., Wetten, S., Li, L., St Jean, P.L., Upmanyu, R., Surh, L., Hosford, D., Barnes, M.R., Briley, J.D., Borrie, M. et al. (2008a). Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. *Arch Neurol* 65, 45–53.

Li, Y., Rowland, C., Catanese, J., Morris, J., Lovestone, S., O'Donovan, M.C., Goate, A., Owen, M., Williams, J. and Grupe, A. (2008b). SORL1 variants and risk of late-onset Alzheimer's disease. *Neurobiol Dis* 29, 293–296.

Meng, Y., Lee, J.H., Cheng, R., St George-Hyslop, P., Mayeux, R. and Farrer, L.A. (2007). Association between SORL1 and Alzheimer's disease in a genome-wide study. *Neuroreport* 18, 1761–1764.

Minster, R.L., DeKosky, S.T. and Kamboh, M.I. (2008). No association of SORL1 SNPs with Alzheimer's disease. *Neurosci Lett* 440, 190–192.

Miyashita, A., Arai, H., Asada, T., Imagawa, M., Shoji, M., Higuchi, S., Urakami, K., Toyabe, S., Akazawa, K., Kanazawa, I., Ihara, Y. and Kuwano, R. GAB2 is not associated with late-onset Alzheimer's disease in Japanese. *Eur J Hum Genet* 2008 October 15; [Epub ahead of print].

Reiman, E.M., Webster, J.A., Myers, A.J., Hardy, J., Dunckley, T., Zismann, V.L., Joshipura, K.D., Pearson, J.V., Hu-Lince, D., Huentelman, M.J. et al. (2007). GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers. *Neuron* 54, 713–720.

Roche, S., Gabelle, A. and Lehmann, S. (2008). Clinical proteomics of the cerebrospinal fluid: Towards the discovery of new biomarkers. *Proteomics Clin Appl* 2, 428–436.

Rogaeva, E., Meng, Y., Lee, J.H., Gu, Y., Kawarai, T., Zou, F., Katayama, T., Baldwin, C.T., Cheng, R., Hasegawa, H. et al. (2007). The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. *Nat Genet* 39, 168–177.

Shibata, N., Ohnuma, T., Baba, H., Higashi, S., Nishioka, K. and Arai, H. (2008). Genetic association between SORL1 polymorphisms and Alzheimer's disease in a Japanese population. *Dement Geriatr Cogn Disord* 26, 161–164.

Wada-Isoe, K., Michio, K., Imamura, K., Nakaso, K., Kusumi, M., Kowa, H. and Nakashima, K. (2007). Serum proteomic profiling of dementia with Lewy bodies: Diagnostic potential of SELDI-TOF MS analysis. *J Neural Transm* 114, 1579–1583.

Webster, J.A., Myers, A.J., Pearson, J.V., Craig, D.W., Hu-Lince, D., Coon, K.D., Zismann, V.L., Beach, T., Leung, D., Bryden, L. et al. (2008). Sorl1 as an Alzheimer's disease predisposition gene? *Neurodegener Dis* 5, 60–64.

Zuchner, S., Gilbert, J.R., Martin, E.R., Leon-Guerrero, C.R., Xu, P.T., Browning, C., Bronson, P.G., Whitehead, P., Schmechel, D.E., Haines, J.L. and Pericak-Vance, M.A. (2008). Linkage and association study of late-onset Alzheimer disease families linked to 9p21.3. *Ann Hum Genet* 72, 725–731.

REFERENCES

Barker, W.W., Luis, C.A., Kashuba, A., Luis, M., Harwood, D.G., Loewenstein, D., Waters, C., Jimison, P., Shepherd, E., Sevush, S. et al. (2002). Relative frequencies of Alzheimer disease, Lewy body, vascular and frontotemporal dementia, and hippocampal sclerosis in the State of Florida Brain Bank. *Alzheimer Dis Assoc Disord* 16, 203–212.

Bertram, L., McQueen, M.B., Mullin, K., Blacker, D. and Tanzi, R.E. (2007). Systematic meta-analyses of Alzheimer disease genetic association studies: The AlzGene database. *Nat Genet* 39, 17–23.

Braak, H., Tredici, K.D., Rub, U., de Vos, R.A., Jansen Steur, E.N. and Braak, E. (2003). Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol Aging* 24, 197–211.

Braak, H., Ghebremedhin, E., Rub, U., Bratzke, H. and Del Tredici, K. (2004). Stages in the development of Parkinson's disease-related pathology. *Cell Tissue Res* 318, 121–134.

Cabrejo, L., Guyant-Marechal, L., Laquerriere, A., Vercelletto, M., De la Fourniere, F., Thomas-Anterion, C., Verny, C., Letournel, F., Pasquier, F., Vital, A. et al. (2006). Phenotype associated with APP duplication in five families. *Brain* 129, 2966–2976.

Chartier-Harlin, M.C., Kachergus, J., Roumier, C., Mouroux, V., Douay, X., Lincoln, S., Levecque, C., Larvor, L., Andrieux, J., Hulihan, M. et al. (2004). Alpha-synuclein locus duplication as a cause of familial Parkinson's disease. *Lancet* 364, 1167–1169.

Citron, M., Westaway, D., Xia, W., Carlson, G., Diehl, T., Levesque, G., Johnson-Wood, K., Lee, M., Seubert, P., Davis, A. et al. (1997). Mutant presenilins of Alzheimer's disease increase production of 42-residue amyloid beta-protein in both transfected cells and transgenic mice. *Nat Med* 3, 67–72.

Coon, K.D., Myers, A.J., Craig, D.W., Webster, J.A., Pearson, J.V., Lince, D.H., Zismann, V.L., Beach, T.G., Leung, D., Bryden, L. et al. (2007). A high-density whole-genome association study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease. *J Clin Psychiatr* 68, 613–618.

Corder, E.H., Saunders, A.M., Strittmatter, W.J., Schmechel, D.E., Gaskell, P.C., Small, G.W., Roses, A.D., Haines, J.L. and Pericak-Vance, M.A. (1993). Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 261, 921–923.

Dichgans, M., Mayer, M., Uttner, I., Bruning, R., Muller-Hocker, J., Rungger, G., Ebke, M., Klockgether, T. and Gasser, T. (1998). The phenotypic spectrum of CADASIL: Clinical findings in 102 cases. *Ann Neurol* 44, 731–739.

Dotti, M.T., Federico, A., Mazzei, R., Bianchi, S., Scali, O., Conforti, F.L., Sprovieri, T., Guidetti, D., Aguglia, U., Consoli, D. et al. (2005). The spectrum of Notch3 mutations in 28 Italian CADASIL families. *J Neurol Neurosurg Psychiatr* 76, 736–738.

Dugue, M., Neugroschl, J., Sewell, M. and Marin, D. (2003). Review of dementia. *Mt Sinai J Med* 70, 45–53.

Eckman, C.B. and Eckman, E.A. (2007). An update on the amyloid hypothesis. *Neurol Clin* 25, 669–682.

Eggenberger, E. (2007). Prion disease. *Neurol Clin* 25, 833–842.

Ertekin-Taner, N. (2007). Genetics of Alzheimer's disease: A centennial review. *Neurol Clin* 25, 611–667.

Farrer, M., Kachergus, J., Forno, L., Lincoln, S., Wang, D.S., Hulihan, M., Maraganore, D., Gwinn-Hardy, K., Wszolek, Z., Dickson, D. et al. (2004). Comparison of kindreds with parkinsonism and alpha-synuclein genomic multiplications. *Ann Neurol* 55, 174–179.

Federico, A., Bianchi, S. and Dotti, M.T. (2005). The spectrum of mutations for CADASIL diagnosis. *Neurol Sci* 26, 117–124.

Finckh, U. (2003). The future of genetic association studies in Alzheimer disease. *J Neural Transm* 110, 253–266.

Finckh, U., Muller-Thomsen, T., Mann, U., Eggers, C., Marksteiner, J., Meins, W., Binetti, G., Alberici, A., Hock, C., Nitsch, R.M. et al. (2000). High prevalence of pathogenic mutations in patients with early-onset dementia detected by sequence analyses of four different genes. *Am J Hum Genet* 66, 110–117.

Gambetti, P., Parchi, P., Petersen, R.B., Chen, S.G. and Lugaresi, E. (1995). Fatal familial insomnia and familial Creutzfeldt–Jakob disease: Clinical, pathological and molecular features. *Brain Pathol* 5, 43–51.

Gatz, M., Fratiglioni, L., Johansson, B., Berg, S., Mortimer, J.A., Reynolds, C.A., Fiske, A. and Pedersen, N.L. (2005). Complete ascertainment of dementia in the Swedish Twin Registry: The HARMONY study. *Neurobiol Aging* 26, 439–447.

Ghetti, B., Piccardo, P., Frangione, B., Bugiani, O., Giaccone, G., Young, K., Prelli, F., Farlow, M.R., Dlouhy, S.R. and Tagliavini, F. (1996). Prion protein amyloidosis. *Brain Pathol* 6, 127–145.

Grossman, M., Libon, D.J., Forman, M.S., Massimo, L., Wood, E., Moore, P., Anderson, C., Farmer, J., Chatterjee, A., Clark, C.M. et al. (2007). Distinct antemortem profiles in patients with pathologically defined frontotemporal dementia. *Arch Neurol* 64, 1601–1609.

Haugarvoll, K., Wszolek, Z.K. and Hutton, M. (2007). The genetics of frontotemporal dementia. *Neurol Clin* 25, 697–715.

Hogan, S. and Hogan, S. (2003). How an aging population will affect health care. In *The Fiscal Sustainability of Health Care in Canada* (G.P. Marchildon, T. McIntosh and P-G. Forest, eds) vol. 1. University of Toronto, Toronto, Canada, pp. 51–80.

Hutton, M. (2001). Missense and splice site mutations in tau associated with FTDP-17: Multiple pathogenic mechanisms. *Neurology* 56, S21–S25.

Jasinska-Myga, B., Opala, G., Goetz, C.G., Tustanowski, J., Ochudlo, S., Gorzkowska, A. and Tyrpa, J. (2007). Apolipoprotein E gene polymorphism, total plasma cholesterol level, and Parkinson disease dementia. *Arch Neurol* 64, 261–265.

Joutel, A. and Tournier-Lasserve, E. (1998). Notch signalling pathway and human diseases. *Semin Cell Dev Biol* 9, 619–625.

Knopman, D.S., Petersen, R.C., Cha, R.H., Edland, S.D. and Rocca, W.A. (2006). Incidence and causes of nondegenerative nonvascular dementia: A population-based study. *Arch Neurol* 63, 218–221.

Kruger, R., Kuhn, W., Muller, T., Woitalla, D., Graeber, S., Kosel, S., Przuntek, H., Epplen, J.T., Schols, L. and Riess, O. (1998). Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. *Nat Genet* 18, 106–108.

Kukull, W.A. and Ganguli, M. (2000). Epidemiology of dementia: Concepts and overview. *Neurol Clin* 18, 923–950.

Kukull, W.A., Higdon, R., Bowen, J.D., McCormick, W.C., Teri, L., Schellenberg, G.D., van Belle, G., Jolley, L. and Larson, E.B. (2002). Dementia and Alzheimer disease incidence: A prospective cohort study. *Arch Neurol* 59, 1737–1746.

Lai, F. and Williams, R.S. (1989). A prospective study of Alzheimer disease in Down syndrome. *Arch Neurol* 46, 849–853.

Lautenschlager, N.T., Cupples, L.A., Rao, V.S., Auerbach, S.A., Becker, R., Burke, J., Chui, H., Duara, R., Foley, E.J., Glatt, S.L. et al. (1996). Risk of dementia among relatives of Alzheimer's disease patients in the MIRAGE study: What is in store for the oldest old?. *Neurology* 46, 641–650.

Lee, J.H., Cheng, R., Schupf, N., Manly, J., Lantigua, R., Stern, Y., Rogaeva, E., Wakutani, Y., Farrer, L., St George-Hyslop, P. et al. (2007). The association between genetic variants in SORL1 and Alzheimer disease in an urban, multiethnic, community-based cohort. *Arch Neurol* 64, 501–506.

Lippa, C.F., Duda, J.E., Grossman, M., Hurtig, H.I., Aarsland, D., Boeve, B.F., Brooks, D.J., Dickson, D.W., Dubois, B., Emre, M. et al. (2007). DLB and PDD boundary issues: Diagnosis, treatment, molecular pathology, and biomarkers. *Neurology* 68, 812–819.

Lunetta, K.L., Erlich, P.M., Cuenco, K.T., Cupples, L.A., Green, R.C., Farrer, L.A. and Decarli, C. (2007). Heritability of magnetic resonance imaging (MRI) traits in Alzheimer disease cases and their siblings in the MIRAGE study. *Alzheimer Dis Assoc Disord* 21, 85–91.

Mahley, R.W., Weisgraber, K.H. and Huang, Y. (2006). Apolipoprotein E4: A causative factor and therapeutic target in neuropathology, including Alzheimer's disease. *Proc Natl Acad Sci USA* 103, 5644–5651.

Martinez, M., Campion, D., Brice, A., Hannequin, D., Dubois, B., Didierjean, O., Michon, A., Thomas-Anterion, C., Puel, M., Frebourg, T. et al. (1998). Apolipoprotein E epsilon4 allele and familial aggregation of Alzheimer disease. *Arch Neurol* 55, 810–816.

Maslow, K. (2006). Early Onset Dementia: A National Challenge, a Future Crisis. Alzheimer's Association.

Mastrianni, J.A. and Roos, R.P. (2000). The prion diseases. *Semin Neurol* 20, 337–352.

Mayeux, R., Denaro, J., Hemenegildo, N., Marder, K., Tang, M.X., Cote, L.J. and Stern, Y. (1992). A population-based investigation of Parkinson's disease with and without dementia. Relationship to age and gender. *Arch Neurol* 49, 492–497.

McKeith, I.G., Dickson, D.W., Lowe, J., Emre, M., O'Brien, J.T., Feldman, H., Cummings, J., Duda, J.E., Lippa, C., Perry, E.K. et al. (2005). Diagnosis and management of dementia with Lewy bodies: Third report of the DLB Consortium. *Neurology* 65, 1863–1872.

McMurtray, A., Clark, D.G., Christine, D. and Mendez, M.F. (2006). Early-onset dementia: Frequency and causes compared to late-onset dementia. *Dement Geriatr Cogn Disord* 21, 59–64.

Mindham, R.H., Biggins, C.A., Boyd, J.L., Harrop, F.M., Madeley, P., Randall, J.I. and Spokes, E.G. (1993). A controlled study of dementia in Parkinson's disease over 54 months. *Adv Neurol* 60, 470–474.

Mohr, J.P. (1983). Lacunes. *Neurol Clin* 1, 201–221.

Murrell, J.R., Price, B., Lane, K.A., Baiyewu, O., Gureje, O., Ogunniyi, A., Unverzagt, F.W., Smith-Gamble, V., Gao, S., Hendrie, H.C. et al. (2006). Association of apolipoprotein E genotype and Alzheimer disease in African Americans. *Arch Neurol* 63, 431–434.

Nussbaum, R.L. and Ellis, C.E. (2003). Alzheimer's disease and Parkinson's disease. *N Engl J Med* 348, 1356–1364.

Nussbaum, R.L., Mcinnes, R.R., Willard, H.F. (2007). Genetics of common disorders with complex inheritance. In *Thompson and Thompson Genetics in Medicine, 7th edition*. Saunders Elsevier, Philadelphia, PA, pp. 151–174.

Opherk, C., Peters, N., Herzog, J., Luedtke, R. and Dichgans, M. (2004). Long-term prognosis and causes of death in CADASIL: A retrospective study in 411 patients. *Brain* 127, 2533–2539.

Pedersen, N.L., Gatz, M., Berg, S. and Johansson, B. (2004). How heritable is Alzheimer's disease late in life? Findings from Swedish twins. *Ann Neurol* 55, 180–185.

Petersen, R.C., Smith, G.E., Ivnik, R.J., Tangalos, E.G., Schaid, D.J., Thibodeau, S.N., Kokmen, E., Waring, S.C. and Kurland, L.T. (1995). Apolipoprotein E status as a predictor of the development of Alzheimer's disease in memory-impaired individuals. *JAMA* 273, 1274–1278.

Polymeropoulos, M.H., Lavedan, C., Leroy, E., Ide, S.E., Dehejia, A., Dutra, A., Pike, B., Root, H., Rubenstein, J., Boyer, R. et al. (1997). Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science* 276, 2045–2047.

Rao, V.S., van Duijn, C.M., Connor-Lacke, L., Cupples, L.A., Growdon, J.H. and Farrer, L.A. (1994). Multiple etiologies for Alzheimer disease are revealed by segregation analysis. *Am J Hum Genet* 55, 991–1000.

Ratnavalli, E., Brayne, C., Dawson, K. and Hodges, J.R. (2002). The prevalence of frontotemporal dementia. *Neurology* 58, 1615–1621.

Razvi, S.S., Davidson, R., Bone, I. and Muir, K.W. (2005). The prevalence of cerebral autosomal dominant arteriopathy with subcortical infarcts and leucoencephalopathy (CADASIL) in the west of Scotland. *J Neurol Neurosurg Psychiatr* 76, 739–741.

Rogaeva, E., Kawarai, T. and George-Hyslop, P.S. (2006). Genetic complexity of Alzheimer's disease: Successes and challenges. *J Alzheimers Dis* 9, 381–387.

Rogaeva, E., Meng, Y., Lee, J.H., Gu, Y., Kawarai, T., Zou, F., Katayama, T., Baldwin, C.T., Cheng, R., Hasegawa, H. et al. (2007). The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. *Nat Genet* 39, 168–177.

Ross, O.A., Toft, M., Whittle, A.J., Johnson, J.L., Papapetropoulos, S., Mash, D.C., Litvan, I., Gordon, M.F., Wszolek, Z.K., Farrer, M.J. et al. (2006). Lrrk2 and Lewy body disease. *Ann Neurol* 59, 388–393.

Rovelet-Lecrux, A., Hannequin, D., Raux, G., Le Meur, N., Laquerriere, A., Vital, A., Dumanchin, C., Feuillette, S., Brice, A., Vercelletto, M. et al. (2006). APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat Genet* 38, 24–26.

Rovelet-Lecrux, A., Frebourg, T., Tuominen, H., Majamaa, K., Campion, D., Remes, A.M. (2007). APP locus duplication in a Finnish family with dementia and intracerebral haemorrhage. *J Neurol Neurosurg Psychiatr* 78, 1158–1159.

Saunders, A.M., Strittmatter, W.J., Schmechel, D., George-Hyslop, P.H., Pericak-Vance, M.A., Joo, S.H., Rosi, B.L., Gusella, J.F., Crapper-MacLachlan, D.R., Alberts, M.J. et al. (1993). Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology* 43, 1467–1472.

Schroder, J.M., Sellhaus, B. and Jorg, J. (1995). Identification of the characteristic vascular changes in a sural nerve biopsy of a case with cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL). *Acta Neuropathol (Berl)* 89, 116–121.

Schupf, N., Patel, B., Pang, D., Zigman, W.B., Silverman, W., Mehta, P.D. and Mayeux, R. (2007). Elevated plasma beta-amyloid peptide Abeta(42) levels, incident dementia, and mortality in Down syndrome. *Arch Neurol* 64, 1007–1013.

Seshadri, S., Drachman, D.A. and Lippa, C.F. (1995). Apolipoprotein E epsilon 4 allele and the lifetime risk of Alzheimer's disease. What physicians know, and what they should know. *Arch Neurol* 52, 1074–1079.

Singleton, A.B., Farrer, M., Johnson, J., Singleton, A., Hague, S., Kachergus, J., Hulihan, M., Peuralinna, T., Dutra, A., Nussbaum, R. et al. (2003). alpha-Synuclein locus triplication causes Parkinson's disease. *Science* 302, 841.

Steinbart, E.J., Smith, C.O., Poorkaj, P. and Bird, T.D. (2001). Impact of DNA testing for early-onset familial Alzheimer disease and frontotemporal dementia. *Arch Neurol* 58, 1828–1831.

Tang, M.X., Stern, Y., Marder, K., Bell, K., Gurland, B., Lantigua, R., Andrews, H., Feng, L., Tycko, B. and Mayeux, R. (1998). The APOE-epsilon4 allele and the risk of Alzheimer disease among African Americans, whites, and Hispanics. *JAMA* 279, 751–755.

Tuominen, S., Juvonen, V., Amberla, K., Jolma, T., Rinne, J.O., Tuisku, S., Kurki, T., Marttila, R., Poyhonen, M., Savontaus, M.L. et al. (2001). Phenotype of a homozygous CADASIL patient in comparison to 9 age-matched heterozygous patients with the same R133C Notch3 mutation. *Stroke* 32, 1767–1774.

van den Boom, R., Lesnik Oberstein, S.A., van Duinen, S.G., Bornebroek, M., Ferrari, M.D., Haan, J. and van Buchem, M.A. (2002). Subcortical lacunar lesions: An MR imaging finding in patients with cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy. *Radiology* 224, 791–796.

van den Boom, R., Lesnik Oberstein, S.A., Spilt, A., Behloul, F., Ferrari, M.D., Haan, J., Westendorp, R.G. and van Buchem, M.A. (2003). Cerebral hemodynamics and white matter hyperintensities in CADASIL. *J Cereb Blood Flow Metab* 23, 599–604.

Walker, Z., Costa, D.C., Walker, R.W., Shaw, K., Gacinovic, S., Stevens, T., Livingston, G., Ince, P., McKeith, I.G. and Katona, C.L. (2002). Differentiation of dementia with Lewy bodies from Alzheimer's disease using a dopaminergic presynaptic ligand. *J Neurol Neurosurg Psychiatr* 73, 134–140.

Zarranz, J.J., Alegre, J., Gomez-Esteban, J.C., Lezcano, E., Ros, R., Ampuero, I., Vidal, L., Hoenicka, J., Rodriguez, O., Atares, B. et al. (2004). The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia. *Ann Neurol* 55, 164–173.

RECOMMENDED RESOURCES

AlzGene website. Go to http://www.alzforum.org/res/com/gen/alzgene/default.asp

GeneReviews. Online Reviews of AD, FTD, CADASIL, and fCJD. Go to http://www.geneclinics.org/and click on GeneReviews.

Useful Reviews

Dugue, M., et al. (2003). Review of dementia. *Mt Sinai J Med* 70, 45–53.

Ertekin-Taner, N. (2007). Genetics of Alzheimer's disease: A centennial review. *Neurol Clin* 25, 611–667.

CHAPTER

Genetics and Genomics of Parkinson's Disease

Shushant Jain and Andrew B. Singleton

INTRODUCTION

Parkinson's disease (PD) is the second most common neurodegenerative disease with at least 5 million people affected globally (Twelves et al., 2003). PD was first described in 1817 by James Parkinson in the *Essay of Shaking Palsy*. Subsequently Brissaud noted lesions within the substantia nigra (SN) and, together with Meynerts previous observation that the basal ganglia is involved in movement, concluded that injury of this region was responsible for the symptoms in PD.

Nearly half a century later, dopamine was identified as a neurotransmitter in the basal ganglia and individuals with PD showed a deficiency in dopamine. This subsequently led to the administration of levodopa (*L*-dopa, metabolic precursor of dopamine) which remains the most effective symptomatic treatment for PD (Birkmayer and Hornykiewicz, 1962).

Although the underlying physiological cause of the major movement symptoms of PD had been exposed, the explanation for specific nigral degeneration remains unclear. For many years it was believed that PD was primarily the effect of environmental insult as studies had recognized that individuals with exposure to certain chemicals such as 1-methyl-4-phenyl-1, 2, 3, 6-tetrahydropyridine (MPTP) could lead to a disease with parkinsonism features. Further studies showed that a PD phenotype could arise from multiple different etiologies including vascular insults, infections (post-encephalitic Parkinsonism caused by the influenza virus) and frontal lobe tumors.

Research into PD was revolutionized when a genetic basis for PD was established with the identification of monogenic forms (Table 54.1). Elucidating the genetics and environmental causes of PD has highlighted biological pathways critical in disease pathogenesis. This in turn could allow the subdivision of the disease based on genetic rather than phenotypic information which could help explain the wide variation in this disease, including differences in clinical course, and response to treatment, and it might help clarify the role of environmental factors in disease cause or susceptibility.

CLINICAL CHARACTERISTICS OF PD

PD belongs to a heterogeneous family of diseases referred to as parkinsonism syndromes. Within this group there are many diseases such as progressive supranuclear palsy (PSP), diffuse Lewy body disease (DLB) and environmentally induced parkinsonism, such as exposure to MPTP or other environmental insults. A clinical diagnosis of PD as opposed to other parkinsonism syndromes requires the presence of tremor, rigidity, and akinesia. In addition, there are inclusion and exclusion criteria which also must be fulfilled: (1) no detectable cause, (2) no cerebella deficits, (3) limited pyramidal signs, (4) no lower motor dysfunction, (5) limited gaze palsy, and (6) minor autonomic deficits. A pathological diagnosis dictates loss of dopaminergic cells in the SN and also the presence of intracytoplasmic eosinophilic inclusions called Lewy bodies (LB) in surviving neurons (Hardy and Lees, 2005).

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Table 54.1 Genetic loci implicated in Parkinson's disease

Locus	Protein name	Inheritance pattern	Phenotype
PARK1 PARK4	α-SYNUCLEIN	AD	AOO~40–50 years Features of dementia with Lewy bodies
PARK2	PARKIN	AR	AOO~35 years Slow disease progression *L*-dopa responsive No Lewy bodies
PARK3	Unknown	AD with reduced penetrance	AOO~50 years Typical PD *L*-dopa responsive No pathology available
PARK5	UCHL-1	AD	AOO~50 years Typical PD *L*-dopa responsive No pathology available
PARK6	PINK1	AR	AOO~35–45 years Slow disease progression *L*-dopa responsive No pathology available
PARK7	DJ-l	AR	AOO~30–40 years Slow disease progression *L*-dopa responsive No pathology available
PARK8	LRRK2	AD	AOO~50 years Typical PD*L*-dopa responsive Variable pathology
PARK9	ATP13A2	AR	AOO~16 years Levodopa-responsive parkinsonism with pyramidal degeneration, supranuclear gaze palsy, and dementia
PARK10	Unknown	Risk factor	AOO~50–60 years Typical PD No pathology available
PARK11	Unknown	Risk factor	AOO~50–60 years Typical PD No pathology available
PARK12	Unknown	Risk factor	AOO~50–60 years Typical PD No pathology available
Chromosome 5 (5q23)	Unknown	Risk factor	AOO~50–60 years Typical PD No pathology available

Abbreviations: AOO: Average age of onset, AD: Autosomal dominant, AR: Autosomal recessive.

PD is a late-onset disease, primarily occurring in the fifth or sixth decades, although some forms, particularly the recessive genetic diseases, can begin in childhood. Disease in individuals where a specific etiology is not known and where there is no clear family history of PD are classified as sporadic PD. The early symptoms of PD are usually non-specific and may be seen in many other neurological syndromes or as a part of normal aging. To begin with, there is a general lethargy with possible

mood and mild cognitive impairment. Subsequently, an intermittent tremor often present only under stress can develop with asymmetrical rigidity moving to the other side of the body within 3–5 years. Within 5 years bradykinesia and postural instability ensue (Hughes et al., 1993).

The average mortality rate in PD is approximately 1.5 above the general population. On average disease duration is 13 years, and the mean age at death has been reported at 73 years. The most common causes of death in PD patients is pneumonia through lack of activity, cardiovascular disease, or severe injury through falling (Hughes et al., 1993).

GENETICS OF PD

Many diseases have a genetic component, whether it is due to inherited mutations or as a result of genetic variation controlling the response to environmental stresses such as viruses or toxins. The identification of the genetic causes of a disease allows one to isolate the primary pathogenic mechanism and/or contributors to a disease. The ultimate goal is to use this information to identify and develop new ways to treat, cure, or even prevent the disease.

A common methodology used in the determination of the relative contribution of genetics to disease is the twin study, this is performed by comparing concordance of disease in monozygotic (MZ) twins (who share all genes) and dizygotic (DZ) twins (who share, on average, 50% of autosomal genes). Recent data from twin studies using the uptake of ^{18}F-dopa and positron emission tomography (PET) imaging have suggested that genetics does play a role in disease (Piccini et al., 1997) but also suggest genetics is not the sole determinant of disease. These data are consistent with the most widely held hypothesis, that the majority of typical PD cases are a result of a complex interplay between genetic variability and environmental exposures.

Many genes have been implicated in PD but the analysis of multiple nuclear families or isolated populations has led to the identification of multiple genes and loci that cause mendelian PD or increase risk for PD (Table 54.1).

SNCA (*PARK1*; *PARK4*; *α-SYNUCLEIN*)

α-synuclein was previously cloned as the non-Abeta (β) component of Alzheimer's disease (AD) amyloid plaques. However, its role in PD became evident in 1997 when a mutation (A53T) within a Greek kindred was shown to cause autosomal dominant PD (Polymeropoulos et al., 1997). Subsequently, a further two additional missense mutations (A30P and E64K) have been identified as rare causes of disease as it has multiplication of the genomic segment containing the gene encoding α-synuclein (*SNCA*) (Singleton et al., 2003).

Soon after the discovery of mutations in the gene encoding α-synuclein as the first genetic cause of PD, this protein was found to be the major component of LB, the pathological hallmark of PD (Spillantini et al., 1997). However, the pathology in individuals with α-synuclein mutations is not typical of idiopathic PD; the pathology is usually more extensive with LB not only in the SN but also throughput the cortex, striatum and locus ceruleus; additionally α-synuclein pathology may also be seen outside of neuronal cells, within glia, somewhat similar to the glial cytoplasmic inclusions noted in multiple system atrophy (Mukaetova-Ladinska and McKeith, 2006). The parkinsonism associated with α-synuclein mutations presents at a relatively early age (30s to 50s) and is rapidly progressive, in many cases the disease in patients with α-synuclein mutations progresses to include a prominent dementia, likely a reflection of the extensive cortical pathology noted in these patients (Kruger et al., 2001).

α-synuclein is part of a gene family including γ and β synuclein. The function of α-synuclein is not well understood but many hypothesizes exist regarding its role in PD pathogenesis. α-synuclein is primarily located at synaptic membranes and therefore may have a role in maintaining synaptic function which in part has been supported by animal modeling; analysis of α-synuclein knockout mice has suggested a role for α-synuclein of long-term regulation and/or maintenance of presynaptic function (Kaplan et al., 2003).

Because *SNCA* was the first gene implicated in PD and because its protein product is the major deposited species in the hallmark lesion of this disease, considerable resources have been used in an attempt to understand the pathophysiological process that results from α-synuclein mutation. Firstly α-synuclein can aggregate under a number of different conditions, the central hydrophobic region, where the missense mutations reside, tends to self-aggregate. The end product of α-synuclein aggregation is the formation of heavily insoluble polymers of protein known as fibrils, which is promoted by both the A53T mutation and overexpression of α-synuclein (Figure 54.1) (Volles and Lansbury, 2003). Conversely, A30P slows the rate of fibril accumulation but increases the rate of α-synuclein protofibril formation. Because of this and other data it is now believed that the protofibril species of α-synuclein are the toxic species (Cookson, 2005). α-synuclein protofibrils have the ability to form pores-like structures (Volles and Lansbury, 2003) which can cause leakage of vesicles. Furthermore, PD associated mutations are able to increase the permeabilizing activity of α-synuclein by increasing protofibril formation. The subsequent binding and formation of pores in the mitochondrial or vesicular membranes (Volles and Lansbury, 2003) or at the cell surface, could lead to disruption of numerous cellular activities and cell death.

Another mechanism of α-synuclein toxicity could be mediated in part by its post-translational modification. α-synuclein is phosphorylated at Ser-129 and it is this form that is primarily deposited in LB. In addition, altering this residue to either prevent or mimic phosphorylation suppresses or enhances α-synuclein toxicity respectively in Drosophila transgenic models (Chen and Feany, 2005). If phosphorylation of α-synuclein is a necessary event in its pathogenesis then inhibiting the kinases responsible would be a good target. Numerous studies are also underway to determine if synuclein levels are a good correlate for disease status and hence a biomarker for pre-symptomatic individuals (Miller et al., 2004).

PRKN (*PARK2*; *PARKIN*)

The gene encoding *PARKIN* (*PRKN*) was the first gene to be identified with mutations that underlie autosomal recessive PD

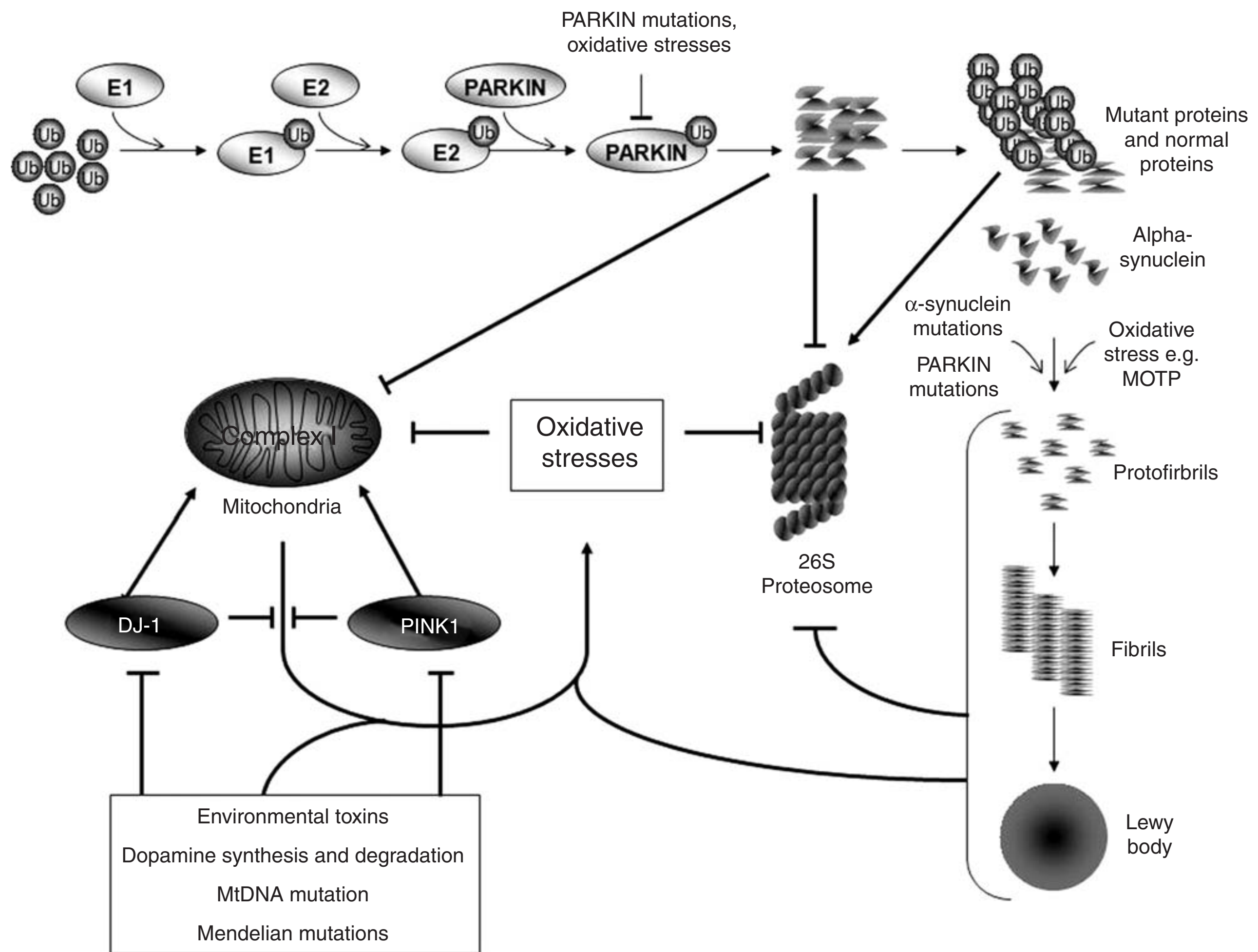


Figure 54.1 A proposed model for mechanisms of cellular toxicity in PD. *PRKN* mutations and oxidative stress can inhibit PARKIN mediated ubiquitination of specific substrates leading to their accumulation. These substrates may inhibit both the proteasome and mitochondria. The formation of α-synuclein protofibrils and aggregates can be toxic to both the mitochondria and proteasome. PINK1 and DJ-1 promote cell survival, either directly or indirectly by protecting mitochondria from oxidative stress.

and represents the most common known cause of early-onset parkinsonism (Kitada et al., 1998). *PRKN* mutations account for nearly 50% familial cases where the age of onset is below 40 years and the probability of a patients having *PRKN*-linked disease increases as age of onset decreases. The clinical picture of *PRKN*-linked disease is not that of typical PD patients commonly present with dystonia, other atypical features include hyperreflexia, slow progression, and early compilations from *L*-dopa treatment. Patients survive an average of 10 to 20 years and have a more symmetrical onset (Lohmann et al., 2003).

Despite the relative abundance of *PRKN*-linked disease there remains a paucity of neuropathological analysis in patients with disease unequivocally caused by *PRKN* mutation. The majority of reports indicate a lack of LB pathology and because the presence of nigral LB is a hallmark pathological feature of PD and a large amount of resources have been invested in an attempt to determine the molecular pathway of disease related to *PRKN* mutation, there has been considerable debate over the relevance of *PRKN*-linked parkinsonism to typical PD (Hardy and Lees, 2005). However, while the nature of LB pathology in this form of parkinsonism is unclear, there is certainly degeneration and dysfunction of the dopaminergic neurons and as such establishing the mechanism of the preferential vulnerability of this neuronal system in *PRKN* disease is likely to be directly relevant to typical PD.

There are multiple types of mutations that cause loss of function mutations within *PRKN*: point mutations to large scale deletions spanning multiple exons, multiplications, frameshift (alters reading frame), truncating, and splice-site mutations. The scale and assortment of mutations within *PRKN* makes genetic diagnosis difficult as extensive screening has to be performed. Furthermore, there are several point mutations where pathogenicity is currently equivocal (von Coelln et al., 2004).

PRKN encodes an E3 ubiquitin ligase which is responsible for the addition of ubiquitin molecules to specific target proteins that are subsequently recognized by the proteasome and degraded (Figure 54.1). As a consequence of the large deletions

and multiple mutations throughout the gene, lack of *PARKIN* may lead to the accumulation of one or more of its substrates and subsequently to cell death. In support of this hypothesis, when suspected substrates of *PARKIN* are overexpressed they can lead to dopaminergic cell loss which can subsequently be rescued by *PARKIN* but not its mutants (Kahle and Haass, 2004). As the accumulation of proteasomal substrates or proteasome inhibition is implicated in PD pathogenesis (McNaught and Olanow, 2006), individuals involved in small molecule proteasome inhibitor clinical trials (e.g. NPI-0052; Chauhan et al., 2006), will have to be carefully monitored to ensure they do not develop parkinsonism features.

Another hypothesis suggests that loss of *PARKIN* may result in mitochondrial damage and apoptosis, as knockout of *PARKIN* homologs from both mice and Drosophila cause decreases in mitochondrial respiratory capacity demonstrated by reduced lifespan, locomotor defects due to apoptotic cell death and male sterility due to spermatid individualization defects; this hypothesis is gaining impetus as the genes mutated in other recessive Mendelian PD cases appear to link mitochondria to neuronal cell loss (Shen and Cookson, 2004).

DJ1 (*PARK7*; *DJ-1*)

DJ-1 mutations are found in young-onset autosomal recessive parkinsonism but are the rarest known genetic cause of parkinsonism (<1–2% of familial PD; Bonifati et al., 2003); however, mutation of *DJ-1* remains the rarest known genetic cause of parkinsonism. Given the rarity of *DJ-1* mutations there is limited clinical data and to date no pathological data available in *DJ-1*-linked patients. Patients with *DJ-1* mutations present with a young-onset disorder (mid-30s) and follows a relatively benign course (Bonifati et al., 2003). Consistent with loss of function, *DJ-1* is recessively inherited and one of the original families possessed a large deletion encompassing the start codon of *DJ-1*. Subsequently, multiple point and splice mutations have been described within *DJ-1*. Although the current function of *DJ-1* is unclear, some mutations (e.g. L166P) destabilize *DJ-1* thus leading to increased degradation by the proteasome. As a consequence there is insufficient *DJ-1* which is hypothesized to increase neuronal vulnerability to toxic insult and apoptosis (Figure 54.1).

Several hypothesis exist as to how *DJ-1* protects cells from toxic insult. *DJ-1* is a 189 amino acid member of the ThiJ/PfpI/DJ1 superfamily, ubiquitously expressed and localizes to the cytosol and mitochondria as well as the nucleus in dividing cells. Under oxidative stress conditions, such as exposure to paraquat or MPTP, *DJ-1* undergoes an acidic shift in pI by modifying the side chain of cysteine 106 to form a sulfinic acid. This is correlated with the protein relocating from the cytosol to the outer mitochondria membrane. Overexpression of *DJ-1* in culture can decrease sensitivity to specific stressors, such as paraquat and MPTP (Canet-Aviles et al., 2004). Conversely, loss of *DJ-1* in mice, cell culture, and Drosophila models leads to increased sensitivity to oxidative stresses (Cookson, 2005; Shen and Cookson, 2004).

DJ-1 only has a weak ability to scavenge free radicals and thus, *DJ-1* is unlikely to primarily function as an anti-oxidant protein. Consequently, a role for *DJ-1* as an oxidative stress sensor has been suggested. *DJ-1* may have an analogous role to the DNA damage sensing enzymes (e.g. ATM, ATR, and RAD proteins) where specific enzymes recognize different types of DNA damage and are able to mediate the appropriate response (e.g. apoptosis, cell cycle arrest, and transcription). As *DJ-1* was cloned as part of a RNA protein-binding complex, it is postulated that *DJ-1* may control transcription and/or translation of particular RNA species in response to oxidative stress (Abou-Sleiman et al., 2003).

Alternatively, *DJ-1* has been shown to bind to numerous proteins such as DAXX, preventing it from activating the apoptotic pathway and decreasing cell sensitivity to oxidative stresses. However, many of these *DJ-1* interactors still require validation *in vivo*, both to confirm the interaction and to establish that they play specific roles in *DJ-1* mediated cell survival (Abou-Sleiman et al., 2003).

PINK1 (*PARK6*; Pten Induced Kinase 1)

Mutations in the gene *PINK1* were identified in four Italian families with recessive early-onset PD (Valente et al., 2004). Initial screens for *PINK1* mutations in early-onset familial cases reveals that *PINK1* mutations are a more common cause of young-onset PD than *DJ-1* mutation, but not as prevalent as *PRKN* mutation. *PINK1* mutations are estimated to cause 4% of familial recessive PD. The clinical course of individuals with *PINK1* mutations resembles that of sporadic PD except the age of onset is earlier (approximately 35–45 years of age) and disease progression is slower. Similar to *PRKN* disease dystonia at onset appears to be more frequent in individuals with *PINK1* mutations. No pathology data is available from any affected individuals (Healy et al., 2004a).

PINK1 is predicted to be a serine-threonine kinase that is targeted to the mitochondria. Once *PINK1* enters the mitochondria, the *N*-terminal mitochondrial targeting motif is cleaved. Although no substrates of *PINK1* have been identified, the recessive nature of the disease and the presence of truncating mutations in *PINK1*-linked cases, suggest loss of kinase activity may result in cell loss. As *PINK1* is a mitochondrial kinase and can protect cells against oxidative stresses such as paraquat and MPTP (Figure 54.1) (Beilina et al., 2005; Valente et al., 2004). PINK1 may phosphorylate multiple proteins to maintain mitochondrial function and inhibit apoptosis. In support of this observation, knockout of Drosophila *PINK1* results in male sterility, apoptotic muscle degeneration, defects in mitochondrial morphology, and increased sensitivity to oxidative stress.

As mutations within *PINK1* were only recently identified in PD, more work is needed to determine what the endogenous function of *PINK1* is and how mutations within *PINK1* can cause selective degeneration of the SN.

LRRK2 (*PARK8*; DARDARIN)

In 2002, autosomal dominant PD within a large Japanese kindred from Sagamihara was linked to the pericentromeric region of chromosome 12 (Funayama et al., 2002). Affected members of this family presented with a clinically typical *L*-dopa responsive PD with an age at onset of approximately

50 years. Neuropathologically, individuals exhibited pure nigral degeneration in the absence of LB. In addition the disease associated haplotype was carried by many unaffected individuals suggesting incomplete penetrance of the mutation (Funayama et al., 2002).

Further linkage of multiple families to the region confirmed the locus and suggested that the gene could be a common cause of PD. However, the pathology of these families was heterogeneous with some individuals presenting with LB pathology while others presented with *TAU* pathology. In 2004, the gene for PARK8-linked PD was identified, *LRRK2* was originally identified as part of the kinome project and the protein (DARDARIN-the Basque word for tremor) is predicted to be a tyrosine kinase (Paisan-Ruiz et al., 2004; Zimprich et al., 2004).

DARDARIN is a very large protein (2527 amino acids) with multiple functional motifs. DARDARIN primarily belongs to a newly identified family of proteins referred to as ROCO that contains two conserved domains (1) a ROC (Ras in complex proteins) domain that belongs to the Ras GTPase superfamily and (2) a COR domain (*C*-terminal of ROC). In addition DARDARIN contains multiple protein interaction motifs such as WD40, armadillo, and a leucine rich repeat (Jain et al., 2005).

The wild-type function of DARDARIN is unknown but from preliminary functional studies it appears that mutations within DARDARIN alter kinase activity and it is required with the formation of intracellular aggregates as well as with toxicity.

In terms of genetic testing, *LRRK2* is arguably the most important gene linked to PD as one single mutation (G2019S) accounts for 1–2% of sporadic PD and 5–6% of familial European PD (Jain et al., 2005). However, there is considerable variability associated with *LRRK2*-linked disease; this includes not only the pathological variability noted above but also variability in clinical presentation, age at onset and penetrance. Initial estimates of penetrance suggested that carrying a G2019S mutation resulted in an 80% chance of having disease; however, the bias of family based recruitment means that this estimate was high and it appears that the penetrance of the G2019S mutation is closer to 30%. This variability in the disease occurrence, presentation, progression, and endpoint suggests that there are other genetic, environmental or stochastic events that modulate the disease process. Unlike mutations in other genes that cause PD, the frequency with which mutations in *LRRK2* occur affords us the opportunity to investigate these specific modulators of the disease, and one would hope these will also be relevant to idiopathic PD (Singleton, 2005).

As common as *LRRK2* mutations may be, one has to question the utility of clinical genetic testing in PD. Multiple mutations within *LRRK2* (Brice, 2005) and several other genes implicated in PD have been described where the pathogenicity of the variants remain in doubt. Furthermore, identification of a mutation does not necessitate an individual will develop disease, or alter its prognosis or clinical treatment. At present, genetic testing in PD should only be used to confirm a clinical diagnosis of PD (McInerney-Leo et al., 2005).

GENETICS OF SPORADIC PD

Mutations in known genes account for less than 10% of all PD thus research has been aimed at identifying genes associated with typical PD and has focused on the role of common genetic variation in modulating lifetime risk for disease. As many of the genes identified in mendelian forms of PD implicate the mitochondria in disease pathogenesis (Shen and Cookson, 2004), numerous studies have questioned if mutation of the mitochondrial genome (mtDNA) or the various components encoded by the nuclear genome, contribute to PD development or progression.

Many lines of evidence support a role of mitochondrial damage in the pathogenesis of PD (Muqit et al., 2006). Mitochondrial complex I activity is systematically decreased in human PD brains and administration of complex I inhibitors (MPTP and rotenone) to rats, mice, and monkeys recapitulates many aspects of PD, including selective neurodegeneration of the SN and formation of LB type pathology (Betarbet et al., 2000). Recently two studies implicated specific mutation of SN mitochondria as causes for impairment of cellular respiration, specific neuronal vulnerability, and age-dependent risk associated with PD (Bender et al., 2006; Kraytsberg et al., 2006). Amplification of mitochondrial DNA revealed more somatic deletions within SN mitochondria than mitochondria from other brain regions and that deletions in SN mitochondria were higher in PD cases than controls, although this difference did not reach statistical significance. By the age 70, nearly all the SN neurons had elevated levels of mtDNA deletions, implying that these types of deletions might contribute to the age-dependent pathogenic processes seen in PD.

It is feasible that mutation and damage of mitochondria contribute to the preferential vulnerability of SN neurons and disease progression, but it remains unclear if accumulation of mitochondrial mutations is the fundamental pathogenic event in the majority of PD. It is probable that genetic variability at different loci contributes and predisposes some individuals to accumulating higher levels of mitochondrial mutations and damage, thus leading to sufficient neuronal loss and clinical manifestation of disease.

More traditional approaches have identified risk factors for PD. This involves a candidate gene association analysis, where typically a gene is chosen based on its function, expression or genomic position, common variants are assayed within the gene, and the frequency of these variants are compared between cases and controls. The ease and low cost of this approach has resulted in hundreds of candidate gene association studies being published in PD. The well-characterized genes in terms of genetic association with typical PD are likely those encoding α-synuclein and *TAU*.

When the *SNCA* triplication was discovered, not only did it validate many overexpression studies but also asked the question if smaller increases in *SNCA* could increase the risk for sporadic disease. Many studies have attempted to address this question but as with most studies looking at risk factors in complex diseases, they have been largely inconclusive. Even though a polymorphic multi-allelic repeat in the promoter of

SNCA (Rep1) can negatively regulate α-synuclein expression, genetic analysis of this marker has not determined if more subtle increases in α-synuclein expression can increase risk for sporadic PD. Examination of common variability in other genes involved in monogenic forms of PD has failed to reveal a consistent association with sporadic PD (Jain et al., 2005).

Perhaps the most robust genetic association with increased risk for PD comes from analysis of the microtubule associated protein *TAU*. Mutations in this gene cause frontotemporal dementia and Parkinsonism linked to chromosome 17 (FTDP-17) (Hutton et al., 1998). *TAU* forms intraneuronal inclusion referred to neurofibrillary tangles (NFTs) in many diseases pathologically referred to as tauopathies such as AD, PSP and corticobasal degeneration (CBD) (Rademakers et al., 2004). Consequently *TAU* was considered as a candidate gene for these diseases and a specific haplotype (referred to as the H1 haplotype) has been consistently associated with increased risk for PSP and CBD. As a common feature of both PSP and CBD is an initial presentation of parkinsonism, a role for *TAU* in PD has been evaluated. Thus far, numerous studies demonstrate that individuals that are homozygous for the H1 haplotype are approximately 1.5 times at greater risk of developing PD (Healy et al., 2004b).

Although *TAU* and *SNCA* have been extensively investigated, the role of common genetic variability in these genes in risk for PD is still arguable. There are many reasons why elucidating the genetics of typical PD has been problematic; some of these lie in study design, many studies are often underpowered, there may be poor selection of control populations and lack of correction for population stratification or multiple testing. All these problems result in the publication of positive association that is usually followed by failure to replicate papers (Botstein and Risch, 2003). Although statistical power will always be a problem, the advent of the HapMap project (www.hapmap.org) and the availability of technology for genome-wide association studies promise a more complete genetic analysis of PD (Maraganore et al., 2005). These data should reveal common genetic variability underlying disease and, in the absence of association, give a reasonable indication of single common genetic variants not underlying disease (Farrall and Morris, 2005).

Therapeutics

Currently the most effective non-surgical treatment for PD is the administration of *L*-dopa. The metabolic precursor to dopamine was first administered in 1960s to a group of individuals affected with post-encephalitic Parkinsonism (Birkmayer and Hornykiewicz, 1962). Within a short period of time individuals improved from a near catatonic state to relatively normal lives. This drug initially offers a remarkable treatment option to most patients with PD; however, as effective as *L*-dopa is in the early stages of PD, with continued use, approximately 60% of patients develop response fluctuations and dyskinesias. The half life of *L*-dopa and therefore the therapeutic benefit can be improved by the co-administration of inhibitors that break down *L*-dopa to non-useful, potentially harmful metabolites such as 3-MT or dopa quinine (COMT and MAO (monoamine oxidase) inhibitors, Figure 54.2).

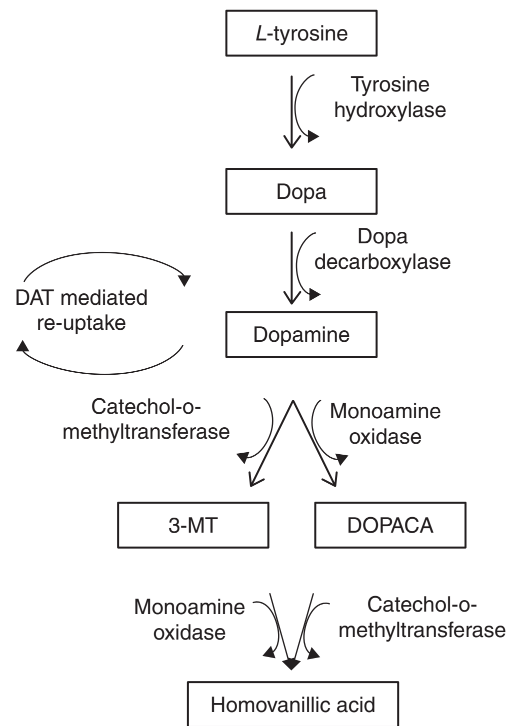


Figure 54.2 Mechanisms of dopamine synthesis and metabolism. 3-MT: 3-methoxytyramine, DOPAC: 3, 4-dihydroxyphenylacetic acid, COMT: Catechol-o-methyltransferase, MAO: Monoamine oxidase.

An alternative to dopamine replacement has been surgical intervention. The loss of dopaminergic cells causes the general deregulation of neurotransmission. The over activation of the subthalamic nucleus (STN) leads to an excessive inhibitory output from the globus pallidus interna (GPi) and SN to the cortex and motor systems (Figure 54.3). As a consequence bradykinesia, tremor, and rigidity arise. The aim of surgical intervention is to disrupt the striatum and other areas to allow increased signaling from the SN. This treatment can be effective at treating *L*-dopa induced dyskinesias, tremor, and rigidity. A more recent advancement in surgical intervention is deep brain stimulation (DBS). This is essentially the same as disrupting the STN and GPi, but as an alternative to ablation, electrodes are implanted to block the signaling via a high-frequency electrical current (Garcia et al., 2005). As a consequence, there is more flexibility in controlling neurotransmission. Current treatments only temporarily manage the symptoms of PD and do not halt or slow down the progression of dopaminergic cell loss.

Since the primary movement disorder associated with PD involves the relative loss of a specific neuronal population, this aspect of the disease represents an excellent target for cell-based therapies. Transplantation of fetal mesencephilic dopaminergic neurons into the STN of PD patients has proven clinical effective with sustained relief from the symptoms (Correia et al., 2005). This therapy has the potential to cure PD if the

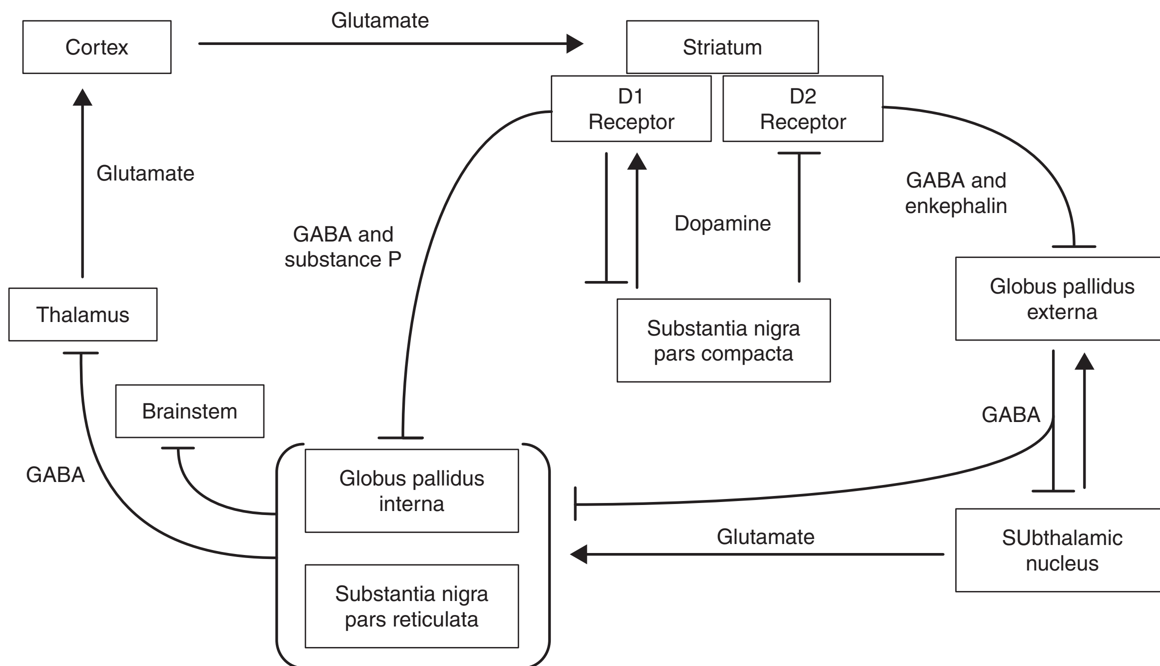


Figure 54.3 Neuronal pathways in the basal ganglia. The overall effect of striatal dopamine release is to reduce basal ganglia output, leading to increased activity of thalamocortical projection neurons. Lack of dopamine results in increased activity of GPi, SN pars reticulata and the STN. This ultimately leads to disruption and inhibition of brain stem motor areas and thalamocortical motor system.

transplanted cells are not afflicted by the same mechanism of neuronal death as the patient's neurons. However, at present a prohibitive number of neurons are required (3–6 fetuses per patient) per patient and current methods of delivery do not lend themselves to treatment of large numbers of patients. Research is underway to grow and differentiate and stem cells into dopamine cells but problems remain. The population of cells produced are not of a pure cell type (i.e. dopaminergic cells) and therefore when implanted into the striatum may actually be detrimental. Studies into creating a pure enriched population of dopamine neurons *ex vivo* and targeting endogenous stem cells for differentiation into dopaminergic neurons are advancing at a significant pace and represent a realistic therapeutic goal for PD.

The Role of Genetics in the Development of Future Therapeutic Strategies

In the near future it will become practical to analyze a person's genome, in a high throughput standardized manner. As a result, identifying genetic variability that alters response to a drug and thus identifying markers that are likely predictors of drug response will gain momentum. Based on this information an individual's drug regime can be tailored to their biological characteristics. This avenue of research should eventually be able to optimize the dose of a drug but also prevent the administration of drugs to individuals who would normally have an adverse potentially lethal reaction. From the opposite side, individuals that respond better to certain drugs may help elucidate biological components of a disease process and identify other potential drug targets. As *L*-dopa is such an effective treatment for PD many of the preliminary pharmacogenomic approaches have aimed to identify biological factors which augment and prolong its result. Accordingly, genes involved in dopamine metabolism have been assessed for their genetic contribution to *L*-dopa efficacy, most notably including analysis of genetic variability within the genes encoding MAO, dopa decarboxylase and COMT among others (Figure 54.2; Maimone et al., 2001). While the initial results of these studies are inconclusive, the application of genome-wide genotyping techniques holds the promise of identifying easily assayable variants that offer predictive value in terms of response to treatment.

The most successful strategies for treating PD will be those that aim to stop or significantly delay the underlying molecular disease processes. Etiology-based treatment will require not only the elucidation of these processes and identification of viable targets but also early diagnosis of the disease. This latter point ensures that protective therapy can be applied before too much irreversible damage has occurred and in the event that the successful therapy is one that slows down but does not halt the pathological process, the therapy can be applied early enough to ensure there is little effect on quality of life for the patient in later years of treatment.

Numerous studies have sought to identify biological or genetic markers that may aid in the pre-symptomatic diagnosis or prognosis of PD sufferer's. As yet no such genetic factors have been identified but ^{18}F-dopa uptake with PET imaging (Khan et al., 2005) and levels of phospho-synuclein have the potential to be used as

pre-symptomatic diagnostic aids (Miller et al., 2004). However, before both can be put into clinical practice, they have to be put to rigorous and extensive testing in large patient cohorts to determine if they are truly diagnostic for disease development and prognosis.

As noted above, it is hoped that highlighting genes involved in the pathogenesis of PD will allow effective modeling of disease in both cell and animal based systems. This in turn will provide a greater understanding of the underlying molecular mechanisms of the disease process and highlight pathways for therapeutic intervention. When these models produce a readily quantifiable endpoint believed to be related to the pathological processes of the disease, they also allow high-throughput screening of molecular libraries of compounds for inhibitors of pathogenic processes. Importantly, the recent discovery of *LRRK2* mutations as a cause of PD has not only provided us with another tool to create model systems of disease, but has also provided clinical researchers with a large patient pool to study disease onset, progression, and response to treatment. Large cohorts of asymptomatic subjects will be relatively easy to identify by assessing siblings and children of patients with G2019S linked disease. This group of subjects affords us the opportunity to not only identify signs and symptoms of disease that may be used as specific early indicators of PD, but also provide a cohort of patients in whom the efficacy of neuroprotective agents can be tested.

2009 UPDATE

Genetic Approaches to PD

A large follow-up study on *LRRK2* (*PARK8*) by the International *LRRK2* Consortium found six pathogenic mutations with frequencies of 1% and 4% in sporadic and hereditary PD, respectively (Healy et al., 2008). This consortium recommended that mutations in *LRRK2* merit testing in patients with hereditary PD.

Genome-wide linkage analysis using a 500K SNP resulted in linkage to chromosome 22 in a large pedigree with Parkinsonian-pyramidal syndrome (PPS). Sequencing of candidate genes revealed a variant (R378G) in *FBX07*, a gene that encodes an F-box protein implicated in protein degradation by the ubitquitin proteosome pathway – a pathway that has been implicated in a variety of neurodegenerative diseases (Shojaee et al., 2008). The role of *FBX07* in neurons is currently not known. Following up on these results, two additional families with early onset, progressive parkinsonism with pyramidal tract signs were assessed for *FBX07*. In one family, a homozygous truncating mutation (A498Stop) was found, whereas in another a variety of compound heterozygous mutations were found. *FBX07* was designated as *PARK15* (Di Fonzo et al., 2009).

Copy number variation (CNV) is becoming a more common finding underlying complex disease. At least two studies reported CNV in α-synuclein (*SCNA*) in sporadic PD (Ahn et al., 2008; Ross et al., 2008). Ross et al. (2008) characterized five families and found that *SCNA* dosage was responsible for the severity of clinical presentation of PD, autonomic dysfunction, and dementia. They screened over 1100 patients and found evidence for *SCNA* multiplication in sporadic PD, suggesting that CNV at this locus should be considered as part of screening for PD.

At least one study investigated the statistical interaction between variants associated with PD. Gao et al. (2008) found a joint effect between polymorphisms in the *FGF20* and *MAOB* genes on PD, suggesting that interactions between *FGF20* and *MAOB* should be considered in investigating PD risk. These genes may interact in regulating dopamine levels in the brain.

The biological interaction of genes may provide another way to define groups of gene variants that may provide greater effect sizes or attributable risk than single gene variants alone. Lesnick et al. (2007) mined large datasets of DNA variations in persons with and without PD and identified SNPs related to axon-guidance pathway genes. The effect sizes and statistical significance for the pathway in predicting disease susceptibility to PD, survival free of PD, and age of onset of PD was greater than any single gene. This approach may have implications not only for studies designed to determine genetic risk for PD but also for the design and analysis of studies of other complex diseases.

Genomic Approaches to PD

An important pilot study was published in 2007 that used gene expression analysis of peripheral blood mononuclear cells to develop "noninvasive" biomarkers of PD (Scherzer et al., 2007). Using a training set of 66 patients and health controls and an independent validation set of 39 patients and controls, a 22 gene model was built that could classify PD (or *de novo* PD) versus a number of other neurodegenerative diseases. A follow-up study of these data showed that a number of neuroimmune signaling pathways were expressed in early PD (Soreq et al., 2008).

Finally, a number of reviews pointed to the potential of neuroproteomics and blood-based proteomic analyses to define novel biomarkers for PD (Goldknopf, 2008; Pienaar et al., 2008). Proteomics provides a "systems approach" that allows for the identification of key proteins and signaling cascades. Examination of disease tissue using this approach can give rise to targets that can be sought in blood that can be used to distinguish PD from other neurodegenerative diseases such as Alzheimers and ALS, provides information that can be used to monitor patients' response to therapy and/or progressive disease, and finally offers the potential to tailor therapies for personalized treatment of PD. (*Update prepared by editors*)

2009 UPDATE REFERENCES

Ahn, T.B., Kim, S.Y., Kim, J.Y., Park, S.S., Lee, D.S., Min, H.J., Kim, Y.K., Kim, S.E., Kim, J.M., Kim, H.J. et al. (2008). Alpha-synuclein gene duplication is present in sporadic Parkinson disease. *Neurology* 70(1),43–49.

Di Fonzo, A., Dekker, M.C., Montagna, P., Baruzzi, A., Yonova, E.H., Correia Guedes, L., Szczerbinska, A., Zhao, T., Dubbel-Hulsman, L.O., Wouters, C.H., et al. FBXO7 mutations cause autosomal recessive, early-onset parkinsonian-pyramidal syndrome. *Neurology* 2008 November 26; [Epub ahead of print].

Gao, X., Scott, W.K., Wang, G., Mayhew, G., Li, Y.J., Vance, J.M. and Martin, E.R. (2008). Gene–gene interaction between FGF20 and MAOB in Parkinson disease. *Ann Hum Genet* 72(Pt 2), 157–162.

Goldknopf, I.L. (2008). Blood-based proteomics for personalized medicine: Examples from neurodegenerative disease. *Expert Rev Proteomics* 5(1), 1–8.

Healy, D.G., Falchi, M., O'Sullivan, S.S., Bonifati, V., Durr, A., Bressman, S., Brice, A., Aasly, J., Zabetian, C.P., Goldwurm, S. et al. (2008). Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: A case-control study. *Lancet Neurol* 7(7), 583–590.

Lesnick, T.G., Papapetropoulos, S., Mash, D.C., Ffrench-Mullen, J., Shehadeh, L., de Andrade, M., Henley, J.R., Rocca, W.A., Ahlskog, J.E. and Maraganore, D.M. (2007). A genomic pathway approach to a complex disease: Axon guidance and Parkinson disease. *PLoS Genet* 3(6), e98.

Pienaar, I.S., Daniels, W.M. and Götz, J. (2008). Neuroproteomics as a promising tool in Parkinson's disease research. *J Neural Transm* 115(10), 1413–1430.

Ross, O.A., Braithwaite, A.T., Skipper, L.M., Kachergus, J., Hulihan, M.M., Middleton, F.A., Nishioka, K., Fuchs, J., Gasser, T., Maraganore, D.M. et al. (2008). Genomic investigation of alpha-synuclein multiplication and parkinsonism. *Ann Neurol* 63(6), 743–750.

Scherzer, C.R., Eklund, A.C., Morse, L.J., Liao, Z., Locascio, J.J., Fefer, D., Schwarzschild, M.A., Schlossmacher, M.G., Hauser, M.A., Vance, J.M. et al. (2007). Molecular markers of early Parkinson's disease based on gene expression in blood. *Proc Natl Acad Sci U S A* 104(3), 955–960.

Shojaee, S., Sina, F., Banihosseini, S.S., Kazemi, M.H., Kalhor, R., Shahidi, G.A., Fakhrai-Rad, H., Ronaghi, M. and Elahi, E. (2008). Genome-wide linkage analysis of a Parkinsonian-pyramidal syndrome pedigree by 500 K SNP arrays. *Am J Hum Genet* 82(6), 1375–1384.

Soreq, L., Israel, Z., Bergman, H. and Soreq, H. (2008). Advanced microarray analysis highlights modified neuro-immune signaling in nucleated blood cells from Parkinson's disease patients. *J Neuroimmunol* 201–202, 227–236.

REFERENCES

Abou-Sleiman, P.M., Healy, D.G., Quinn, N., Lees, A.J. and Wood, N.W. (2003). The role of pathogenic DJ-1 mutations in Parkinson's disease. *Ann Neurol* 54, 283–286.

Beilina, A., Van Der Brug, M., Ahmad, R., Kesavapany, S., Miller, D.W., Petsko, G.A. and Cookson, M.R. (2005). Mutations in PTEN-induced putative kinase 1 associated with recessive parkinsonism have differential effects on protein stability. *Proc Natl Acad Sci USA* 102, 5703–5708.

Bender, A., Krishnan, K.J., Morris, C.M., Taylor, G.A., Reeve, A.K., Perry, R.H., Jaros, E., Hersheson, J.S., Betts, J., Klopstock, T. et al. (2006). High levels of mitochondrial DNA deletions in substantia nigra neurons in aging and Parkinson disease. *Nat Genet* 38, 515–517.

Betarbet, R., Sherer, T.B., MacKenzie, G., Garcia-Osuna, M., Panov, A.V. and Greenamyre, J.T. (2000). Chronic systemic pesticide exposure reproduces features of Parkinson's disease. *Nat Neurosci* 3, 1301–1306.

Birkmayer, W. and Hornykiewicz, O. (1962). The *L*-dihydroxyphenylalanine (*L*-DOPA) effect in Parkinson's syndrome in man: On the pathogenesis and treatment of Parkinson akinesis.. *Arch Psychiatr Nervenkr Z Gesamte Neurol Psychiatr* 203, 560–574.

Bonifati, V., Rizzu, P., Squitieri, F., Krieger, E., Vanacore, N., van Swieten, J.C., Brice, A., van Duijn, C.M., Oostra, B., Meco, G. et al. (2003). DJ-1(PARK7), a novel gene for autosomal recessive, early onset parkinsonism. *Neurol Sci* 24, 159–160.

Botstein, D. and Risch, N. (2003). Discovering genotypes underlying human phenotypes: Past successes for mendelian disease, future approaches for complex disease. *Nat Genet* 33, 228–237.

Brice, A. (2005). Genetics of Parkinson's disease: LRRK2 on the rise. *Brain* 128, 2760–2762.

Canet-Aviles, R.M., Wilson, M.A., Miller, D.W., Ahmad, R., McLendon, C., Bandyopadhyay, S., Baptista, M.J., Ringe, D., Petsko, G.A. and Cookson, M.R. (2004). The Parkinson's disease protein DJ-1 is neuroprotective due to cysteine-sulfinic acid-driven mitochondrial localization. *Proc Natl Acad Sci USA* 101, 9103–9108.

Chen, L. and Feany, M.B. (2005). Alpha-synuclein phosphorylation controls neurotoxicity and inclusion formation in a Drosophila model of Parkinson disease. *Nat Neurosci* 8, 657–663.

Cookson, M.R. (2005). The biochemistry of Parkinson's disease. *Annu Rev Biochem* 74, 29–52.

Correia, A.S., Anisimov, S.V., Li, J.Y. and Brundin, P. (2005). Stem cell-based therapy for Parkinson's disease. *Ann Med* 37, 487–498.

Farrall, M. and Morris, A.P. (2005). Gearing up for genome-wide gene-association studies. *Hum Mol Genet* 14(Spec No. 2), R157–R162.

Funayama, M., Hasegawa, K., Kowa, H., Saito, M., Tsuji, S. and Obata, F. (2002). A new locus for Parkinson's disease (PARK8) maps to chromosome 12p11.2-q13.1. *Ann Neurol* 51, 296–301.

Garcia, L., D'Alessandro, G., Bioulac, B. and Hammond, C. (2005). High-frequency stimulation in Parkinson's disease: More or less? *Trends Neurosci* 28, 209–216.

Hardy, J. and Lees, A.J. (2005). Parkinson's disease: A broken nosology. *Mov Disord* 20(Suppl 12), S2–S4.

Healy, D.G., Abou-Sleiman, P.M., Gibson, J.M., Ross, O.A., Jain, S., Gandhi, S., Gosal, D., Muqit, M.M., Wood, N.W. and Lynch, T. (2004a). PINK1 (PARK6) associated Parkinson disease in Ireland. *Neurology* 63, 1486–1488.

Healy, D.G., Abou-Sleiman, P.M., Lees, A.J., Casas, J.P., Quinn, N., Bhatia, K., Hingorani, A.D. and Wood, N.W. (2004b). Tau gene and Parkinson's disease: A case-control study and meta-analysis. *J Neurol Neurosurg Psychiatr* 75, 962–965.

Hughes, A.J., Daniel, S.E. and Lees, A.J. (1993). The clinical features of Parkinson's disease in 100 histologically proven cases. *Adv Neurol* 60, 595–599.

Hutton, M., Lendon, C.L., Rizzu, P., Baker, M., Froelich, S., Houlden, H., Pickering-Brown, S., Chakraverty, S., Isaacs, A., Grover, A. et al. (1998). Association of missense and 5′-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature* 393, 702–705.

Jain, S., Wood, N.W. and Healy, D.G. (2005). Molecular genetic pathways in Parkinson's disease: A review. *Clin Sci (Lond)* 109, 355–364.

Kahle, P.J. and Haass, C. (2004). How does parkin ligate ubiquitin to Parkinson's disease?. *EMBO Rep* 5, 681–685.

Kaplan, B., Ratner, V. and Haas, E. (2003). Alpha-synuclein: Its biological function and role in neurodegenerative diseases. *J Mol Neurosci* 20, 83–92.

Khan, N.L., Jain, S., Lynch, J.M., Pavese, N., Abou-Sleiman, P., Holton, J.L., Healy, D.G., Gilks, W.P., Sweeney, M.G., Ganguly, M. et al. (2005). Mutations in the gene LRRK2 encoding dardarin (PARK8) cause familial Parkinson's disease: Clinical, pathological, olfactory and functional imaging and genetic data. *Brain* 128, 2786–2796.

Kitada, T., Asakawa, S., Hattori, N., Matsumine, H., Yamamura, Y., Minoshima, S., Yokochi, M., Mizuno, Y. and Shimizu, N. (1998). Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature* 392, 605–608.

Kraytsberg, Y., Kudryavtseva, E., McKee, A.C., Geula, C., Kowall, N.W. and Khrapko, K. (2006). Mitochondrial DNA deletions are abundant and cause functional impairment in aged human substantia nigra neurons. *Nat Genet* 38, 518–520.

Kruger, R., Kuhn, W., Leenders, K.L., Sprengelmeyer, R., Muller, T., Woitalla, D., Portman, A.T., Maguire, R.P., Veenma, L., Schroder, U. et al. (2001). Familial parkinsonism with synuclein pathology: Clinical and PET studies of A30P mutation carriers. *Neurology* 56, 1355–1362.

Lohmann, E., Periquet, M., Bonifati, V., Wood, N.W., De Michele, G., Bonnet, A.M., Fraix, V., Broussolle, E., Horstink, M.W., Vidailhet, M. et al. (2003). How much phenotypic variation can be attributed to parkin genotype?. *Ann Neurol* 54, 176–185.

Maimone, D., Dominici, R. and Grimaldi, L.M. (2001). Pharmacogenomics of neurodegenerative diseases. *Eur J Pharmacol* 413, 11–29.

Maraganore, D.M., de Andrade, M., Lesnick, T.G., Strain, K.J., Farrer, M.J., Rocca, W.A., Pant, P.V., Frazer, K.A., Cox, D.R. and Ballinger, D.G. (2005). High-resolution whole-genome association study of Parkinson disease. *Am J Hum Genet* 77, 685–693.

McInerney-Leo, A., Hadley, D.W., Gwinn-Hardy, K. and Hardy, J. (2005). Genetic testing in Parkinson's disease. *Mov Disord* 20, 1–10.

McNaught, K.S. and Olanow, C.W. (2006). Proteasome inhibitor-induced model of Parkinson's disease. *Ann Neurol* 60, 243–247.

Miller, D.W., Hague, S.M., Clarimon, J., Baptista, M., Gwinn-Hardy, K., Cookson, M.R. and Singleton, A.B. (2004). Alpha-synuclein in blood and brain from familial Parkinson disease with SNCA locus triplication. *Neurology* 62, 1835–1838.

Mukaetova-Ladinska, E.B. and McKeith, I.G. (2006). Pathophysiology of synuclein aggregation in Lewy body disease. *Mechanisms of Ageing and Development Dementias and Cognitive Disorders: New Insights and Approaches* 127, 188–202.

Muqit, M.M., Gandhi, S. and Wood, N.W. (2006). Mitochondria in Parkinson disease: Back in fashion with a little help from genetics. *Arch Neurol* 63, 649–654.

Paisan-Ruiz, C., Jain, S., Evans, E.W., Gilks, W.P., Simon, J., van der Brug, M., Lopez de Munain, A., Aparicio, S., Gil, A.M., Khan, N. et al. (2004). Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. *Neuron* 44, 595–600.

Piccini, P., Morrish, P.K., Turjanski, N., Sawle, G.V., Burn, D.J., Weeks, R.A., Mark, M.H., Maraganore, D.M., Lees, A.J. and Brooks, D.J. (1997). Dopaminergic function in familial Parkinson's disease: A clinical and 18F-dopa positron emission tomography study. *Ann Neurol* 41, 222–229.

Polymeropoulos, M.H., Lavedan, C., Leroy, E., Ide, S.E., Dehejia, A., Dutra, A., Pike, B., Root, H., Rubenstein, J., Boyer, R. et al. (1997). Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science* 276, 2045–2047.

Rademakers, R., Cruts, M. and van Broeckhoven, C. (2004). The role of tau (MAPT) in frontotemporal dementia and related tauopathies. *Hum Mutat* 24, 277–295.

Shen, J. and Cookson, M.R. (2004). Mitochondria and dopamine: New insights into recessive parkinsonism. *Neuron* 43, 301–304.

Singleton, A.B., Farrer, M., Johnson, J., Singleton, A., Hague, S., Kachergus, J., Hulihan, M., Peuralinna, T., Dutra, A., Nussbaum, R. et al. (2003). alpha-Synuclein locus triplication causes Parkinson's disease. *Science* 302, 841.

Singleton, A.B. (2005). Altered alpha-synuclein homeostasis causing Parkinson's disease: The potential roles of dardarin. *Trends Neurosci* 28, 416–421.

Spillantini, M.G., Schmidt, M.L., Lee, V.M., Trojanowski, J.Q., Jakes, R. and Goedert, M. (1997). Alpha-synuclein in Lewy bodies. *Nature* 388, 839–840.

Twelves, D., Perkins, K.S. and Counsell, C. (2003). Systematic review of incidence studies of Parkinson's disease. *Mov Disord* 18, 19–31.

Valente, E.M., Abou-Sleiman, P.M., Caputo, V., Muqit, M.M., Harvey, K., Gispert, S., Ali, Z., Del Turco, D., Bentivoglio, A.R., Healy, D.G. et al. (2004). Hereditary early-onset Parkinson's disease caused by mutations in PINK1. *Science* 304, 1158–1160.

Volles, M.J. and Lansbury, P.T., Jr. (2003). Zeroing in on the pathogenic form of alpha-synuclein and its mechanism of neurotoxicity in Parkinson's disease. *Biochemistry* 42, 7871–7878.

von Coelln, R., Dawson, V.L. and Dawson, T.M. (2004). Parkin-associated Parkinson's disease. *Cell Tissue Res* 318, 175–184.

Zimprich, A., Biskup, S., Leitner, P., Lichtner, P., Farrer, M., Lincoln, S., Kachergus, J., Hulihan, M., Uitti, R.J., Calne, D.B. et al. (2004). Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron* 44, 601–607.

RECOMMENDED RESOURCES

1. http://www.alzforum.org/ – A website primarily for current research in Alzheimer's disease, it also discusses the most important current findings in many other neurological disorders including Parkinson's disease
2. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=omim – A link to the Online Mendelian Inheritance in Man based on the NCBI website. This website gives brief descriptions of various diseases including Parkinson's disease (#168600). From this, one can access all genetic factors associated with a specific disease.
3. Healy, D.G., Abou-Sleiman, P.M. and Wood, N.W. (2004c). PINK, PANK, or PARK? A clinicians' guide to familial parkinsonism. *Lancet Neurol*, 3, 652–662 – An excellent review on the impact of genetics on the clinical diagnosis of Parkinson's disease.

CHAPTER

Genomic Considerations in Ophthalmology

Janey L. Wiggs

INTRODUCTION

The function of the eye is to transduce light into an electrical signal and then transmit the electrical signal to the brain. A variety of tissues and specialized cells carry out these complex processes. The ocular globe is divided into two fluid-filled compartments, called the anterior and posterior chambers (Figure 55.1). The anterior chamber is filled with an aqueous fluid called aqueous humor and the posterior chamber is filled with a viscous substance called the vitreous humor. The globe is supported by a tough outer shell, the sclera, that also supports the optic nerve as it exits the eye.

The cornea is a transparent tissue located on the anterior ocular surface that allows light to enter the eye and also helps focus the light on the retina. Inside the eye are a number of structures including the iris and pupil (regulates the amount of light entering the eye), the lens (focuses light on the retina), the ciliary body (makes aqueous humor) and trabecular meshwork (drains aqueous humor). Under normal circumstances the rate of production of aqueous humor equals the rate of removal.

Light traveling through the cornea, pupil and lens is focused on the retina, which carries out the phototransduction of light to produce an electrical signal that is transmitted through the optic nerve to the brain. The retina is a complex tissue made up of 10 distinct layers (Figure 55.2). The most external cell layer is the retinal pigment epithelium, which provides metabolic support and is attached to a basement membrane (Bruch's membrane). Next to the retinal pigment epithelium are the rod and cone photoreceptors, which are the cells where phototransduction occurs. Connected to the photoreceptors are the amacrine, bipolar and horizontal cells that modulate the signal output from the rods (dim light) and cones (bright light). The signal from the photoreceptor goes through the bipolar cells

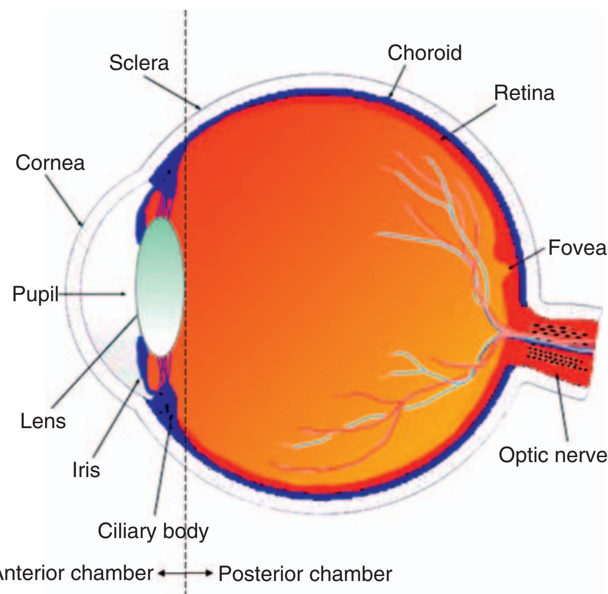


Figure 55.1 Vertical sagittal section of the adult human eye.

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

Schematic diagram of the human retina.
Abbreviations: PE: pigment epithelium; OS: outer segments; IS: inner segment; ONL: outer nuclear layer; OPL: outer plexiform layer; INL: inner nuclear layer; IPL: inner plexiform layer; GCL: ganglion cell layer; NFL: nerve fiber layer.

ROBO3 (Chan et al., 2006) and *HOXA1* (Tischfield et al., 2005). A number of chromosomal syndromes are also associated with strabismus, including Rubenstein–Taybi syndrome (Allanson and Hennekam, 1997), Cornelia de Lange syndrome (Nallasamy et al., 2006) and Down's syndrome (Yurdakul et al., 2006). Screening patients with eye movement disorders for mutations in these genes will help define the diagnosis.

Duane's syndrome is a congenital eye movement disorder characterized by abnormalities of horizontal eye movements, and indicates a problem with cranial nerve VI, the abducens nerve (Bagheri and Repka, 2005). The syndrome can be inherited as an autosomal dominant trait, and using three affected pedigrees a genome-wide scan identified a locus for the syndrome on chromosome 20q13 (Al-Baradie et al., 2002). A novel gene, *LOC57167*, subsequently identified as a new member of the SAL family (*SALL4*) was identified as the causative gene. *SALL4* is a zinc finger protein that shares significant homology with *SALL1* coding for a transcriptional repressor and responsible for a related developmental disorder Towne–Brocke syndrome (Botzenhart et al., 2005). *SALL4* mutations causing Duane's syndrome mainly result in truncated protein products and cause a loss-of-function of the protein (Terhal et al., 2006). The Duane's phenotype suggests that this protein is likely to have an important role in abducens motor neuron development.

A related condition, congenital fibrosis of the extraocular muscles type 1 (CFEOM1) is associated with absence of the superior division of the oculomotor nerve (cranial nerve III) and corresponding midbrain motoneurons and profound atrophy of the two muscles normally innervated by this nerve, the levator palpebrae superioris and superior rectus (Traboulsi, 2004). These two muscles elevate the eyelid and the globe, and their dysfunction accounts for the primary features of the CFEOM1 phenotype. Mutations in KIF21A, a kinesin, have been identified in patients affected with this condition (Yamada et al., 2003). Kinesins are molecular motors responsible for microtubule-dependent transport of cargo; in neurons, they are responsible for anterograde and retrograde axonal transport. CFEOM1 probably results from the inability of mutated KIF21A to successfully deliver a cargo essential to the development of the oculomotor axons, neuromuscular junction or extraocular muscles.

CORNEA

The function of the cornea is dependent on the transparency of the tissue. Any process that causes opacification of the cornea results in a decrease in vision. The cornea can lose its transparency as a consequence of infection (both bacterial and viral), trauma (scar tissue formation) and inherited corneal dystrophies. Genetic causes of corneal disease can be inherited as autosomal dominant and autosomal recessive traits and are caused by mutations in at least nine genes (*ARSC1*, *CHST6*, *COL8A2*, *GLA*, *GSN*, *KRT3*, *KRT12*, *M1S1* and *TGFBI* [*BIGH3*]). Three of these genes (*GSN*, *M1S1*, *TGFBI*) are associated with amyloid deposition in the cornea (review see Klintworth, 2003).

TABLE 55.1 Selected ocular disorders and associated genes discussed in this chapter

Ocular tissue	Specific disorder	Inheritance	Gene – function
Ocular motor (eye movement)	Duane's syndrome	AD	*SALL4* – zinc finger protein; transcriptional regulator
	CFEOM1 (congenital fibrosis of extraocular eye muscles type 1)	AD	*KIF21A* – kinesin; axonal transport
Cornea (dystrophy)	Groenouw's (granular) dystrophy	AD	*TGFB1/BIGH3* – keratoepithelin; extracellular matrix protein
	Lattice type 1 dystrophy	AD	
	Avellino's (combined granular-lattice) dystrophy	AD	
	ReisBückler's dystrophy	AD	
Lens (cataract)	Pulverulent cataract	AD	Alpha3 (C × 46) and Alpha8 (C × 50) – connexins; gap junctions
	Zonular cataract Posterior polar cataract	AD	Alpha A- and alpha B-crystallins – lens proteins
Iris	Abnormal development: Peter's Anomaly Corneal keratitis Aniridia	AD	*PAX6* – transcriptional regulator; gene expression
Trabecular meshwork	Early-onset glaucoma	AD	*MYOC* – Myocilin; extracelluar matrix
Optic nerve	Leber's Hereditary Optic neuropathy	Mitochondrial	Missense mutations in complex I of the respiratory chain (mitochondrial DNA)
	Kjer's optic neuropathy	AD	*OPA1* – dynamin-related GTPase
Retina	Retinitis pigmentosa	AD*	*RHO* – rhodopsin; rod photoreceptor pigment
	Leber's congenital amaurosis	AR	*RPE65* – enzyme responsible for regeneration of 11-cis retinal
	Retinoblastoma	AD	*RB1* – tumor suppressor protein
	Age-related macular degeneration (AMD)	Complex	*CFH* – complement factor H; *LOC387715/HTRA1* – unknown function

*Retinitis pigmentosa may also exhibit autosomal recessive, X-linked and digenic inheritance patterns.

Screening patients with inherited corneal opacities for mutations in this panel of genes will help define the diagnosis and the prognosis. Because some of these disorders are inherited as dominant traits and others as recessive traits, defining the causative gene in an affected individual will also define the disease risk in other family members.

The four most common autosomal dominant corneal dystrophies are: Groenouw's (granular) type 1 (Moller, 1989), lattice type 1 (Klintworth, 1967), Avellino's (combined granular-lattice) (Folberg et al., 1988; Rosenwasser et al., 1993), and Reis–Bückler's (Kuchle et al., 1995). Although all four of these corneal dystrophies affect the anterior stromal layer of the cornea and cause the formation of discrete white localized deposits that progressively obscure vision, the detailed clinical and pathologic features differ. All four dystrophies had been genetically mapped to a common interval on chromosome 5q31 (Eiberg et al., 1994; Gregory et al., 1995; Small et al., 1996; Stone et al., 1994), and mutations in a single gene, *TGFB1/BIGH3*, were subsequently identified in individuals affected with one of the four conditions (Munier et al., 1997). An abnormal protein product of this gene, keratoepithelin, accumulates in patients carrying mutations. The normal protein product is probably an extracellular matrix protein that modulates cell adhesion. Four different missense mutations occurring at two different arginine codons, R124 and R555 have been found. Interestingly, different mutations at the codon R124 cause lattice dystrophy type I or Avellino's dystrophy, the two dystrophies characterized by amyloid deposits (Korvatska et al., 1998). The mutations

that cause Avellino's and lattice dystrophies abolish a putative phosphorylation site that is probably required for the normal structure of keratoepithelin. Destruction of this aspect of the protein structure leads to the formation of the amyloid deposits that cause opacification of the cornea. As a result, the mutant protein is destructive to the normal tissue. Mutations at the R555 appear to result in either granular dystrophy or Reis–Bücklers dystrophy. These phenotype–genotype correlations demonstrate the variable expressivity of mutations in this gene and the significance of alteration of the arginine residues 124 and 555. Of interest, pathologic deposits caused by keratoepithelin accumulation have only been observed in the cornea and not in other tissues or organs (El Kochairi et al., 2006). Because the *TGFB1/BIGH3* gene is expressed in other tissues (Sandgren et al., 1990), these results suggest a cornea-specific mechanism causing the accumulation of mutant keratoepithelin.

LENS

The ocular lens is transparent and focuses the light coming into the eye from the cornea and pupil onto the retina. Cataract is an opacification of the lens that causes a loss of vision and can develop congenitally during childhood and most commonly as an adult (age-related cataract). Congenital cataract is a leading cause of visual disability in children and recently many causative genetic mutations have been identified. Inherited cataract is clinically and genetically heterogeneous with at least 11 different defined phenotypes. Cataracts may be inherited as autosomal dominant, autosomal recessive, or X-linked recessive traits, and at least 12 loci and 15 specific genes are currently identified, with more genes remaining to be discovered (for reviews see Hejtmancik and Kantorow, 2004; Reddy et al., 2004). Two general classes of proteins, the crystallins and the connexins have been associated with early onset cataracts. Screening for mutations in the causative genes will help define the diagnosis in affected individuals. Some forms of early onset cataract severely interfere with vision while others do not. Because of this spectrum of disease severity, in addition to clarification of the diagnosis, genetic screening could help establish a prognosis and also cataract-risk in family members.

Intercellular gap junction channels, consisting of alpha3 (C × 46), alpha8 (C × 50) and alpha6 (C × 43) connexin subunits have been implicated in lens development and maintenance. These channels probably transport metabolites, secondary messages and ions between lens cells (Krutovskikh and Yamasaki, 2000; Xia et al., 2006). Missense changes in these genes have been associated with cataract phenotypes, with changes in the alpha3 (C × 46) and alpha8 (C × 50) associated with a central pulverulent cataract than can have a better prognosis for vision (Berry et al., 1999; Li et al., 2004).

Alpha A- and alpha B-crystallins are the major components of the lens. These proteins are also members of the small heat-shock protein family, and they possess chaperone-like function. Interestingly, mutations in these genes can cause cataract and also various forms of myopathy (Horwitz, 2003). For example, missense and frame-shift mutations in alpha B-crystallin cause autosomal dominant congenital lamellar cataract (Liu et al., 2006), while other missense mutations have been associated with dilated cardiomyopathy (Inagaki et al., 2006). The underlying mechanism appears to be a dominant negative effect that interferes with the normal chaperone-like function of these proteins.

IRIS

The iris contains the pupil, which functions as an aperture regulating the amount of light entering the eye. Disorders causing iris and pupil dysfunction are typically ocular developmental syndromes that lead to abnormal formation of the iris and pupil. These developmental abnormalities can lead to an elevation of intraocular pressure and glaucoma (see below) and also cataract. Most of the genes responsible for these conditions are transcriptional regulatory factors and other proteins that play regulatory roles in ocular developmental processes including *PAX6*, *FOXC1*, *PITX2* and *LMX1B* (for review see Gould et al., 2004; Idrees et al., 2006). Most of the developmental syndromes resulting from mutations in these genes are inherited as autosomal dominant traits. Mutations in these genes are associated with variable expressivity, and there is extensive overlap between the phenotypes associated with the different causative genes. Screening a panel of these genes for mutations in affected patients will define the diagnosis and help identify other family members at risk.

The *PAX6* gene is a transcription factor essential for the development of tissues including the eyes, central nervous system and endocrine glands of vertebrates and invertebrates. It regulates the expression of a broad range of molecules, including other transcription factors, cell adhesion and short-range cell–cell signaling molecules, hormones and structural proteins. It has been implicated in a number of key biological processes including other cell proliferation, migration, adhesion and signaling both in normal development and in oncogenesis (Simpson and Price, 2002). Mutations in *PAX6* are associated with a range of human phenotypes including aniridia (absence of the iris) (Tzoulaki et al., 2005), Peter's anomaly (Singh et al., 1998) and corneal keratitis (Sale et al., 2002). Most mutations cause a truncated polypeptide or disruption of the critical paired box homeodomain (van Heyningen and Williamson, 2002).

TRABECULAR MESHWORK

The trabecular meshwork and related outflow pathways remove aqueous humor from the anterior chamber of the eye. The intraocular pressure is dependent on the rate of fluid removal by the trabecular meshwork which under normal conditions matches the rate of formation. The intraocular pressure can become elevated when the trabecular meshwork no longer

keeps pace with the rate of fluid formation and an elevation of intraocular pressure is a major risk factor for glaucoma (Gordon et al., 2002). Glaucoma is a common blinding disease that results in an irreversible degeneration of the optic nerve (see below). Despite the important function of the trabecular meshwork, very little is known about the molecular pathways involved in fluid removal. A number of genes have been associated with trabecular meshwork dysfunction and elevated intraocular pressure including those that cause abnormal development of the iris and anterior segment (see above and Wiggs, 2005 for review).

The degeneration of the optic nerve associated with high intraocular pressure is an insidious process and significant vision can be irreversibly lost before the patient is aware of the symptoms. Screening for risk factors, including genetic risk factors, is necessary to identify individuals at risk so that pressure-lowering treatment can be initiated before irreversible damage to the optic nerve occurs. Although a major genetic risk factor for glaucoma or elevated intraocular pressure has yet to be discovered, genes that can confer risk in a small number of patients have been described. Missense mutations in the gene coding for myocilin (*MYOC*) are associated with an early onset of elevated intraocular pressure and severe glaucoma (Fingert et al., 2002). Some mutations cause more significant disease than others and probably cause a gain-of-function resulting in retention of the mutant protein in the cell and subsequent cell death (Liu and Vollrath, 2004). Mutations in another gene, *WDR36* are insufficient to cause glaucoma, but patients who have glaucoma and also have changes in this gene have a more severe phenotype (Hauser et al., 2006). The function of the WDR36 protein is not known but may influence immune mechanisms suggesting a possible role for immune response in elevated intraocular pressure (Monemi et al., 2005).

OPTIC NERVE

The optic nerve is the only path for communication between the eye and the brain. When the optic nerve is damaged, irreversible blindness occurs. The optic nerve contains the axons from the retinal ganglion cells that travel through a supporting structure, the lamina cribrosa, before reaching their first synapse in the lateral geniculate body. Inherited disorders of the optic nerve include degenerative processes (primarily glaucoma described above) as well as primary disorders causing optic nerve atrophy (for review see Newman, 2005). Mitochondrial function is a critical element in optic nerve disease: Leber's hereditary optic neuropathy is caused by missense mutations in mitochondrial DNA (Valentino et al., 2004), while Kjer's autosomal dominant optic atrophy is caused by mutations in *OPA1* that also affect mitochondrial function (Olichon et al., 2006). Screening for mutations in genes known to contribute to optic nerve disease can identify individuals at risk, but therapeutic options are currently limited (Johns and Colby, 2002).

RETINA

The phototransduction of light is carried out by the retina. Phototransduction occurs in the rod and cone photoreceptors and is dependent on rhodopsin and the cone opsins. The photoreceptors signal a cascade of retinal neural cells with the terminal axons of the retinal ganglion cells forming the optic nerve. Inherited blinding disorders that affect the structure and physiology of retinal cells participating in this complex process include: retinitis pigmentosa and other retinal degenerations (for review see Kennan et al., 2005); retinoblastoma (for review see Knudson, 2001); and age-related macular degeneration (for review see Tuo et al., 2004; Wiggs, 2006). The role of selected genetic factors in these retinal disorders will be discussed below.

The first gene to be recognized as a cause of retinitis pigmentosa was the gene coding for rhodopsin, the primary pigment in rod photoreceptors. Missense mutations in rhodopsin cause an autosomal dominant form of retinitis pigmentosa (Dryja et al., 1991). In general the rhodopsin mutations associated with retinitis pigmentosa can be placed into two groups: those that affect rhodopsin synthesis, folding, or transport from the rod cell, and those that have detrimental effects on rhodopsin's functions, such as photobleaching, photoactivation and deactivation (Kisselev, 2005). P23H, is the most common mutation associated with autosomal dominant retinitis pigmentosa, and this mutation causes protein misfolding or missorting in the endoplasmic reticulum (Olsson et al., 1992). Missense mutations that affect rhodopsin function, such as K296E cause congenital night blindness in addition to autosomal dominant retinitis pigmentosa (Robinson et al., 1992). Although definitive treatment for these conditions is not currently possible, screening for mutations will help identify individuals at risk.

Leber's congenital amaurosis is a severe blinding condition that leaves infants and children without any useful vision. The most common cause of this condition are mutations in *RPE65* coding for a protein located in the retinal pigment epithelium that is a critical enzyme responsible for regeneration of 11-cis retinal, the chromophore needed for visual pigments (Galvin et al., 2005; Takahashi et al., 2005). Mutations in the gene cause a loss of function of the protein, and restoration of vision has been accomplished in both a canine model and a mouse model of the disease using an rAAV vector containing the normal *RPE65* gene (Acland et al., 2005; Chen et al., 2006). Clinical trials are underway to evaluate the efficacy of injecting *RPE65* viral constructs into human eyes.

Retinoblastoma is the most common primary intraocular malignancy of childhood. Predisposition to the disease can be inherited as an autosomal dominant trait, and the mean age of diagnosis for inherited cases is 12 months of age. Affected children can present with a visible white reflex at the pupil (leukokoria) but tumors can also be located in the retina such that they are not detected by a typical clinical exam. If the tumors are not treated, nearly all patients die of intracranial extension and disseminated disease within 2 years. Careful examination of the retina is

difficult in young children, and usually requires general anesthesia. Children at risk must be examined every 3 months until they have reached an age where tumor development is unlikely (about age 5). The discovery of the *RB1* gene as the causative gene for retinoblastoma made it possible to develop molecular diagnostic tests to detect carriers of mutant forms of the gene (Wiggs et al., 1988). Current methods include the protein truncation test (Tsai et al., 2004) and sequence detection strategies (Houdayer et al., 2004). Individuals identified as carriers of mutant forms of the gene can undergo increased surveillance and timely treatment which involves removal of the eye or treatment of the tumor with proton beam radiation (Abramson and Schefler, 2004).

Age-related macular degeneration (AMD) affects over 10 million Americans and is the leading cause of blindness among the elderly. AMD is a complex disease that results from interactions between genetic and environmental factors. Studies have shown that the risk of macular degeneration increases with age, smoking and excess dietary lipids (Hyman and Neborsky, 2002). Significant genetic contributions have been demonstrated by increased concordance between identical twins, familial clustering and increased risk to first-degree relatives (Stone et al., 2001). Recently variants of two genes, complement factor H (Edwards et al., 2005; Haines et al., 2005; Klein et al., 2005) and a novel gene *LOC387715* (Rivera et al., 2005) have been shown to substantially increase the risk of macular degeneration. These risks are further increased in individuals who smoke (Schmidt et al., 2006). These important results suggest that screening for these genetic factors can identify individuals at risk for the disease and identify individuals who can reduce their risk by behavior modification.

GENETIC TESTING FOR OCULAR DISORDERS

The identification of genes responsible for inherited eye disease and the development of powerful molecular techniques for detecting genetic defects makes it likely that genetic testing for inherited ocular disorders will become commonplace over the next decade. Current genetic tests are available for many ocular disorders, including those discussed in this chapter. A number of CLIA-certified laboratories perform genetic testing for eye diseases in the United States (Stone, 2007), and a group of CLIA laboratories are participating in a genotype/phenotype project sponsored by the National Eye Institute (eyeGENE; http://www.nei.nih.gov). Current tests mainly provide diagnostic information, however, future testing is expected to provide prognostic information. With the additional development of pharmacogenomic tests predicting therapeutic responsiveness, gene-based tests could become an increasingly important component of an individualized treatment plan.

Genetic testing is currently most useful for the numerous rare Mendelian ocular disorders. Recent advances have identified genetic risk factors for two common complex ocular conditions, macular degeneration and one form of glaucoma, pseudoexfoliation syndrome (Edwards et al., 2005; Fan et al., 2007; Haines et al., 2005; Klein et al., 2005; Rivera et al., 2005; Thorleifsson et al., 2007). As more genetic and environmental risk factors for complex diseases are identified, genetic testing will become an important part of establishing risk estimates for these common blinding conditions.

2009 UPDATE

A recent genome-wide association study using Icelandic glaucoma patients and controls demonstrated a very significant association between SNPs in the *LOXL1* gene and pseudoexfoliation glaucoma – one type of age-related, late-onset glaucoma (Thorleifsson et al., 2007). The association in the Icelandic patients has been confirmed in more heterogeneous populations from the United States (Challa et al., 2008; Fan et al., 2008b; Fingert et al., 2007; Yang et al., 2008) and other populations throughout the world, including Australia (Hewitt et al., 2007), Germany (Pasutto et al., 2008), Italy (Pasutto et al., 2008), China (Gong et al., 2008), Japan (Ozaki et al., 2008), and India (Ramprasad et al., 2008). In all the populations studied, the risk genotype at rs3825942 (G153D) accounts for over 90% of the population attributable risk. However, use of this genotype for screening purposes is not yet practical, as the risk genotype is also prevalent in unaffected individuals, significantly lowering the specificity of the test. The high frequency of the risk genotype in controls also suggests that additional factors, both genetic and environmental, are likely to contribute to this disease.

Pseudoexfoliation glaucoma is a specific type of glaucoma characterized by the deposition of microfibrillar material throughout the eye. The composition of the PXFS-related material, although not completely defined, appears to be a complex glycoprotein structure containing elements of basement membranes and the elastic fiber system (Martone et al., 2007). The biological processes that cause this material to accumulate in ocular structures are not known, although recent studies indicate that vascular compromise may be a contributing factor (Parodi et al., 2008). The product of the *LOXL1* gene, *lysyl oxidase like 1*, is a member of the lysyl oxidase family of proteins that catalyze the polymerization of tropoelastin to form the mature elastin polymer. Mice lacking *LOXL1* have abnormal elastic tissues including the basement membranes of blood vessels (Liu et al., 2004). Several studies have suggested that pseudoexfoliation patients have modest elevations of serum homocysteine, which could add further vasculature insult in affected tissues (Altinta et al., 2005; Roedl et al., 2007). Genes coding for proteins that regulate homocysteine

metabolism do not appear to contribute to this syndrome (Fan et al., 2008a), suggesting that the homocysteine elevation seen in these patients may be environmental in origin.

In addition to homocysteine, other candidates for secondary factors influencing the disease have been investigated including *clusterin* (Burdon et al., 2008) and most recently *elastin* (Fan et al., 2009). Variants in these genes were not significantly associated with the disease and did not show any evidence for gene–gene interactions with *LOXL1*, arguing that these genes do not significantly contribute to the condition. Further studies evaluating genes coding for other proteins that interact with *LOXL1* or function in elastic tissues could help identify additional genetic factors contributing to this form of glaucoma.

The *LOXL1* glycine allele specified by rs3825942 is significantly associated with disease in all populations studied; however, it is not yet known if this amino acid substitution is directly responsible for the disease or if it is in linkage disequilibrium with other causal variants. Initial studies have suggested that the condition results from a loss of lyslyl oxidase 1 function, and the glycine at position 153 could affect the catalytic activity of the protein through modifications of propeptide cleavage and binding to substrates such as tropoelastin and fibulin-5 (Schlötzer-Schrehardt et al., 2008).

2009 UPDATE REFERENCES

Altintaş, O., Maral, H., Yüksel, N., Karabaş, V.L., Dillioğlugil, M.O. and Cağlar, Y. (2005). Homocysteine and nitric oxide levels in plasma of patients with pseudoexfoliation syndrome, pseudoexfoliation glaucoma, and primary open-angle glaucoma. *Graefes Arch Clin Exp Ophthalmol* 243(7), 677–683.

Burdon, K.P., Sharma, S., Hewitt, A.W., McMellon, A.E., Wang, J.J., Mackey, D.A., Mitchell, P. and Craig, J.E. (2008). Genetic analysis of the clusterin gene in pseudoexfoliation syndrome. *Mol Vis* 14, 1727–1736.

Challa, P., Schmidt, S., Liu, Y., Qin, X., Vann, R.R., Gonzalez, P., Allingham, R.R. and Hauser, M.A. (2008). Analysis of LOXL1 polymorphisms in a United States population with pseudoexfoliation glaucoma. *Mol Vis* 14, 146–149.

Fan, B.J., Chen, T., Grosskreutz, C., Pasquale, L., Rhee, D., DelBono, E., Haines, J.L. and Wiggs, J.L. (2008a). Lack of association of polymorphisms in homocysteine metabolism genes with pseudoexfoliation syndrome and glaucoma. *Mol Vis* 14, 2484–2491.

Fan, B.J., Pasquale, L., Grosskreutz, C.L., Rhee, D., Chen, T., DeAngelis, M.M., Kim, I., del Bono, E., Miller, J.W., Li, T. et al. (2008b). DNA sequence variants in the LOXL1 gene are associated with pseudoexfoliation glaucoma in a U.S. clinic-based population with broad ethnic diversity. *BMC Med Genet* 6(9), 5.

Fan, B.J., Chen, T., Grosskreutz, C., Pasquale, L., Rhee, D., DelBono, E., Haines, J.L. and Wiggs, J.L. (2009). Lack of association of polymorphisms in Elastin with pseudoexfoliation syndrome and glaucoma. *Mol Vis*, submitted.

Fingert, J.H., Alward, W.L., Kwon, Y.H., Wang, K., Streb, L.M., Sheffield, V.C. and Stone, E.M. (2007). LOXL1 mutations are associated with exfoliation syndrome in patients from the midwestern United States. *Am J Ophthalmol* 144(6), 974–975.

Gong, W.F., Chiang, S.W., Chen, L.J., Tam, P.O., Jia, L.Y., Leung, D.Y., Geng, Y.Q., Tham, C.C., Lam, D.S., Ritch, R. et al. (2008). Evaluation of LOXL1 polymorphisms in primary open-angle glaucoma in southern and northern Chinese. *Mol Vis* 14, 2381–2389.

Hewitt, A.W., Sharma, S., Burdon, K.P., Wang, J.J., Baird, P.N., Dimasi, D.P., Mackey, D.A., Mitchell, P. and Craig, J.E. (2007). Ancestral LOXL1 variants are associated with pseudoexfoliation in Caucasian Australians but with markedly lower penetrance than in Nordic people. *Hum Mol Genet* 17(5), 710–716.

Liu, X., Zhao, Y., Gao, J., Pawlyk, B., Starcher, B., Spencer, J.A., Yanagisawa, H., Zuo, J. and Li, T. (2004). Elastic fiber homeostasis requires lysyl oxidase-like 1 protein. *Nat Genet* 36(2), 178–182.

Martone, G., Casprini, F., Traversi, C., Lepri, F., Pichierri, P. and Caporossi, A. (2007). Pseudoexfoliation syndrome: *In vivo* confocal microscopy analysis. *Clin Experiment Ophthalmol* 35(6), 582–585.

Ozaki, M., Lee, K.Y., Vithana, E.N., Yong, V.H., Thalamuthu, A., Mizoguchi, T., Venkatraman, A. and Aung, T. (2008). Association of LOXL1 gene polymorphisms with pseudoexfoliation in the Japanese. *Invest Ophthalmol Vis Sci* 49(9), 3976–3980.

Parodi, M.B., Liberali, T., Iacono, P., De Nicola, M., Ballone, E. and Ravalico, G. (2008). The spectrum of iris angiography abnormalities in pseudoexfoliation syndrome. *Eye* 22(1), 49–54.

Pasutto, F., Krumbiegel, M., Mardin, C.Y., Paoli, D., Lämmer, R., Weber, B.H., Kruse, F.E., Schlötzer-Schrehardt, U. and Reis, A. (2008). Association of LOXL1 common sequence variants in German and Italian patients with pseudoexfoliation syndrome and pseudoexfoliation glaucoma. *Invest Ophthalmol Vis Sci* 49(4), 1459–1463.

Ramprasad, V.L., George, R., Soumittra, N., Sharmila, F., Vijaya, L. and Kumaramanickavel, G. (2008). Association of non-synonymous single nucleotide polymorphisms in the LOXL1 gene with pseudoexfoliation syndrome in India. *Mol Vis* 14, 318–322.

Roedl, J.B., Bleich, S., Reulbach, U., Rejdak, R., Naumann, G.O., Kruse, F.E., Schlötzer-Schrehardt, U., Kornhuber, J. and Jünemann, A.G. (2007). Vitamin deficiency and hyperhomocysteinemia in pseudoexfoliation glaucoma. *J Neural Transm* 114(5), 571–575.

Schlötzer-Schrehardt, U., Pasutto, F., Sommer, P., Hornstra, I., Kruse, F.E., Naumann, G.O., Reis, A. and Zenkel, M. (2008). Genotype-correlated expression of lysyl oxidase-like 1 in ocular tissues of patients with pseudoexfoliation syndrome/glaucoma and normal patients. *Am J Pathol* 173(6), 1724–1735.

Thorleifsson, G., Magnusson, K.P., Sulem, P., Walters, G.B., Gudbjartsson, D.F., Stefansson, H., Jonsson, T., Jonasdottir, A., Jonasdottir, A., Stefansdottir, G. et al. (2007). Common sequence variants in the LOXL1 gene confer susceptibility to exfoliation glaucoma. *Science* 317(5843), 1397–1400.

Yang, X., Zabriskie, N.A., Hau, V.S., Chen, H., Tong, Z., Gibbs, D., Farhi, P., Katz, B.J., Luo, L., Pearson, E. et al. (2008). Genetic association of LOXL1 gene variants and exfoliation glaucoma in a Utah cohort. *Cell Cycle* 7(4), 521–524.

REFERENCES

Abramson, D.H. and Schefler, A.C. (2004). Update on retinoblastoma. *Retina* 24(6), 828–848.

Acland, G.M., Aguirre, G.D., Bennett, J., Aleman, T.S., Cideciyan, A.V., Bennicelli, J., Dejneka, N.S., Pearce-Kelling, S.E., Maguire, A.M., Palczewski, K. et al. (2005). Long-term restoration of rod and cone vision by single dose rAAV-mediated gene transfer to the retina in a canine model of childhood blindness. *Mol Ther* 12(6), 1072–1082.

Al-Baradie, R., Yamada, K., St Hilaire, C., Chan, W.M., Andrews, C., McIntosh, N., Nakano, M., Martonyi, E.J., Raymond, W.R., Okumura, S. et al. (2002). Duane radial ray syndrome (Okihiro syndrome) maps to 20q13 and results from mutations in SALL4, a new member of the SAL family. *Am J Hum Genet* 71(5), 1195–1199.

Allanson, J.E. and Hennekam, R.C. (1997). Rubinstein-Taybi syndrome: Objective evaluation of craniofacial structure. *Am J Med Genet* 71(4), 414–419.

Bagheri, A. and Repka, M.X. (2005). Association of Duane retraction syndrome and Brown syndrome. *J Pediatr Ophthalmol Strabismus* 42(4), 235–237.

Berry, V., Mackay, D., Khaliq, S., Francis, P.J., Hameed, A., Anwar, K., Mehdi, S.Q., Newbold, R.J., Ionides, A., Shiels, A. et al. (1999). Connexin 50 mutation in a family with congenital zonular nuclear pulverulent cataract of Pakistani origin. *Hum Genet* 105(1-2), 168–170.

Botzenhart, E.M., Green, A., Ilyina, H., Konig, R., Lowry, R.B., Lo, I.F., Shohat, M., Burke, L., McGaughran, J., Chafai, R. et al. (2005). SALL1 mutation analysis in Townes-Brocks syndrome: Twelve novel mutations and expansion of the phenotype. *Hum Mutat* 26(3), 282.

Chan, W.M., Traboulsi, E.I., Arthur, B., Friedman, N., Andrews, C. and Engle, E.C. (2006). Horizontal gaze palsy with progressive scoliosis can result from compound heterozygous mutations in ROBO3. *J Med Genet* 43(3), e11.

Chen, Y., Moiseyev, G., Takahashi, Y. and Ma, J.X. (2006). RPE65 gene delivery restores isomerohydrolase activity and prevents early cone loss in Rpe65-/- mice. *Invest Ophthalmol Vis Sci* 47(3), 1177–1184.

Dryja, T.P., Hahn, L.B., Cowley, G.S., McGee, T.L. and Berson, E.L. (1991). Mutation spectrum of the rhodopsin gene among patients with autosomal dominant retinitis pigmentosa. *Proc Natl Acad Sci USA* 88(20), 9370–9374.

Edwards, A.O., Ritter, R., 3rd, Abel, K.J., Manning, A., Panhuysen, C. and Farrer, L.A. (2005). Complement factor H polymorphism and age-related macular degeneration. *Science* 308(5720), 421–424.

Eiberg, H., Moller, H.U. and Berendt, I. (1994). Assignment of granular corneal dystrophy Groenouw type I locus to within a 2 cM interval. *Eur J Hum Genet* 2, 132–138.

El Kochairi, I., Letovanec, I., Uffer, S., Munier, F.L., Chaubert, P. and Schorderet, D.F. (2006). Systemic investigation of keratoepithelin deposits in TGFBI/BIGH3-related corneal dystrophy. *Mol Vis* 12, 461–466.

Engle, E.C. (2006). The genetic basis of complex strabismus. *Pediatr Res* 59(3), 343–348.

Fan, B.J., Pasquale, L., Grosskreutz, C.L., Rhee, D., Chen, T., DeAngelis, M.M., Kim, I., DelBono, E., Miller, J.W., Li, T., et al. (2008). DNA sequence variants in the LOXL1 gene are associated with pseudoexfoliation glaucoma in a US clinic-based population with broad ethnic diversity. *BMC Med Genet* 9, 5.

Fingert, J.H., Stone, E.M., Sheffield, V.C. and Alward, W.L. (2002). Myocilin glaucoma. *Surv Ophthalmol* 47(6), 547–561.

Folberg, R., Alfonso, E., Croxatto, J.O. et al. (1988). Clinically atypical granular corneal dystrophy with pathologic features of lattice-like amyloid deposits. *Ophthalmology* 95, 46–51.

Galvin, J.A., Fishman, G.A., Stone, E.M. and Koenekoop, R.K. (2005). Evaluation of genotype-phenotype associations in Leber congenital amaurosis. *Retina* 25(7), 919–929.

Gordon, M.O., Beiser, J.A., Brandt, J.D., Heuer, D.K., Higginbotham, E.J., Johnson, C.A., Keltner, J.L., Miller, J.P., Parrish, R.K., 2nd, Wilson, M.R. et al. (2002). The Ocular Hypertension Treatment Study: Baseline factors that predict the onset of primary open-angle glaucoma. *Arch Ophthalmol* 120(6), 714–720.

Gould, D.B., Smith, R.S. and John, S.W. (2004). Anterior segment development relevant to glaucoma. *Int J Dev Biol* 48(8-9), 1015–1029.

Gregory, C.Y., Evans, K. and Bhattacharya, S.S. (1995). Genetic refinement of the chromosome 5q lattice corneal dystrophy to within a 2 cM interval. *J Med Genet* 32, 224–226.

Haines, J.L., Hauser, M.A., Schmidt, S., Scott, W.K., Olson, L.M., Gallins, P., Spencer, K.L., Kwan, S.Y., Noureddine, M., Gilbert, J.R. et al. (2005). Complement factor H variant increases the risk of age-related macular degeneration. *Science* 308(5720), 419–421.

Hauser, M.A., Allingham, R.R., Linkroum, K., Wang, J., Larocque-Abramson, K., Figueiredo, D., Santiago-Turla, C., Del Bono, E.A., Haines, J.L., Pericak-Vance, M.A. et al. (2006). Distribution of WDR36 DNA sequence variants in patients with primary open-angle glaucoma. *Invest Ophthalmol Vis Sci* 47(6), 2542–2546.

Hejtmancik, J.F. and Kantorow, M. (2004). Molecular genetics of age-related cataract. *Exp Eye Res* 79(1), 3–9.

Horwitz, J. (2003). Alpha-crystallin. *Exp Eye Res* 76(2), 145–153.

Houdayer, C., Gauthier-Villars, M., Lauge, A., Pages-Berhouet, S., Dehainault, C., Caux-Moncoutier, V., Karczynski, P., Tosi, M., Doz, F., Desjardins, L. et al. (2004). Comprehensive screening for constitutional RB1 mutations by DHPLC and QMPSF. *Hum Mutat* 23(2), 193–202.

Hyman, L. and Neborsky, R. (2002). Risk factors for age-relatedmacular degeneration: An update. *Curr Opin Ophthalmol* 13(3), 171–175.

Idrees, F., Vaideanu, D., Fraser, S.G., Sowden, J.C. and Khaw, P.T. (2006). A review of anterior segment dysgeneses. *Surv Ophthalmol* 51(3), 213–231.

Inagaki, N., Hayashi, T., Arimura, T., Koga, Y., Takahashi, M., Shibata, H., Teraoka, K., Chikamori, T., Yamashina, A. and Kimura, A. (2006). Alpha B-crystallin mutation in dilated cardiomyopathy. *Biochem Biophys Res Commun* 342(2), 379–386.

Johns, D.R. and Colby, K.A. (2002). Treatment of Leber's hereditary optic neuropathy: Theory to practice. *Semin Ophthalmol* 17(1), 33–38.

Kennan, A., Aherne, A. and Humphries, P. (2005). Light in retinitis pigmentosa. *Trends Genet* 21(2), 103–110.

Kisselev, O.G. (2005). Focus on molecules: Rhodopsin. *Exp Eye Res* 81(4), 366–367.

Klein, R.J., Zeiss, C., Chew, E.Y., Tsai, J.Y., Sackler, R.S., Haynes, C., Henning, A.K., SanGiovanni, J.P., Mane, S.M., Mayne, S.T. et al. (2005). Complement factor H polymorphism in age-related macular degeneration. *Science* 308(5720), 385–389.

Klintworth, G.K. (1967). Lattice corneal dystrophy: An inherited variety of amyloidosis restricted to the cornea. *Am J Pathol* 50, 371–399.

Klintworth, G.K. (2003). The molecular genetics of the corneal dystrophies – current status. *Front Biosci* 8, d687–d713.

Knudson, A.G. (2001). Two genetic hits (more or less) to cancer. *Nat Rev Cancer* 1(2), 157–162.

Korvatska, E., Munier, F.L., Djemai, A., Wang, M.X., Frueh, B., Chiou, A.G., Uffer, S., Ballestrazzi, E., Braunstein, R.E., Forster, R.K. et al. (1998). Mutation hot spots in 5q31-linked corneal dystrophies. *Am J Hum Genet* 62(2), 320–324.

Krutovskikh, V. and Yamasaki, H. (2000). Connexin gene mutations in human genetic diseases. *Mutat Res* 462(2-3), 197–207.

Kuchle, M., Green, W.R., Volcker, H.E. et al. (1995). Reevaluation of corneal dystrophies of Bowman's layer and the anterior stroma (Reis Bücklers and Thiel Behnke types): A light and electron microscopic study of eight corneas and a review of the literature. *Cornea* 14, 333–354.

Li, Y., Wang, J., Dong, B. and Man, H. (2004). A novel connexin46 (GJA3) mutation in autosomal dominant congenital nuclear pulverulent cataract. *Mol Vis* 10, 668–671.

Liu, Y. and Vollrath, D. (2004). Reversal of mutant myocilin non-secretion and cell killing: Implications for glaucoma. *Hum Mol Genet* 13(11), 1193–1204.

Liu, Y., Zhang, X., Luo, L., Wu, M., Zeng, R., Cheng, G., Hu, B., Liu, B., Liang, J.J. and Shang, F. (2006). A novel alphaB-crystallin mutation associated with autosomal dominant congenital lamellar cataract. *Invest Ophthalmol Vis Sci* 47(3), 1069–1075.

Moller, H.U. (1989). Inter-familial variability and intra-familial similarities of granular corneal dystrophy Groenouw type I with respect to biomicroscopical appearance and symptomatology. *Act Ophthalmol* 67, 669–677.

Monemi, S., Spaeth, G., DaSilva, A., Popinchalk, S., Ilitchev, E., Liebmann, J., Ritch, R., Heon, E., Crick, R.P., Child, A. et al. (2005). Identification of a novel adult-onset primary open-angle glaucoma (POAG) gene on 5q22.1. *Hum Mol Genet* 14(6), 725–733.

Munier, F.L., Korvatska, E., Djemai, A. et al. (1997). Kerato-epithelin mutations in four 5q31-linked corneal dystrophies. *Nat Genet* 15, 247–251.

Nakano, M., Yamada, K., Fain, J., Sener, E.C., Selleck, C.J., Awad, A.H., Zwaan, J., Mullaney, P.B., Bosley, T.M. and Engle, E.C. (2001). Homozygous mutations in ARIX(PHOX2A) result in congenital fibrosis of the extraocular muscles type 2. *Nat Genet* 29(3), 315–320.

Nallasamy, S., Kherani, F., Yaeger, D., McCallum, J., Kaur, M., Devoto, M., Jackson, L.G., Krantz, I.D. and Young, T.L. (2006). Ophthalmologic findings in Cornelia de Lange syndrome: A genotype-phenotype correlation study. *Arch Ophthalmol* 124(4), 552–557.

Newman, N.J. (2005). Hereditary optic neuropathies: From the mitochondria to the optic nerve. *Am J Ophthalmol* 140(3), 517–523.

Olichon, A., Guillou, E., Delettre, C., Landes, T., Arnaune-Pelloquin, L., Emorine, L.J., Mils, V., Daloyau, M., Hamel, C., Amati-Bonneau, P. et al. (2006). Mitochondrial dynamics and disease, OPA1. *Biochim Biophys Acta* 1763, 500–509.

Olsson, J.E., Gordon, J.W., Pawlyk, B.S., Roof, D., Hayes, A., Molday, R.S., Mukai, S., Cowley, G.S., Berson, E.L. and Dryja, T.P. (1992). Transgenic mice with a rhodopsin mutation (Pro23His): A mouse model of autosomal dominant retinitis pigmentosa. *Neuron* 9(5), 815–830.

Reddy, M.A., Francis, P.J., Berry, V., Bhattacharya, S.S. and Moore, A.T. (2004). Molecular genetic basis of inherited cataract and associated phenotypes. *Surv Ophthalmol* 49(3), 300–315.

Rivera, A., Fisher, S.A., Fritsche, L.G., Keilhauer, C.N., Lichtner, P., Meitinger, T. and Weber, B.H. (2005). Hypothetical LOC387715 is a second major susceptibility gene for age-related macular degeneration, contributing independently of complement factor H to disease risk.. *Hum Mol Genet* 14(21), 3227–3236.

Robinson, P.R., Cohen, G.B., Zhukovsky, E.A. and Oprian, D.D. (1992). Constitutively active mutants of rhodopsin. *Neuron* 9(4), 719–725.

Rosenwasser, G.O., Sucheski, B.M., Rosa, N. et al. (1993). Phenotypic variation in combined granular-lattice (Avellino) corneal dystrophy. *Arch Ophthalmol* 111, 1546–1552.

Sale, M.M., Craig, J.E., Charlesworth, J.C., FitzGerald, L.M., Hanson, I.M., Dickinson, J.L., Matthews, S.J., Heyningen Vv, V., Fingert, J.H. and Mackey, D.A. (2002). Broad phenotypic variability in a single pedigree with a novel 1410delC mutation in the PST domain of the PAX6 gene. *Hum Mutat* 20(4), 322.

Sandgren, O., Holmgren, G. and Lundgren, E. (1990). Vitreous amyloidosis associated with homozygosity for the transthyretin methionine-30 gene. *Arch Ophthalmol* 108, 1584–1586.

Schmidt, S., Hauser, M.A., Scott, W.K., Postel, E.A., Agarwal, A., Gallins, P., Wong, F., Chen, Y.S., Spencer, K., Schnetz-Boutaud, N. et al. (2006). Cigarette smoking strongly modifies the association of LOC387715 and age-related macular degeneration. *Am J Hum Genet* 78(5), 852–864.

Simpson, T.I. and Price, D.J. (2002). Pax6; a pleiotropic player in development. *Bioessays* 24(11), 1041–1051.

Singh, S., Tang, H.K., Lee, J.Y. and Saunders, G.F. (1998). Truncation mutations in the transactivation region of PAX6 result in dominant-negative mutants. *J Biol Chem* 273(34), 21531–21541.

Small, K.W., Mullen, L., Barletta, J. et al. (1996). Mapping of Reis–Bücklers' corneal dystrophy to chromosome 5q. *Am J Ophthalmol* 121, 384–390.

Stone, E.M. (2007). Genetic testing for inherited eye disease. *Arch Ophthal* 125, 205–212.

Stone, E.M., Mathers, W.D., Rosenwasser, G.O. et al. (1994). Three autosomal dominant corneal dystrophies map to chromosome 5q. *Nat Genet* 6, 47–51.

Stone, E.M., Sheffield, V.C. and Hageman, G.S. (2001). Molecular genetics of age-related macular degeneration. *Hum Mol Genet* 10(20), 2285–2292.

Takahashi, Y., Moiseyev, G., Chen, Y. and Ma, J.X. (2005). Identification of conserved histidines and glutamic acid as key residues for isomerohydrolase activity of RPE65, an enzyme of the visual cycle in the retinal pigment epithelium. *FEBS Lett* 579(24), 5414–5418.

Terhal, P., Rosler, B. and Kohlhase, J. (2006). A family with features overlapping Okihiro syndrome, hemifacial microsomia and isolated Duane anomaly caused by a novel SALL4 mutation. *Am J Med Genet A* 140(3), 222–226.

Thorleifsson, G., Magnusson, K.P., Sulem, P., Walters, G.B., Gudbjartsson, D.F., Stefansson, H., Jonsson, T., Jonasdottir, A., Jonasdottir, A., Stefansdottir, G. et al. (2007). Common sequence variants in the LOXL1 gene confer susceptibility to exfoliation glaucoma. *Science* 317(5843), 1397–1400.

Tischfield, M.A., Bosley, T.M., Salih, M.A., Alorainy, I.A., Sener, E.C., Nester, M.J., Oystreck, D.T., Chan, W.M., Andrews, C., Erickson, R.P. et al. (2005). Homozygous HOXA1 mutations disrupt human brainstem, inner ear, cardiovascular and cognitive development. *Nat Genet* 37(10), 1035–1037.

Traboulsi, E.I. (2004). Congenital abnormalities of cranial nerve development: Overview, molecular mechanisms, and further evidence of heterogeneity and complexity of syndromes with congenital limitation of eye movements. *Trans Am Ophthalmol Soc* 102, 373–389.

Tsai, T., Fulton, L., Smith, B.J., Mueller, R.L., Gonzalez, G.A., Uusitalo, M.S. and O'Brien, J.M. (2004). Rapid identification of

germline mutations in retinoblastoma by protein truncation testing. *Arch Ophthalmol* 122(2), 239–248.

Tuo, J., Bojanowski, C.M. and Chan, C.C. (2004). Genetic factors of age-related macular degeneration. *Prog Retin Eye Res* 23(2), 229–249.

Tzoulaki, I., White, I.M. and Hanson, I.M. (2005). PAX6 mutations: Genotype-phenotype correlations. *BMC Genet* 6(1), 27.

Valentino, M.L., Barboni, P., Ghelli, A., Bucchi, L., Rengo, C., Achilli, A., Torroni, A., Lugaresi, A., Lodi, R., Barbiroli, B. et al. (2004). The ND1 gene of complex I is a mutational hot spot for Leber's hereditary optic neuropathy. *Ann Neurol* 56(5), 631–641.

Wiggs, J., Nordenskjold, M., Yandell, D., Rapaport, J., Grondin, V., Janson, M., Werelius, B., Petersen, R., Craft, A., Riedel, K. et al. (1988). Prediction of the risk of hereditary retinoblastoma, using DNA polymorphisms within the retinoblastoma gene. *N Engl J Med* 318(3), 151–157.

Wiggs, J.L. (2005). Genes associated with human glaucoma. *Ophthalmol Clin North Am* 18(3), 335–343.

Wiggs, J.L. (2006). Complement factor H and macular degeneration: The genome yields an important clue. *Arch Ophthalmol* 124(4), 577–578.

Xia, C.H., Cheng, C., Huang, Q., Cheung, D., Li, L., Dunia, I., Benedetti, L.E., Horwitz, J. and Gong, X. (2006). Absence of alpha3 (Cx46) and alpha8 (Cx50) connexins leads to cataracts by affecting lens inner fiber cells. *Exp Eye Res* May 9. [Epub ahead of print].

Yamada, K., Andrews, C., Chan, W.M., McKeown, C.A., Magli, A., de Berardinis, T., Loewenstein, A., Lazar, M., O'Keefe, M., Letson, R. et al. (2003). Heterozygous mutations of the kinesin KIF21A in congenital fibrosis of the extraocular muscles type 1 (CFEOM1). *Nat Genet* 35(4), 318–321.

Yurdakul, N.S., Ugurlu, S. and Maden, A. (2006). Strabismus in Down syndrome. *J Pediatr Ophthalmol Strabismus* 43(1), 27–30.

van Heyningen, V. and Williamson, K.A. (2002). PAX6 in sensory development. *Hum Mol Genet* 11(10), 1161–1167.

CHAPTER

Genomics in the Diagnosis and Management of Depression

Brigitta Bondy

INTRODUCTION

Depression is a highly prevalent, potentially life-threatening condition that affects hundreds of millions of people all over the world. It can occur at any age from childhood to late life and exerts a tremendous cost upon society. The psychopathological state involves a triad of symptoms with low or depressed mood, anhedonia and low energy or fatigue. Other symptoms, such as sleep and psychomotor disturbances, feelings of guilt, low self-esteem, suicidal tendencies as well as autonomous and gastrointestinal disturbances are also often present.

Depression is not a homogenous disorder but a complex phenomenon with many subtypes. The differences in symptomatology range from mild symptoms to severe symptoms with or without psychotic features. Interactions with other psychiatric and somatic disorders are quite common. An interaction between depression and cardiovascular disease, with worsened prognosis after myocardial infarction and increased mortality rates, is especially well documented.

The etiology of depression is multi-factorial, including non-genetic and genetic factors. As with most human complex disorders, the relationship between genes and the disorder is not straightforward but the result of complicated interactions between multiple genes and the environment. Similarly, the response to antidepressant medication varies among patients and is also at least partly influenced by genetic variants.

DIAGNOSIS, PREVALENCE AND COURSE OF DEPRESSION

The clinical course of major depression (formerly unipolar depression) is characterized by one or more major depressive episodes without a history of manic, mixed or hypomanic episodes. According to the diagnostic criteria of DSMS-IV (American Psychiatric Association, 1994), five of the following symptoms have to be present for a minimum of two weeks: depressed mood, loss of interest or pleasure, significant alteration in weight and appetite, insomnia or hyposomnia, disturbances in psychomotor activity with either agitation or retardation, fatigue or loss of energy, feelings of worthlessness, diminished ability to think or concentrate and, last but not least, recurrent thoughts of death and suicidal ideation or acts. Recurrent episodes of major depressive disorders may differ in symptomatology and thus show pleomorphic manifestations within one individual (Oquendo et al., 2004).

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

The lifetime prevalence of depression is between 10% and 20% in the general population worldwide, with a female to male ratio of about 5:2. Typically, the course of the disease is recurrent and most patients recover from major depressive episodes. However, a substantial proportion of the patients become chronic, and after 5 or 10 years of prospective follow-up, 12% and 7%, respectively, are still depressed (Keller et al., 1997). But patients who recover also have a high rate of recurrence, as approximately 75% of patients experience more than one episode of major depression within 10 years. There is a high level of comorbidity between anxiety and depressive disorders, with co-occurrence rates up to 60% (Gorman, 1996). This suggests that comorbid anxiety and depression are the rule rather than exception. Furthermore, there is high co-occurrence of neuroticism, which is characterized by dysphoria, tension and emotional reactivity and is often a premorbid personality structure and a robust predictor for future onset of depression. It is estimated that both anxiety and neuroticism share about 50% of the genetic factors with depression.

PATHOPHYSIOLOGICAL MECHANISMS

Depressive disorders are complex phenomena in terms of symptomatology and multi-factorial in etiopathogenesis. Although the specific causes have not been elucidated in detail, there is overwhelming evidence that depression is caused by the interaction of multiple genetic risk factors and environmental and neurobiological factors (Mann and Currier, 2006) (Figure 56.1). Cross-influences can be seen between all these pathways, and genes contribute to all of them (Kendler et al., 2006).

The Neurobiological Basis of Depression

Most theories about the neurobiology of depression involve functional deficiencies of the brain monoaminergic transmitters, especially serotonin (5-HT) and norepinephrine (NE). Both neurotransmitters modulate many behavioral symptoms that are disturbed in depression, such as mood, vigilance, motivation, fatigue and psychomotor agitation or retardation. Abnormal function may arise from altered synthesis, storage or release of the neurotransmitters as well as from disturbed sensitivity of their receptors or sub-cellular messenger functions within the synapse (Figure 56.2). Transport proteins play a crucial role in neural transmission, as they reduce the availability of the neurotransmitters in the synaptic cleft and thus terminate their effect on pre- and postsynaptic receptors. It is also noteworthy that most of the available antidepressants have an impact on the 5-HT transporter. On the basis of recent evidence, which showed that the proteins of the postsynapse are involved in long-term adaptive mechanisms in response to altered transmission during disease or when drugs take effect, these proteins are now considered to be the main modulators of neuronal activity and pathophysiology of mental disorders (Manji and Duman, 2001). Newer hypotheses stress the importance of this adaptation or plasticity of neuronal systems and propose that depression could result from an inability to make the appropriate adaptive responses to stress or other aversive stimuli, and that antidepressants may act by correcting this dysfunction or by directly inducing the appropriate adaptive responses (Duman, 2004). Neurotrophic factors (e.g., the brain-derived neurotrophic factor (BDNF), and the BDNF receptor (trkB)) are involved in these processes and promote the function and growth of neurones in the adult brain that contain 5-HT.

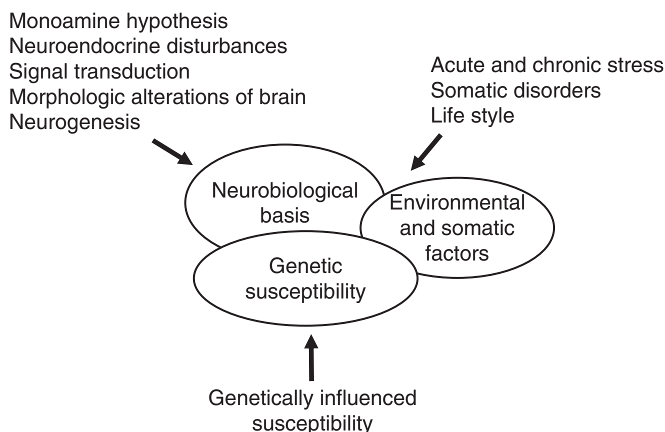


Figure 56.1 Hypothetical model for the etiology of depression.

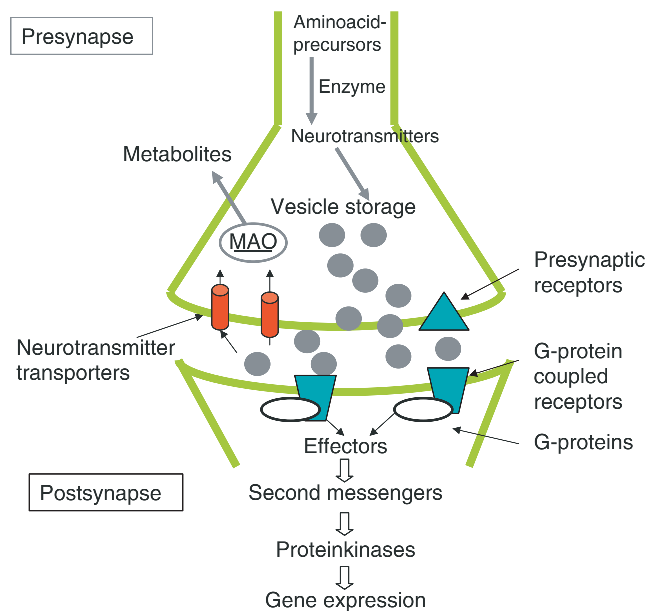


Figure 56.2 The signal transduction cascade in the synapse. All compartments may be disturbed, beginning with neurotransmitter production, storage and release within the presynapse, down to postsynaptic events, mediated via neurotransmitter–receptor coupling and the activation of postsynaptic effectors. According to recent knowledge, the long-term adaptation processes with regulation of gene expression are crucially involved in pathophysiology of disorders and drug response.

Despite these transmission processes within neurones, the stress-response system involving the hypothalamus-pituitary-adrenal (HPA) axis is of paramount importance in the development of depression. One of the consistent findings in psychiatry is that a significant proportion of depressed patients hypersecretes hypothalamic-corticotropin-releasing hormone (CRH), which stimulates adrenocorticotropin (ACTH) secretion from the pituitary and finally leads to increased cortisol levels and inadequate glucocorticoid feedback (Holsboer, 2000). Furthermore, various stress hormones interact with the serotonergic system on several levels, as sustained CRH or cortisol overdrive down-regulate the serotonergic system in terms of turnover, activity of the neurons and alterations on the receptor basis (van Praag, 2005). Thus, a possible conclusion might be that disturbances of the 5-HT and stress hormone systems are of pathophysiological relevance in depression and are not merely an epiphenomenon of the condition.

Genetic Basis of Major Depression

It is well known that depression runs in families. Numerous studies have documented that first-degree relatives of mood disorder patients have an approximately threefold risk of developing depression. There is further evidence that these relatives are also more susceptible to anxiety, substance abuse or social impairment compared to the offspring of non-depressed parents (Weissman et al., 2006). However, the familial loading could also be the result of shared environmental factors, thus suggesting that the vulnerability to depression could be due to nurture rather than nature. A number of studies comparing the prevalence of depression among mono- and dizygotic twins gave sufficient evidence that major depression (as a disease) is moderately heritable, and on the basis of these twin studies, estimated that heritability may be 40–50% (Malhi et al., 2000). Furthermore, there is no doubt that more than one single gene is responsible for the increased vulnerability and that considerable gene–gene and gene–environment interactions complicate the identification of relevant susceptibility genes (Levinson, 2005).

Genetic Linkage Analyses

As heritability of depression does not follow the Mendelian pattern, the available linkage analyses that investigated a major gene effect are less convincing, and genome scan results are not yet available. Nevertheless, in the last decade a number of different results were obtained for depression and related traits. Selected regions on chromosomes 3, 4, 6, 12, 15 and 18 (for review see Levinson, 2005), especially in combination with depression-related personality traits such as neuroticism and harm avoidance, have received support from more than one study. However, as is the case for all linkage analyses for complex disorders, one cannot predict which will be true positives in the long run and thus until now none of these chromosomal regions has really been accepted as a susceptibility gene for depression. As statistical power is one of the fundamental problems of most studies, the full impact of genetic linkage findings on the search for depression susceptibility genes might come from combined analyses of multiple datasets (Levinson, 2005).

As major depression is such a complex disorder, association studies, despite all their well-known pitfalls, appeared to be more suitable than linkage analyses and it was hoped that this method might unravel minor effects of genes. The selection of candidate genes followed the neurobiological hypotheses of depression, which will be summarized briefly here.

Association Studies with Different Candidate Genes

Due to the prominent importance of the serotonergic system, association studies have so far focused on polymorphisms in genes encoding for the rate-limiting synthesizing enzyme tryptophan hydroxylase (TPH), the serotonin transporter (5-HTT), the different 5-HT receptors and the degrading enzyme, monoamine oxidase-A (MAOA).

The majority of studies were carried out with the serotonin transporter, which is a crucial protein within the synapse as it cleaves 5-HT from the synaptic cleft and is the target protein of the selective 5-HT reuptake inhibitors and other antidepressants (see section: Psychopharmacogenetics). A polymorphism within the promoter region of the 5-HTT (5-HTTLPR), comprising a 44-base-pair insertion/deletion resulting in short (S) and long (L) alleles, was shown to affect the transcription and thus the function of the gene (Heils et al., 1995). Many studies followed the first findings that the S-allele is associated with anxiety (Lesch et al., 1996), but the results were discrepant and even meta-analyses yielded inconclusive results as to whether the S-allele confers a risk for depression. Methodological issues might be one explanation for the failure to find such an association, because in most studies the samples were small and clinically heterogeneous. In this respect it is remarkable that a recent study, which investigated a large and clinically well characterized sample, clearly demonstrated that the S-allele of the 5-HTTLPR was significantly more frequent in depressed patients than in controls (Hoefgen et al., 2005). This finding could be independently replicated in a large cohort of the Spanish population (Cervilla et al., 2006). On the basis of these new and methodologically subtle studies, one might be tempted to speculate that the S-allele of the 5-HTTLPR might have a small but independent effect on the vulnerability for depression.

A further interesting and promising gene of the serotonergic pathway is the recently identified, brain-specific TPH2 gene (Walther et al., 2003), which is the limiting enzyme in serotonin synthesis of the central nervous system (CNS). There is now evidence that a region in or around introns 5 of the TPH2 gene might be a susceptibility locus for depression. Zill and colleagues were the first to associate a haplotype in intron 5 with major depression (Zill et al., 2004a) and completed suicide (Zill et al., 2004b). Although the exact location and characterization of the relevant mutations have yet to be established, there are now three independent studies showing that haplotype blocks in regions overlapping with that reported by Zill et al. are associated with major depression, suicidality and bipolar disorder patients of different ethnic groups (Harvey et al., 2004; Van Den et al., 2006; Zhou et al., 2005).

Many other genes of the serotonergic pathway were investigated by several groups, but although some positive results emerged the data were contradictory and even some meta-analyses

TABLE 56.1 Important association studies in major depression

Gene symbol	Gene	Result
BDNF	Brain derived neurotrophic factor	270CT, negative result
CRHR1	CRH receptor	One positive result, not replicated
DRD4	Dopamine receptor D4	48 bp repeat, questionable association
GNB3	G-protein β3	Questionable, partly replicated
5-HT1A	Serotonin receptor 1A	−1019CG polymorphism affect serotonin system Predispose depression
5-HT2A	Serotonin receptor 2A	No association
5-HTTLPR	Serotonin transporter	S-allele associated with anxiety, neuroticism S-allele associated with depression in response to stress
TPH2	Tryptophan hydroxylase, brain specific	SNPs in introns 5 associated with depression

Adapted from Levinson, 2005.

did not give convincing evidence for a significant interaction between any of these investigated genes and the genetic liability for major depression (for review see Leonardo and Hen, 2006; Levinson, 2005). A summary of findings of genetic association studies is listed in Table 56.1. Genome-wide association studies could be very valuable and are currently being carried out, but the results are not yet available.

Evaluating the results of both linkage and association studies it emerges that, besides the established heritability for depression, a susceptibility gene or set of genes could not be identified so far. Beyond the fact that mood disorders are complex in nature, may represent a family of related but distinct conditions and are probably polygenic, the definition of the clinical phenotype of depression is a serious problem in molecular genetic studies. Although the diagnostic criteria seem to be clearly defined by the DSM-IV criteria, the co-occurrence of other psychiatric disorders, such as anxiety or alcohol/drug constitutes a major challenge in genetic studies.

Genetics of Emotional Regulation and the Response to Stress

The identification of biological mechanisms through which genes lead to differences in emotional behavior is paramount to understanding how genes confer risk for psychiatric disorders such as depression (Hariri and Holmes, 2006). As 5-HT is a key modulator of emotional behavior sub-serving anxiety and depression, the effects of gene variations in the serotonin transporter gene were investigated for their influence on temperament, character, and the regulation of emotion. After the first observation almost a decade ago (Lesch et al., 1996) that individuals carrying at least one S-allele of the 5-HTTLPR displayed higher levels of anxiety, neuroticism and harm avoidance, many studies were performed to confirm this observation. Although the results of the various studies were not consistent, meta-analyses have demonstrated a significant association between the S-allele and these personality traits (for review see Hariri and Holmes, 2006; Munafo et al., 2005).

Emerging evidence from human and animal studies indicates that a relative loss in the 5-HTT gene function not only increases anxiety but also exerts a negative influence on the capacity to cope with stress, which further increases the risk for mood disorders. This was underlined by a study by Caspi et al., who investigated a representative cohort in a prospective, longitudinal study (Caspi et al., 2003). They could show that individuals with one or two copies of the S-allele exhibited more depressive symptoms, more diagnosable depression and more suicidality in relation to stressful life events than individuals who were homozygous for the L allele. This finding suggests that genetic variants may act to promote one's resistance to environmental pathogens and that the S-allele moderates the "depressiogenic" influence of stressful life events. Similar studies followed and replicated these findings about an increased vulnerability to stress in S-allele carriers, and were further extended to the impact of the dopamine D4 receptor gene (Ebstein, 2006; Hariri and Holmes, 2006). In addition, the study by Grabe et al. (2005) demonstrated this gene–environment interaction in relation to the 5-HTTLPR genotypes in a cohort with severe mental and physical distress (e.g., myocardial infarction, stroke, diabetes and degenerative diseases). They found an interaction between the 5-HTTLPR S-allele and unemployment or chronic disease, but only in females. This finding goes beyond those of the previous studies, and indicates a higher mental vulnerability to social stressors and chronic disease. All these recent findings demonstrate that the 5-HTTLPR affects not only central 5-HTT function, but as a consequence seems also to be involved in the regulation of bio-behavioral characteristics.

Recent studies have begun to investigate how the 5-HTTLPR S-allele might mediate stress reactivity at the level of the neural pathway regulating emotion. The non-invasive method of neuroimaging was used to determine the effects of 5-HTTLPR

on reactivity of the amygdala, a brain region that is crucial in mediating emotion, to stress. An amygdala hyper-reactivity to stress was shown in S-allele carriers in several ethnically different cohorts (Hariri and Holmes, 2006). Thus, these findings identify a possible neuroanatomical substrate for the negative affect associated with the S-allele in healthy volunteers.

PHARMACOGENOMICS OF ANTIDEPRESSANTS

There is substantial unexplained interindividual variability in the response to treatment with psychoactive drugs, as a proportion of patients who receive a regular dose do not respond adequately or experience limiting side effects. Despite the availability of a vast variety of drugs, about 30–50% of patients are non-responders. The nature of drug response is a highly complex phenomenon involving genetic and non-genetic factors, the latter including age, gender, hepatic and renal status, nutrition, smoking and alcohol intake. Within the last decade, the concept of individualized drug therapy on the basis of genetic investigations has become a major issue in psychopharmacology. After early positive results there was much enthusiasm about the identification of a genetic make-up that would be ideal to optimally tailor drug treatment in psychiatry.

Numerous association studies have been carried out since then, with genes coding for either the pharmacokinetic (i.e., the processes that influence bioavailability) or pharmacodynamic (targets of drug action) pathways. Most pharmacodynamic studies investigated candidate genes that were proposed either by the etiopathology of depression or by the putative pharmacological mechanisms of the drugs (Reynolds et al., 2006). However, despite some advances in various fields, the final goal of an optimally tailored therapy still remains elusive.

Polymorphic Drug Metabolizing Enzymes and Pharmacokinetic Aspects

All antipsychotics are subject to extensive metabolism by various enzymes of the cytochrome P-450 (CYP) family, which play a pivotal role in the elimination of these drugs and therefore influence their efficacy and toxicity. Factors that affect CYP function and expression, such as CYP pharmacogenetics and the processes of inhibition and induction, all influence the *in vivo* rates of drug elimination (Murray, 2006). Genetic variants in CYP enzymes constitute multi-allelic systems that express a variety of phenotypes. Patients with these phenotypes can be distinguished as poor (PM), intermediate (IM), extensive (EM) or ultrafast metabolizers (UM) (Brosen, 2004). The PMs lack an active form of the expressed enzyme due to an inactivating allelic variant, IMs have at least one copy of an active gene and UMs contain duplicated or amplified gene copies, thus leading to either increased, maybe toxic, or decreased, maybe ineffective, concentrations of the drug (Oscarson, 2003). Although many CYP enzymes are known, only a few of them are relevant for metabolism of psychoactive drugs, mainly the CYP1A2, CYP2D6, CYP2C19 and CYP3A4.

Earlier studies focused on the interaction between CYP genotype, the plasma concentration of the drugs and the response to treatment. Although a relation between the CYP genotype and plasma concentration could be shown, a clear effect on drug response is missing. The processes of CYP induction and inhibition may have tremendous effects on drug elimination and thus on the incidence of adverse drug effects. Thus, inhibition of the metabolizing enzyme might convert an EM to a PM type. Since many if not most depressives are on a multidrug regimen, there is a large potential for enzyme inhibitory reactions and an increased rate of adverse drug effects. Especially co-medication with antidepressants, but also with some antibiotics or β-blockers, might considerably increase the plasma concentrations of the main drug (Table 56.2); this interaction is

TABLE 56.2 Major cytochrome P450 isoenyzmes (CYP), their substrates, enzyme inhibitors and inductors

Enzyme	Substrate	Inhibitor	Inducer
CYP1A2	TCAs (e.g., amitriptyline, clomipramine, imipramine), duloxetin, fluvoxamine, mirtazapine	Fluvoxamine, grapefruit juice, antibiotics	Carbamazepine, hyperforin, nicotine
CYP2C19	TCAs (e.g., amitriptyline, clomipramine, doxepin, imipramine, trimipramine); SSRIs (e.g., citalopram, fluoxetine, sertraline), moclobemide, venlafaxin,	Fluoxetine, valproic acid, fluvoxamine	Carbamazepine, phenytoin
CYP2D6	TCAs (e.g., amitriptyline, clomipramine, desipramine, imipramine, nortriptyline); SSRIs (e.g., fluoxetine, fluvoxamine, paroxetine),duloxetin, mianserin, venlafaxine,	Fluoxetine, paroxetine, citalopram, duloxetin, fluvoxamine, fluphenazine, moclobemide, haloperidol, perphenazin, propranolol, antibiotics	No inducer known
CYP3A4	TCAs (e.g., amitriptylin, clomipramin, imipramin), citalopram, reboxetine, venlafaxin,	Fluoxetine, fluvoxamine, olanzapine, grapefruit juice nicotine	Carbamazepine, hyperforin, phenytoin

extremely important in those patients who are already PMs due to genetic variants (Grasmader et al., 2004).

On the other hand, some drugs such as carbamazepine, phenytoin or hyperforin are known to up-regulate CYP expression in the liver; the respective substrates are then eliminated more rapidly via this enzyme induction (Grasmader et al., 2004). Together with environmental factors such as smoking and/or alcohol consumption, both of which are inducers of CYP activity, genetic factors might lead to different metabolism of a given drug.

Pharmacodynamic Aspects: Response to Treatment

The term pharmacodynamics encompasses all processes influencing the resulting effect, such as the interaction with target proteins or with mechanisms that are modulated by this interaction. The primary targets of antidepressants are known and are congruent with the etiological hypotheses. For this reason, mostly candidate genes from the monoaminergic pathway have so far been investigated.

One of the most consistent findings in pharmacogenetics is the involvement of the serotonin transporter (5-HTT), which is not only the initial target of the selective serotonin reuptake inhibitors (SSRIs), but is also affected by most of the other antidepressants. The functional variant in the 5'-regulatory region of this it was repeatedly shown in independent that the S-allele is associated with slower response to SSRIs such as paroxetine, citalopram and fluvoxamine, as demonstrated in more than two independent studies (for review see (Serretti et al., 2005). Interestingly, this polymorphism also seems to be involved in response to drugs beyond the SSRIs, for example lithium ions, which are widely used as a mood stabilizer (review (Binder and Holsboer, 2006). Other possible candidate genes of the serotonergic system were those coding for the different 5-HT receptors, the rate-limiting enzyme for the synthesis of tryptophan hydroxylase (TPH1), or monoamine oxidase (MAO), the degrading enzyme. Further investigated genes were of the norepinephrine and dopamine systems (Serretti and Artioli, 2004; Serretti et al., 2005), however, only a few consistent and replicated studies are available (Lefebvre et al., 2006) (see Table 56.3).

TABLE 56.3 Important genes tested for association with antidepressants drug response

Gene symbol	Gene	Association with response
ACE	Angiotensin converting enzyme	Yes, not independently replicated
COMT	Catechol-O-methyltransferase	Yes, not independently replicated
CYP2D6	Cytochrome P450, family2, subfamily 6	Yes, independently replicated
DRD2	Dopamine receptor D2	No
FKBP5	FK506 binding protein	Yes, independently replicated
GNB3	G-protein β3	Yes, independently replicated
HTR1A	Serotonin receptor 1A	Yes, not independently replicated
HTR2A	Serotonin receptor 2A	Yes, independently replicated
MAOA	Monoamine oxidase A	No
HTT	Serotonin transporter	Yes, independently replicated

Adapted from Binder and Holsboer, 2006.

GENE AND PROTEIN EXPRESSION STUDIES

New techniques such as proteomics and the cDNA micro-arrays represent prominent expression profiling techniques to investigate multi-factorial and polygenic complex traits. These techniques allow to study the expression of thousands of genes or proteins in one experiment and are thus powerful instruments for the analysis of pathophysiological mechanisms and for the search for new drug targets (Bertilsson et al., 2002). Most of the work currently carried out covered pharmacogenomic aspects and a recent review demonstrated certain novel candidate genes that may underlie the mechanism of action of antidepressants. Among them are genes affecting neurogenesis (via the HPA axis and related neuroendocrine systems, the cAMP messenger system as well as phopshorylation of CREB), neurotransmitter release and neurite outgrowth via cysteine string protein, as well as several neural cell adhesion molecules (for review see (Yamada et al., 2005)). These findings support the hypothesis that plasticity represents the mechanism of action of drugs. Concerning pathophysiological mechanisms of depression alteration of signaling-, oligodendroglial- and GABA-related genes have been observed and opened up new aspects for studying the pathophysiologic mechanisms and the treatment of the disorder (for review (Sequeira and Turecki, 2006).

Proteomic analyses uses two-dimensional electrophoreses and subsequent mass spectrometric sequencing of proteins to investigate protein changes in relation to the disorder or drug treatment. Due to limited accessibility of postmortem brain samples and several technical problems (as e.g., postmortem delay), only few studies have been carried out with this technique, but they may be seen as a non-hypothesis driven screening method for the detection of new candidates genes in neurobiological research (Johnston-Wilson et al., 2000; Schlicht et al., 2006).

FUTURE CONSIDERATIONS

Although the findings described above are important, none of the positive results can fully account for the heterogeneity in depression or in response to antidepressant treatment. Despite the fact

that most of the genetic studies evaluated small samples and only one or a few genetic variants, it appears almost impossible to identify a susceptibility gene for "depression" as a disorder. One of the main reasons for this is the involvement of complex emergent phenomena, including character, temperament and several neurobiological factors, all of which underlie a genetic contribution. The well-known comorbidity between anxiety, depression and neuroticism, with their interacting genetic contributions, further complicates the identification of a disease susceptibility gene as genes influence all levels of cognitive and emotional behaviors and all levels of biology, and thus transcend phenomenological diagnoses (Hariri and Holmes, 2006).

After years of limited success the field is now conceptually evolving from trying to find a gene or genes that confer a risk for the development of depression or are involved in treatment response. Our increasing knowledge of the basic neurobiology and of the complex interactions and functional circuits will have an increasing impact on the search for relevant genes. Newer hypotheses about depression beyond the monoaminergic theory will produce new sets of candidate genes. According to recent hypotheses about the neurobiology of depression these candidates will emerge from the processes of brain development, the long-term response to stress, regulation of the HPA axis and cortisol secretion as well as from neurotrophic, neurotoxic or inflammatory processes, each of which are also genetically influenced.

Similarly, it is still unclear which neurotransmitter systems are the ultimate target via which drugs produce a clinical effect. This means that the majority of genes responsible for drug response are still unknown. The pharmacogenomic approach uses the recent advances in experimental genomics and proteomics, together with the available sequence information of the Human Genome Project. These developments will not only enable genome-wide screens of several millions of SNPs without the use of a candidate gene strategy, but also functional investigations of gene and/or protein expression on the whole genome or proteome level. The goal of pharmacogenomics is to study the mechanisms by which different genetic factors affect the organism's drug response. This will help to find the most efficacious treatment for patients with specific genetic profiles (Paez-Pereda, 2005).

Although there is no doubt that large-scale gene and/or protein expression analyses or whole genome analyses will provide new insight into the pathophysiology of depression and the mechanisms of therapeutic effects and adverse drug reactions, we are presently lacking a serious impact of genomics in daily clinical management. Considering the well-documented increased familial vulnerability for the disorder, an identification of persons at risk is highly warranted. However, as the vulnerability for the disorder might not directly be caused by a mutant gene but by an altered interaction between the mutant gene and other predisposing factors, future research will be directed toward these interactions. To give an example: the serotonin system plays an important mood mediating role, alterations in genes and their functional consequences on the protein might be among the first to be recognized as vulnerability markers. Although this serotonin innate vulnerability, by itself, is not sufficient to cause a depressive episode (Firk and Markus, 2007), it was shown that additional factors, such as environment might trigger the outburst of the disorder. Thus, healthy individuals with a positive family history of depression and a genetic susceptibility that may be due to a polymorphism in the 5-HT transporter which deteriorates stress coping mechanisms, are more prone to develop depression. This example is the beginning of current and future research and thus opens up interesting options in the future management of depression and the identification of high-risk subjects.

2009 UPDATE

Recent advances in genetic and pharmacogenomic studies

A recent meta-analysis summarized about two hundred papers that met the criteria for statistical evaluation (Lopez-Leon et al., 2008). Investigated were the already known candidate genes for depression, but also several new data were available. The authors concluded that six genes might be relevant for the genetic susceptibility for depression. Among them were those previously discussed, including the serotonin transporter (*5-HTTLPR* and the 40 bp VNTR), the G-protein-ß3 subunit GNB3, as well as the DRD4 receptor polymorphisms. But also other alleles, such as the *APOE* ε2 allele and the 677T-allele of the methylenetetrahydrofolate reductase (*MTHFR*) gene, were shown to have some relevance in the susceptibility for major depression.

New pharmacogenomic data are now available that has been obtained with a sample of 4000 patients being treated with selective serotonin reuptake inhibitors (STAR*D study; Sequenced Treatment Alternatives to Relieve Depression). According to these results, some of the previously discussed genes, such as genes for the FK506 binding protein (FKBP5), the 5-HT2A- and the 5-HT1A-receptors as well as the serotonin-transporter, are now confirmed, at least in the treatment with serotonin reuptake inhibitors. But also some other, so far sparsely investigated genes, entered the focus of interest, such as those for the glutamate receptors or the potassium channels (Lekman et al., 2008, Lin and Chen, 2008).

As pharmacogenomics is a rapidly evolving field, some methodological guidelines have recently been proposed in order to reveal more reliable results (Serretti et al., 2008). Specification of sampling source (inpatients versus outpatients, primary versus tertiary settings), standardization of diagnostic systems and treatments, adequate monitoring of compliance through plasma levels, sufficient length of observation (at least 6 weeks for acute antidepressant treatments, though 3-6 months are preferable), the use of a range of response criteria and the inclusion of possible environmental confounding

variables (life events, social support, temperament) are all potentially important issues when planning pharmacogenetic studies.

Epigenetic mechanisms may explain heterogeneity

Considering the diversity of results in genetic studies, heterogeneity of both mechanisms of disease and response to treatment might be a plausible cause. Recently it was recognized that epigenetic mechanisms may be a potential major complication in genetic and pharmacogenomic studies of depression, as they can produce long-lasting changes in protein availability and function. Epigenetic mechanisms, such as DNA methylation, might ultimately abolish gene function, even without a DNA sequence change, and have thus long-lasting effects within mature neurons (Tsankova et al., 2007). Given that these changes may considerably influence gene-environment interactions, epigenetics can explain several aspects of depression, including the high concordance rates among monozygotic twins or the greater prevalence of depression in women (Krishnan and Nestler, 2008).

Although human data in depression are still lacking and most of our knowledge concerning this concept derives from animal studies, the data and the hypotheses are convincing. According to the first studies with animals, significant stress or primary experiences can influence supporting structures of DNA like chromatin in several regions of the brain, being relevant for developing depression, and these changes in DNA might be directly related to behavioral abnormalities.

2009 UPDATE REFERENCES

Krishnan, V. and Nestler, E.J. (2008). The molecular neurobiology of depression. *Nature* 455, 894–902.

Lekman, M., Paddock, S. and McMahon, F.J. (2008). Pharmacogenetics of major depression: Insights from level 1 of the sequenced treatment alternatives to relieve depression (STAR*D) trial. *Mol Diagn Ther* 12, 321–330.

Lin, E. and Chen, P.S. (2008). Pharmacogenomics with antidepressants in the STAR*D study. *Pharmacogenomics* 9, 935–946.

Lopez-Leon, S., Janssens, A.C., Gonzalez-Zuloeta Ladd, A.M., Del-Favero, J., Claes, S.J., Oostra, B.A. and van Duijn, C.M. (2008). Meta-analyses of genetic studies on major depressive disorder. *Mol Psychiatry* 13, 772–785.

Serretti, A., Kato, M. and Kennedy, J.L. (2008). Pharmacogenetic studies in depression: A proposal for methodologic guidelines. *Pharmacogenomics J.* 8, 90–100.

Tsankova, N., Renthal, W., Kumar, A. and Nestler, E.J. (2007). Epigenetic regulation in psychiatric disorders. *Nat Rev Neurosci* 8, 355–367.

REFERENCES

American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders.* Washington DC.

Bertilsson, L., Dahl, M.L., Dalen, P. and Al Shurbaji, A. (2002). Molecular genetics of CYP2D6: clinical relevance with focus on psychotropic drugs. *Br J Clin Pharmacol* 53, 111–122.

Binder, E.B. and Holsboer, F. (2006). Pharmacogenomics and antidepressant drugs. *Ann Med* 38, 82–94.

Bondy, B., Baghai, T.C., Zill, P., Schule, C., Eser, D., Deiml, T. et al. (2005). Genetic variants in the angiotensin I-converting-enzyme (ACE) and angiotensin II receptor (AT1) gene and clinical outcome in depression. *Prog Neuropsychopharmacol Biol Psychiatry* 29, 1094–1099.

Brosen, K. (2004). Some aspects of genetic polymorphism in the biotransformation of antidepressants. *Therapie* 59, 5–12.

Caspi, A., Sugden, K., Moffitt, T.E., Taylor, A., Craig, I.W., Harrington, H. et al. (2003). Influence of life stress on depression: Moderation by a polymorphism in the 5-HTT gene. *Science* 301, 386–389.

Cervilla, J.A., Rivera, M., Molina, E., Torres-Gonzalez, F., Bellon, J.A., Moreno, B. et al. (2006). The 5-HTTLPR s/s genotype at the serotonin transporter gene (SLC6A4) increases the risk for depression in a large cohort of primary care attendees: The PREDICT-gene study. *Am J Med Genet B Neuropsychiatr Genet* 141B, 912–917.

Duman, R.S. (2004). Role of neurotrophic factors in the etiology and treatment of mood disorders. *Neuromolecular Med* 5, 11–25.

Ebstein, R.P. (2006). The molecular genetic architecture of human personality: beyond self-report questionnaires. *Mol Psychiatr* 11, 427–445.

Firk, C. and Markus, C.R. (2007). Serotonin by stress interaction: a susceptibility factor for the development of depression? *J Psychopharmacol* 21, 538–544.

Gorman, J.M. (1996). Comorbid depression and anxiety spectrum disorders. *Depress Anxiety* 4, 160–168.

Grabe, H.J., Lange, M., Wolff, B., Volzke, H., Lucht, M., Freyberger, H.J. et al. (2005). Mental and physical distress is modulated by a polymorphism in the 5-HT transporter gene interacting with social stressors and chronic disease burden. *Mol Psychiatr* 10, 220–224.

Grasmader, K., Verwohlt, P.L., Rietschel, M., Dragicevic, A., Muller, M., Hiemke, C. et al. (2004). Impact of polymorphisms of cytochrome-P450 isoenzymes 2C9, 2C19 and 2D6 on plasma concentrations and clinical effects of antidepressants in a naturalistic clinical setting. *Eur J Clin Pharmacol* 60, 329–336.

Hariri, A.R. and Holmes, A. (2006). Genetics of emotional regulation: the role of the serotonin transporter in neural function. *Trends Cogn Sci* 10, 182–191.

Harvey, M., Shink, E., Tremblay, M., Gagne, B., Raymond, C., Labbe, M. et al. (2004). Support for the involvement of TPH2 gene in affective disorders. *Mol Psychiatr* 9, 980–981.

Heils, A., Teufel, A., Petri, S., Seemann, M., Bengel, D., Balling, U. et al. (1995). Functional promoter and polyadenylation site mapping of the human serotonin (5-HT) transporter gene. *J Neural Transm Gen Sect* 102, 247–254.

Hoefgen, B., Schulze, T.G., Ohlraun, S., von, W.O., Hofels, S., Gross, M. et al. (2005). The power of sample size and homogenous

sampling: association between the 5-HTTLPR serotonin transporter polymorphism and major depressive disorder. *Biol Psychiatr* 57, 247–251.

Holsboer, F. (2000). The corticosteroid receptor hypothesis of depression. *Neuropsychopharmacol* 23, 477–501.

Johnston-Wilson, N.L., Sims, C.D., Hofmann, J.P., Anderson, L., Shore, A.D., Torrey, E.F. et al. (2000). Disease-specific alterations in frontal cortex brain proteins in schizophrenia, bipolar disorder, and major depressive disorder. The Stanley Neuropathology Consortium. *Mol Psychiatr* 5, 142–149.

Keller, M.B., Hirschfeld, R.M. and Hanks, D. (1997). Double depression: a distinctive subtype of unipolar depression. *J Affect Disord* 45, 65–73.

Kendler, K.S., Gardner, C.O. and Prescott, C.A. (2006). Toward a comprehensive developmental model for major depression in men. *Am J Psychiatr* 163, 115–124.

Lefebvre, R. and Del-Favero, J. (2006). Pharmacogenetics in affective disorders: A druig response approach. *Current Pharmacogenomics* 4, 191–208.

Leonardo, E.D. and Hen, R. (2006). Genetics of affective and anxiety disorders. *Annu Rev Psychol* 57, 117–137.

Lesch, K.P., Bengel, D., Heils, A., Sabol, S.Z., Greenberg, B.D., Petri, S. et al. (1996). Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region [see comments]. *Science* 274, 1527–1531.

Levinson, D.F. (2005). The genetics of depression: A review. *Biol Psychiatr* 60, 84–92.

Malhi, G.S., Moore, J. and McGuffin, P. (2000). The genetics of major depressive disorder. *Curr Psychiatr Rep* 2, 165–169.

Manji, H.K. and Duman, R.S. (2001). Impairments of neuroplasticity and cellular resilience in severe mood disorders: Implications for the development of novel therapeutics. *Psychopharmacol Bull* 35, 5–49.

Mann, J.J. and Currier, D. (2006). Effects of genes and stress on the neurobiology of depression. *Int Rev Neurobiol* 73, 153–189.

Munafo, M.R., Clark, T. and Flint, J. (2005). Does measurement instrument moderate the association between the serotonin transporter gene and anxiety-related personality traits? A meta-analysis. *Mol Psychiatr* 10, 415–419.

Murray, M. (2006). Role of CYP pharmacogenetics and drug–drug interactions in the efficacy and safety of atypical and other antipsychotic agents. *J Pharm Pharmacol* 58, 871–885.

Oquendo, M.A., Barrera, A., Ellis, S.P., Li, S., Burke, A.K., Grunebaum, M. et al. (2004). Instability of symptoms in recurrent major depression: a prospective study. *Am J Psychiatr* 161, 255–261.

Oscarson, M. (2003). Pharmacogenetics of drug metabolising enzymes: importance for personalised medicine. *Clin Chem Lab Med* 41, 573–580.

Paez-Pereda, M. (2005). New drug targets in the signaling pathways activated by antidepressants. *Prog Neuropsychopharmacol Biol Psychiatry* 29, 1010–1016.

Reynolds, G.P., Templeman, L.A. and Godlewska, B.R. (2006). Pharmacogenetics of schizophrenia. *Expert Opin Pharmacother* 7, 1429–1440.

Schlicht, K., Buttner, A., Siedler, F., Scheffer, B., Zill, P., Eisenmenger, W. et al. (2006). Comparative proteomic analysis with postmortem prefrontal cortex tissues of suicide victims versus controls. *J Psychiatr Res* 41, 493–501.

Sequeira, A. and Turecki, G. (2006). Genome wide gene expression studies in mood disorders. *Omics* 10, 444–454.

Serretti, A. and Artioli, P. (2004). From molecular biology to pharmacogenetics: a review of the literature on antidepressant treatment and suggestions of possible candidate genes. *Psychopharmacology (Berl)* 174, 490–503.

Serretti, A., Artioli, P. and Quartesan, R. (2005). Pharmacogenetics in the treatment of depression: Pharmacodynamic studies. *Pharmacogenet Genomics* 15, 61–67.

Van Den, B.A., Sleegers, K., De, Z.S., Heyrman, L., Norrback, K.F., Adolfsson, R. et al. (2006). Association of brain-specific tryptophan hydroxylase, TPH2, with unipolar and bipolar disorder in a Northern Swedish, isolated population. *Arch Gen Psychiatry* 63, 1103–1110.

van Praag, H.M. (2005). Can stress cause depression? *World J Biol Psychiatr* 6(Suppl 2), 5–22.

Walther, D.J., Peter, J.U., Bashammakh, S., Hortnagl, H., Voits, M., Fink, H.et al. (2003). Synthesis of serotonin by a second tryptophan hydroxylase isoform. *Science* 299, 76.

Weissman, M.M., Wickramaratne, P., Nomura, Y., Warner, V., Pilowsky, D. and Verdeli, H. (2006). Offspring of depressed parents: 20 years later. *Am J Psychiatr* 163, 1001–1008.

Yamada, M., Yamada, M. and Higuchi, T. (2005). Antidepressant-elicited changes in gene expression Remodeling of neuronal circuits as a new hypothesis for drug efficacy. *Prog Neuropsychopharmacol Biol Psychiatr* 29, 999–1009.

Zhou, Z., Roy, A., Lipsky, R., Kuchipudi, K., Zhu, G., Taubman, J. et al. (2005). Haplotype-based linkage of tryptophan hydroxylase 2 to suicide attempt, major depression, and cerebrospinal fluid 5-hydroxyindoleacetic acid in 4 populations. *Arch Gen Psychiatr* 62, 1109–1118.

Zill, P., Baghai, T.C., Zwanzger, P., Schule, C., Eser, D., Rupprecht, R. et al. (2004a). SNP and haplotype analysis of a novel tryptophan hydroxylase isoform (TPH2) gene provide evidence for association with major depression. *Mol Psychiatr* 9, 1030–1036.

Zill, P., Buttner, A., Eisenmenger, W., Moller, H.J., Bondy, B. and Ackenheil, M. (2004b). Single nucleotide polymorphism and haplotype analysis of a novel tryptophan hydroxylase isoform (TPH2) gene in suicide victims. *Biol Psychiatr* 56, 581–586.

Clinical (Infectious Disease)

Section 9

57. Genomic Approaches to the Host Response to Pathogens

58. Host Genomics and Bacterial Infections

59. Genomics in the Evaluation and Management of Sepsis

60. Genomics and the Management of Hepatitis

This page intentionally left blank

CHAPTER

Genomic Approaches to the Host Response to Pathogens

M. Frances Shannon

INTRODUCTION

Following a period of relative disinterest in infectious disease research due to the enormous impact of vaccines and antibiotics on the spread of and mortality from these diseases, there is now renewed and growing interest in this area of research. This has been driven by several recent worldwide developments including: (i) the rising incidence of diseases such as Acquired Immune Deficiency Syndrome (AIDS) and antibiotic-resistant tuberculosis; (ii) antibiotic-resistant bacterial strains presenting a severe health threat in hospitals; (iii) the rapid spread of new pathogens such as Severe Acute Respiratory Syndrome (SARS); and (iv) the threat of bioterrorism. Indeed, nearly 25% of annual deaths worldwide are due to infectious disease (Morens et al., 2004). Thus, the need to develop new diagnostic methods, more effective vaccines and better therapeutic strategies is urgent.

In order to effectively deal with infectious disease threats, it is important to understand both the pathogen and the response of the host, since the outcome of infection is determined by complex host–pathogen interactions. Pathogens are initially detected by the surveillance cells of the innate immune system using cell surface receptors known as Toll-like receptors (TLRs) (reviewed in [Cook et al., 2004]). These TLRs recognize specific components of the pathogen, for example, bacterial lipopolysaccharide (LPS) or double-stranded (ds) RNA from viruses. While many cell types express TLRs, cells of the innate immune system such as dendritic cells (DCs) and macrophages play particularly important roles in detecting and responding to pathogens. The response of these cells to a pathogen is determined by the specific pathogen component that interacts with the TLR and the specific TLR family member that is activated. Widespread changes in gene expression are detected following TLR activation and the activated cells produce a plethora of cytokines and chemokines that then activate the adaptive arm of the immune system. The specific cytokines and chemokines produced by the TLR-activated cells tailor the response of the adaptive immune system to deal with the specific pathogen (Cook et al., 2004). Thus, the initial host response to a pathogen through the TLRs determines the outcome of the infection. Host response to infection can be a double-edged sword in that sometimes the response itself can create an adverse outcome for the host. In addition, the aberrant response of the host to self instead of foreign pathogens can create severe pathologies involving chronic inflammatory and autoimmune diseases.

The urgent need to better understand host–pathogen interactions has come at a time when genomics and related technologies are expanding rapidly. The availability of complete genomic sequences of an expanding number of pathogens, the human and mouse genome sequences and the advent of genome-wide genotyping and gene expression profiling has opened up new avenues of investigation in the field.

The genotype of the pathogen plays a major role in the response of the host to infection with more virulent pathogenic

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

strains often possessing the capability to interfere with the host immune response (Fitzgerald and Musser, 2001; Kato-Maeda et al., 2001; Schoolnik, 2002). In addition, different individuals in a population can have very different responses to a genetically identical pathogen. While there are many complex reasons for this, it is clear that part of the differential response is governed by underlying genetic differences between individuals (Clementi and Di Gianantonio, 2006; Zhang and Zhang, 2006). Studies in mouse models of infection have clearly demonstrated that these genetic differences are complex and may involve more than one genetic locus for a given susceptibility or resistance trait (e.g., [Delahaye et al., 2006]). While there are some classic examples of genetic mutations affecting the response of the host to a pathogen (e.g., malaria and sickle cell mutations) there is much to be learned before the genetics of host susceptibility is fully understood. The advent of genome-wide genotyping using single nucleotide polymorphisms (SNPs) or microsatellite markers, leading to major advances in molecular epidemiology, will revolutionize our ability to determine the complexities of the genetic component of pathogen–host interactions (Weiss and Terwilliger, 2000).

It is well known that the cells of the host immune system are activated upon detection of a pathogen by TLRs as described above. This activation process includes widespread changes in the gene expression profile of the cells with hundreds of genes being either switched on or off in response to signals generated from the pathogen-detecting TLRs. The response of individual genes has been studied in minute detail for a handful of genes and while this has produced an understanding of some aspects of host response to infection it by no means gives us the total picture. Understanding the molecular response of the host to infection has been greatly improved by using microarray-based technologies and these technologies are opening up new diagnostic possibilities as well as presenting new therapeutic options (Aderem and Smith, 2004; Bryant et al., 2004; Feezor et al., 2005; Hedeler et al., 2006; Korth and Katze, 2002; Ng et al., 2006; Ricciardi-Castagnoli, 2005; Ricciardi-Castagnoli and Granucci, 2002; Smith and Bolouri, 2005).

This chapter will focus on two aspects of the host response to pathogens where major advances are being made using genomics approaches and will describe the future impact of these approaches on the development of diagnostics and therapeutics for infectious disease. These are (i) defining the basis of genetic susceptibility to infection and (ii) the definition of the system-wide molecular response to a pathogen.

GENETIC SUSCEPTIBILITY TO PATHOGENS

It is now relatively easy to map genes associated with genetic diseases that show a Mendelian pattern of inheritance. However, these diseases account for only a very small proportion of the human disease burden and many of the more common and fatal diseases have a complex etiology with many genetic and environmental contributions. While it is clear that most complex disease has a genetic component, defining that genetic component has been difficult to date since many complex diseases such as coronary heart disease, diabetes and others are polygenic with different genetic loci contributing in major or minor ways to disease susceptibility. In addition, in different populations or under different environmental conditions, distinct but overlapping sets of genetic loci are likely to contribute. The sequencing of the human genome and the genome-wide genetic and functional mapping that has followed have raised hopes of mapping the genetic component of complex disease and there are many large efforts around the world with this aim.

Studies in both animal models and human populations have shown that infectious disease and the response of the host to a specific infection also has a complex genetic component (Clementi and Di Gianantonio, 2006; Lipoldova and Demant, 2006; Marquet et al., 1996; Mira et al., 2004). Thus, inbred mouse models have been developed that clearly show a genetic component to susceptibility for specific pathogens and in some cases at least part of the underlying genetic reason has been defined (Beck et al., 2000; Mak et al., 2001; Rogner and Avner, 2003). Mapping the genetic components of susceptibility to infection in human populations has been much more difficult due to the large natural variation in humans, the polygenic nature of this trait and the low penetrance of many of the susceptibility alleles. For infectious disease, this is complicated even more by the complex nature of the environmental influences particularly the fact that these diseases, unlike other complex diseases, are transmissible in populations. However, a combination of animal and human population studies, combined with the latest genomic technologies, is beginning to unravel the issues of genetic susceptibility to infection.

The use of inbred and congenic strains of mice are well established systems for identifying susceptibility loci (Beck et al., 2000; Rogner and Avner, 2003). In recent years genetic manipulation of specific loci by deletion or mutation has provided many mouse models for screening (Mak et al., 2001). The use of ethylnitrosourea (ENU) mutagenesis to randomly create point mutations in the mouse genome has opened up a new forward genetics approach to identifying susceptibility loci (Papathanasiou and Goodnow, 2005). This chemical mutagen, when used at appropriate doses and at the correct stage of development, can introduce single point mutations into the mouse genome. By screening libraries of mutant mice for susceptibility to specific pathogens, it should be possible to identify genetic loci that dictate susceptibility or resistance to a range of pathogens on a large scale. It is relatively straightforward to identify a chromosomal region involved in susceptibility in these mouse strains by genotyping with microsatellite markers, but identifying the specific gene that is mutated is still very time-consuming. The speed with which this can be achieved depends on the presence of candidate genes within the chromosomal interval or the ability to resequence large amounts of DNA. The latter is becoming achievable with the advent of new rapid sequencing technologies and is set to revolutionize forward genetic

approaches to disease understanding (Bennett et al., 2005; Serre and Hudson, 2006).

The recent explosion in genetic information for the human genome including the complete genome sequence and detailed genetic and physical maps has increased our ability to find variation in the human genome and correlate it with disease. Family studies, especially twin studies, and population studies have clearly shown a genetic component to susceptibility to infectious disease (Frodsham and Hill, 2004; Lipoldova and Demant, 2006; Strunk and Burgner, 2006). Susceptibility generally follows a complex pattern of inheritance and there are two main methods of mapping and identifying genetic loci involved in such complex heredity. These are either association studies or linkage studies. Association studies involve screening populations for specific mutations in a candidate gene(s) in case–control studies or in family studies. This type of approach identified the link between Human Immunodeficiency Virus (HIV) resistance and the chemokine receptor, CCR5 described below (Dean et al., 1996; Samson et al., 1996). Genome-wide association studies although still quite expensive are now becoming feasible and being used to define linkage between specific markers and susceptibility. These linkage studies depend on the availability of markers and the density of these markers is rapidly increasing with the large scale identification of new SNPs across the human genome. The selection of which SNPs to use and the large numbers of samples needed to generate statistically significant associations for low penetrance alleles are still challenges. The Haplotype MAP (HapMap) project is starting to identify haplotypes within different population groups and together with improvements in large scale genotyping technology and bioinformatics should be useful in studies of complex disease inheritance. Table 57.1 summarizes the best studied genetic susceptibility loci for response to different infectious agents in both mouse models and human studies.

One of the classical examples of genetic susceptibility to infection is the role of the hemaglobinopathies in the outcome of malaria infection (Patrinos et al., 2005). There are also certain chromosomal regions and families of genes that have attracted attention in terms of searching for susceptibility alleles or polymorphisms. Because the TLR family of receptors plays a major role in recognizing pathogens, it was speculated that genetic variation in these receptors or their signaling pathways might be responsible for some susceptibility phenotypes (reviewed in [Schroder and Schumann, 2005]). One of the best examples to date is the occurrence of a single polymorphism in the region of the human TLR4 gene encoding the extracellular domain of the

Table 57.1 List of well-described susceptibility loci for resistance or susceptibility to infectious disease.

Pathogen	Genes	References
HIV	*CCR5*	Dean et al. (1996); Samson et al. (1996)
	HLA Class I	Hendel et al. (1999); Li et al. (2007); Selvaraj et al. (2006)
	CCR2	Magierowska et al. (1999); Su et al. (1999)
Malaria	Globin locus	Reviewed in Patrinos et al. (2005)
	HLA Class I	Migot-Nabias et al. (2001); Young et al. (2005); Reviewed in Hill (1996, 1999)
	TNF-α promoter	McGuire et al., (1994); Ubalee et al. (2001)
Leprosy	*TLR2*	Kang and Lee (2002); Bochud et al. (2003)
	HLA Class II	Shaw et al. (2001); Mehra et al. (1995)
	TNF-α promoter	Roy et al. (1997); Shaw et al. (2001)
Legionella	*TLR5*	Hawn et al. (2003); Merx et al. (2006)
Tuberculosis	HLA Class I	Lombard et al. (2006); Vijaya Lakshmi et al. (2006)
	NRAMP1/Slc11a1	Kusuhara et al. (2007); Li et al. (2006)
Typhoid fever	HLA Class II	Dunstan et al. (2001); Dharmana et al. (2002)
Leishmania	*NRAMP1/Slc11a1*	Bucheton et al. (2003); Mohamed et al. (2004)
	HLA	Reviewed in Lipoldova and Demant (2006)
	TNF	Bucheton et al. (2003)
	IFNgR1	Mohamed et al. (2003)
	IL-4	Mohamed et al. (2003)
Salmonella	*NRAMP1/Slc11a1*	Sebastiani et al. (1998)
Inhaled *E. coli* LPS	*TLR4*	Arbour et al. (2000); Feterowski et al. (2003)
Pyogenic bacteria	*IRAK4*	Picard et al. (2003)

This lists includes genes identified in both mouse and human studies.

receptor which confers reduced sensitivity to inhaled *Escherichia coli* (*E. coli*) LPS (Arbour et al., 2000). Interestingly, when septic shock patients were compared with a control group, these lower-responding alleles were found only in the septic shock group and these individuals had a higher incidence of Gram-negative bacterial infection (Feterowski et al., 2003). Such studies need further confirmation since there are also a number of studies that failed to find any linkage between TLR4 mutations and response to various infections (Schroder and Schumann, 2005). There is also enormous variation in the response of individuals to LPS even in the absence of TLR4 mutations implying that variation may occur in other components of the TLR4 signaling system. An example of this is the link between IRAK4 mutations and increased susceptibility to infection with pyogenic bacteria (Picard et al., 2003). Variation in other TLR genes has also been associated with disease susceptibility. For example, a mutation in the extracellular domain of TLR2 is linked to susceptibility to leprosy (Alcais et al., 2005) and a mutation in TLR5 increases susceptibility to *Legionella* (Hawn et al., 2003). Taken together these data support the idea that variation in the innate immune recognition of pathogens play an important part in governing susceptibility to an array of infectious diseases. However, caution needs to be exercised until larger population groups have been studied.

The extensive polymorphism at the chromosomal regions encoding major histocompatibility complex (MHC) proteins is thought to have arisen through natural selection in response to selective pressure from infectious disease. Although human leukocyte antigen (HLA) association with resistance or susceptibility to infectious disease has been difficult to identify because of the complex array of antigenic epitopes involved, a number of studies have implicated this locus in genetic susceptibility to infectious disease (Ghodke et al., 2005; Little and Parham, 1999). MHC molecules fall into two classes, Class I that present foreign antigens to CD8+ cytotoxic T cells and Class II that play a similar role for CD4+ helper T cells. Variation in specific Class I genes has been shown to confer susceptibility to pulmonary tuberculosis and to HIV whereas mutations in other Class I genes confer resistance to HIV and to severe malaria. Class II mutations that confer resistance to hepatitis B or hepatitis C have been identified and susceptibility to typhoid fever and leprosy are also associated with specific Class II mutations. Further molecular analysis of these and other associations may in the future have an impact on the development of new vaccines and immunotherapeutics.

To date the most successful manner of identifying susceptibility genes in human populations has been the candidate gene approach. Candidate genes have emerged from many sources including mouse genetic studies as well as biochemical and function dissection of the immune system. Once a candidate gene is identified, the chromosomal region spanning this gene in the human genome is then scanned for the occurrence of specific mutations or for functional polymorphisms in case–control studies across populations or in linkage studies in family groups. Such studies have identified a number of well described susceptibility loci for infection with various pathogens. One of the most heralded example was the identification of a deletion in the chemokine receptor, CCR5, which was shown to confer resistance to HIV infection (Dean et al., 1996; Samson et al., 1996). Biochemically, this can be explained by the fact that CCR5 is a co-receptor for HIV on the surface of T cells (Dragic et al., 1996). A mutation in another chemokine receptor, CCR2, has also been shown to confer HIV resistance in certain Caucasian populations (O'Brien and Moore, 2000).

Some genes have been associated with susceptibility or resistance to multiple pathogens. For example, variation in the *NRAMP1/Slc11a1* gene is associated with susceptibility to *Leishmania* and to specific intracellular bacteria such as tuberculosis (Barton et al., 1999; Govoni et al., 1996; Lipoldova and Demant, 2006; Sebastiani et al., 1998). Mutations in the tumor necrosis factor (TNF) locus, mainly gene promoter mutations, have been linked with malaria and leprosy susceptibility (Lipoldova and Demant, 2006). Gene promoter or control region mutations have an impact on the level of protein produced from the gene rather than the function of the protein. This is an area of great interest but more difficult to study for several reasons, including the inability to identify control regions simply from sequence information and the complexity and flexibility of transcriptional control. A recent review detailing the genes associated with *Leishmania* susceptibility describes a number of genes that can affect disease outcome including the *interferon-gamma receptor type 1* (*IFNGR1*), the *interleukin-4 (IL-4*) gene and the *NRAMP-1/Slc11a1* gene (Lipoldova and Demant, 2006). These genes and others such as *interleukin-12* (*IL-12*) and its receptor are also linked with *Salmonella* and certain mycobacterial infections (Lipoldova and Demant, 2006). Thus, it is likely that variation in many genes can contribute to disturbing the finely balanced tuning of the immune system and lead to an altered response to a pathogen. It is clear from such studies that the same genes may be involved in susceptibility to an array of pathogens indicating a core immune response critical for any pathogen.

The identification of susceptibility loci for infection with various pathogens will aid in developing new diagnostic screens based on the detection of genetic variants in these loci. It could be envisaged that a person's susceptibility or resistance to a pathogen could be defined by a simple genotyping screen either prior to exposure to any pathogen or upon presentation with an infection. It may also be possible to determine the likely outcome of the infection through a genotyping screen. The definition of susceptibility loci will also contribute to our ability to develop new vaccines and therapeutics.

EXPLORING THE HOST RESPONSE THROUGH EXPRESSION PROFILING

The use of microarray technology to generate expression profiling data is becoming common place in biomedical research (reviewed in [Quackenbush, 2002; Sherlock, 2000]). This

technology allows the documentation of mRNA levels for thousands of genes from total RNA prepared from cells or tissue samples. The data obtained can be compared from sample to sample allowing the changes between samples to be documented and quantified. The "expression profile" for any cell or tissue is simply the list of genes whose expression can be detected using microarrays. The differences in the expression profile from one cell or tissue to the next or in cells treated in a specific manner is a surrogate measure of the cell/tissue phenotype and shows how that phenotype responds to its environment. Expression profiling is most useful when large datasets become available and when the data is combined with other data types and detailed bioinformatics studies. For example, using functional clustering of expression profiling data can help identify pathways that are important for a particular process and co-expression clustering combined with other technologies can help define regulatory networks within the cell.

Over the last 5–6 years, this technology has been applied to identifying the changes in gene expression that occur in response to infection by various pathogens (Aderem and Smith, 2004; Boyce et al., 2004; Bryant et al., 2004; Feezor et al., 2005; Foti et al., 2006; Jenner and Young, 2005; Korth and Katze, 2002; Korth et al., 2005; Ricciardi-Castagnoli and Granucci, 2002). To date there are more than 150 papers in the literature that describe gene expression changes that occur in response to infection with a plethora of pathogens and in many cell types (reviewed in [Jenner and Young, 2005]). Many of these are *in vitro* studies, taking specific cell types and infecting them with specific agents including bacteria, viruses, parasites and yeasts. In addition, cellular responses to bacterial components have also been documented, helping to identify pathogen-specific responses as well as determining the pathogenic component responsible for the major gene expression effects. Virulent or non-virulent strains of specific pathogens as well as mutant organisms have been used to determine the gene expression profile associated with a negative or positive clinical outcome. Few *in vivo* studies have been carried out and have shed light on the more complex responses seen in whole animals and helped to validate the *in vitro* data.

Identifying a Common Host Response to Infection

Several pioneering studies demonstrated that microarrays could be used to determine changes in the gene expression profile of cells in response to virus or bacterial infection (Boldrick et al., 2002; Gao et al., 2002; Huang et al., 2001; Nau et al., 2002). These studies paved the way for the analysis of the host response to a wide variety of infectious agents. The most significant of these studies compared the response of macrophages or DCs to a variety of infectious agents in a single study. In these studies a strong shared response to all infections, be they bacterial, viral or parasitic in nature, was identified. Not only was there commonality from one infectious agent to another but there was also some commonality across cell types. This expression signature has been interpreted as a general "alarm signal" for infection (reviewed in [Jenner and Young, 2005]). Studies of infection with Gram-positive and Gram-negative bacteria also revealed a common expression signature in peripheral blood mononuclear cells. Recently, the Young lab has interrogated all of the publicly available expression profiling data related to the host response to infection (Jenner and Young, 2005). The dataset includes 785 experiments in cells ranging from macrophages and DCs to cells of the adaptive immune response, endothelial and epithelial cells and spanning a wide range of infecting agents. This meta-analysis revealed that a "common host response" can be detected across all of these cell types and infectious agents and show that although cells such as macrophages and DCs specialize in detecting infection, other cells of the body can mount the same "alarm response" as described above (Figure 57.1).

Not surprisingly this expression signature contains many genes associated with the immune system particularly those encoding inflammatory cytokines, chemokines and their receptors. However, some more surprising patterns of expression were also detected. It has been long known that interferon-stimulated genes (ISGs) are regulated by virus infection, but it has only recently been recognized that bacteria and other infecting agents can also illicit the interferon response. This was borne out in these meta-analyses where upregulation of an ISG set is observed across a broad range of infecting agents and cell types.

Not only do these cells change the expression of secreted factors and their receptors during infection but the intracellular milieu is also modified. Once again, there is a common pattern of change observed in all of these studies. The upregulation of signaling and transcription pathways that both augment and attenuate the immune response are observed leading to the interpretation that both positive and negative feedback loops operate within the cell to heighten or dampen the immune

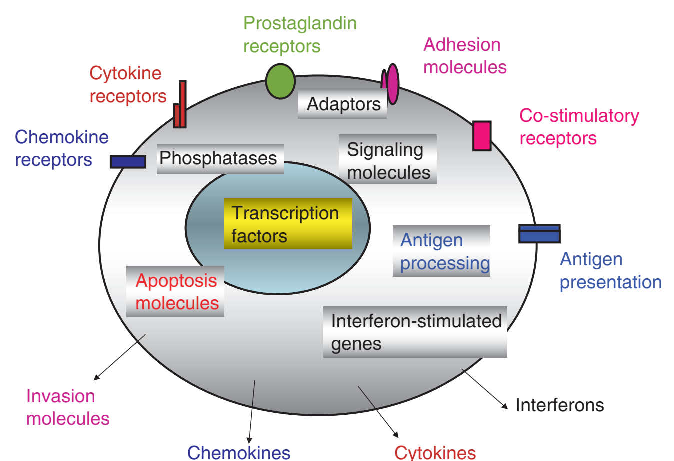


Figure 57.1 Summary of the common host response to infection. This figure is adapted from Jenner and Young (2005) and summarizes their meta-analysis of gene expression changes in response to infection with various pathogens or activation of a variety of cells with agents that mimic infection. Genes are grouped into general functional categories.

response. Temporal profiling can reveal extra layers of complexity and in one study a pro-inflammatory profile followed by an anti-inflammatory profile was identified in macrophages activated with LPS (Wells et al., 2005). Meta-analysis of temporal studies of activation through different TLRs also revealed that the inflammatory chemokines/cytokine signature was an early response while the ISG response was later, presumably reflecting the need for an indirect activation of the ISGs through interferon production (Jenner and Young, 2005). Changes in the expression level of genes involved in both activation and repression of apoptosis also fall under the "common signature" banner and this is interpreted as sending the cells into a state of high alert where apoptosis can be either initiated to eliminate infected cells or terminated if the infection resolves (Jenner and Young, 2005).

Although these *in vitro* studies have provided an overview of the response of isolated cell types to pathogenic infection, *in vivo* studies are needed to validate any of these results before application to clinical medicine. A number of animal models have been used to profile the host response to infection with a variety of agents. These studies are complicated especially if whole tissue samples are used in that changes in gene expression can result not only from genuine changes within the cells of the tissue but also from the recruitment especially of immune cells into the infected or inflamed tissue. Nevertheless, *in vivo* studies have, in some cases, shown good correlation with the expression profiles found from *in vitro* studies. For example, profiling the brains of mice infected with virulent Sindbis virus revealed that ISGs as well as inflammatory chemokines were upregulated (Johnston et al., 2001) and in terms of diagnosis or treatment the exact reason behind these changes in expression signature may be irrelevant. *In vivo* studies also have some advantages in that the gene expression profiles detected in infected tissues will often make more sense when combined with other physiological or cell biology data from studies of the infected host. Thus an "infection signature" would not only describe the altered gene expression of the immune cells that are recruited to the sites of infection but also would include changes in the gene expression of the resident cells of the tissue and may provide a more robust profile of the infection process for use in diagnostic applications.

What can be applied to clinical medicine from these studies? The ability to detect an "infection signature" using focused microarrays could potentially be used as a diagnostic tool. There would be an immediate need to identify a core set of genes with sufficiently robust changes in gene expression to form the basis of a diagnostic array. Arraying technology would need to be priced for diagnostic use and the technology would have to be deemed sufficiently robust to pass all the quality control requirements of a diagnostic laboratory. No doubt progress will be made toward these goals in the near future.

Pathogen-Specific Responses

In addition to the "common host response" described above, microarray studies have revealed that pathogen-specific responses also exist. This is not surprising since it has long been known that different pathogens induce distinct arms of the adaptive immune response. For example, distinct types of helper T cells are activated in response to bacterial and viral infection compared to parasite infection and DC products such as cytokines and chemokines control the differentiation of T helper cell into the correct response type (Mosmann et al., 2005). Thus, the fact that cells of the innate immune system display a pathogen-specific transcriptional response as well as a general alarm signal helps to dictate the subsequent immune response.

Different TLRs are involved in recognizing and responding to different pathogens. For example, TLR2 is responsible for activation in response to Gram-positive bacteria while TLR4 responds to LPS, a component of Gram-negative bacteria (Beutler and Rietschel, 2003; Cook et al., 2004). TLR3 responds to dsRNA and thus dictates the viral immune response for many dsRNA viruses (Beutler and Rietschel, 2003; Cook et al., 2004). Studies using bacterial components such as LPS and flagellin as well as dsRNA have revealed that each TLR induces a specific as well as general transcriptional response. The transcriptional response of macrophages and peripheral blood mononuclear cells is more robust in response to Gram-negative (TLR4) compared with Gram-positive (TLR2) bacteria and the ISG response is considerably reduced for the Gram-positive expression signature (Boldrick et al., 2002; Nau et al., 2002). These differences are further observed when bacterial components are used to activate cells. For example, LPS from Gram-negative bacteria, a specific ligand for TLR4, can activate the ISG profile but TLR2 ligands such as LTA and MDP cannot (Jenner and Young, 2005). Not only the activation of specific gene sets but also the strength of the specific response signature may be important for the immune detection of the type of pathogen involved. *E. coli* infection of DCs strongly upregulates the chemokines/cytokine inflammatory cluster whereas infection with influenza or other single stranded (ss) RNA viruses (through TLR7) has a weaker ability to regulate this cluster but a stronger ability to regulate the ISG signature (Huang et al., 2001; Lund et al., 2004).

These types of results raise the possibility that the diagnosis of the type of pathogen involved in an infection would be helped by the development of customized microarrays that could distinguish the gene expression profiles elicited by particular pathogens. Additionally, arrays that also detect RNAs produced by the pathogen may be even more significant as a diagnostic tool.

Determining the Outcome of the Infection

An infecting agent can either be cleared from the body by the immune system mounting an appropriate response or cause severe or terminal pathology. The outcome depends on a multiplicity of events ranging from the genotype of the host, that is, whether the host displays a resistant or susceptible phenotype, the genotype of the infectious agent, that is, virulent or non-virulent strains and many other less defined environmental factors. Can genomic approaches be used to determine the outcome of infection? Clearly, as discussed above, the technology is developing to define susceptible and resistant host genotypes especially in animal models of infection but the ability to do this routinely in human populations is some way into the future.

Given the smaller genomes of pathogenic organisms, defining virulence genotypes is progressing at a faster rate (Chan, 2003; Dorrell et al., 2005; Fitzgerald and Musser, 2001; Kato-Maeda et al., 2001; MacFarlane et al., 2005; Schoolnik, 2002; Zhang and Zhang, 2006).

Expression profiling studies have been used to investigate the differences in the host response to pathogenic and non-pathogenic strains of specific infectious agents. In one example, mice infected with a pathogenic strain of pneumonia virus upregulated the expected inflammatory chemokines/cytokine profile as well as the ISG profile but an attenuated strain of the same virus could not, although the virus replicated in the lungs of these mice to the same degree (Domachowske et al., 2001). Temporal profiling of the infection process in animals will also help to define the expression signatures associated with the ability of the host to clear specific pathogens.

2009 UPDATE

With the ongoing threat of new or reemerging pathogens, the use of genomic approaches to understand the host response to infectious agents continues to gain momentum. The sequencing of viral pathogens has become routine and there are now more than 800 bacterial genome sequences publicly available (http://www.ncbi.nlm.nih.gov/genomes). This combined with the availability of the human genome and that of an increasing number of model organisms is allowing deeper insight into genomic aspects of host–pathogen interactions.

Genetic susceptibility or resistance to pathogens is still an area of interest with new genetic associations being routinely identified. A number of new studies have again focused on the TLR gene family and their role in susceptibility to infection as well as disease progression. An excellent analysis of all the previous genetic association and functional studies on TLR4 has recently been published (Ferwerda et al., 2008). Some but not all of the association studies indicate a role for TLR4 polymorphisms in pathogen susceptibility (strongest for RSV infection), but many of these studies are small and lack statistical power and the larger studies tend not to find strong associations (Ferwerda et al., 2008). Similarly, functional studies do not always support an important role for TLR4 polymorphisms in controlling cellular responses such as cytokine production (Ferwerda et al., 2008). There are a number of recent association studies that have shown a link between polymorphisms in other TLRs, including TLRs 7, 8 and 9, and HIV susceptibility or disease progression (Bochud et al., 2007; Oh et al., 2007; Soriano-Sarabia et al., 2008). However, these studies will require further verification with larger genetic association studies and more detailed functional studies. Similarly, polymorphisms in the TLR genes continue to emerge as potential players in susceptibility to other pathogens such as TB (Ma et al., 2007). A relatively large association study (1312 individuals) combined with some functional studies revealed a correlation between polymorphisms in the TLR6, TLR1 and TLR10 genes, and the occurrence of TB in certain ethnic populations (Ma et al., 2007). Recent studies of the TLR adaptor protein, TIRAP/MAL, suggest genetic associations between polymorphisms in this protein and TB susceptibility (Hawn et al., 2006; Khor et al., 2007), again implicating the innate immune recognition system in controlling infectious disease responses. For a summary of the literature on genetic association and TB, the reader is referred to a recent review by Berrington and Hawn (Berrington and Hawn, 2007) and for a similar summary of TLR genetic association studies with various pathogens to Misch and Hawn (Misch and Hawn, 2008). It will be very important to clarify the role of the TLR recognition system in pathogen susceptibility and disease progression before any of this information can be applied to clinical screening or drug/vaccine development.

Of particular note, the first whole genome association study to identify the host determinants of HIV-1 infection was recently published (Fellay et al., 2007). Two polymorphisms associated with viral load during the asymptomatic period of infection were located near major histocompatibility allele human leukocyte antigen loci. A second component of the study, examining the time of HIV disease progression, implicated two genes one of which encoded an RNA polymerase I subunit (Fellay et al., 2007). It will be important to follow up these genetic association studies with functional studies of the genes or loci implicated. Given the current availability of genomics resources, we are likely to see many more genome-wide association studies published in this area, a welcome addition to the candidate gene approach.

The recently identified SARS virus has already been the subject of several small scale genetic association studies, especially in relation to the HLA loci. While some associations have been identified, they have not been replicated in different populations and so remain tentative (reviewed in Yang et al., 2008). SNP associations in other genes such as that encoding low serum mannose binding lectin (MBL) and the promoter of the RANTES gene have been replicated in different populations and may stand the test of time (Ng et al., 2007; Zhang et al., 2005).

Expression profiling continues to be a useful tool in the analysis of the host–pathogen interaction. Some examples include a study of *in vivo* human rhinovirus infection (Proud et al., 2008), innate host reponses to Ebola virus and the consequences of a mutation in the virus VP35 protein (Hartman et al., 2008) and Leishmania infection of human macrophages (Guerfali et al., 2008). These and many other studies are leading to the identification of general molecular consequences

of infection as well as the specific consequences of individual pathogens. A recent review outlines the many insights gained into virus–host interaction using functional genomics as well as the use of these approaches in developing and evaluating vaccines (Katze et al., 2008). Studies in model organisms such as *C elegans* support many of the findings in humans (Wong et al., 2007) and provide a useful tool for detailed functional investigation. The "systems biology" approach to study infectious disease is being extended by the use of proteomic profiling. Two recent articles illustrate the usefulness of this approach in examining the response of a human cell line to the avian H9N2 influenza virus (Liu et al., 2008) and of human macrophages to *M tuberculosis* lipids (Shui et al., 2009). Systems biology approaches are moving rapidly in this and other biomedical fields and no doubt the next few years, given the advent of deep sequencing, will see enormous new strides in understanding host–pathogen interactions (for a recent review in relation to virus infection see Tan et al., 2007). Other developments will undoubtedly lead to easier application of these technologies to the clinic in the foreseeable future.

REFERENCES

Berrington, W.R. and Hawn, T.R. (2007). *Mycobacterium tuberculosis*, macrophages, and the innate immune response: Does common variation matter? *Immunol Rev* 219, 167–186.

Bochud, P.Y. et al. (2007). Polymorphisms in Toll-like receptor 9 influence the clinical course of HIV-1 infection. *AIDS* 21(4), 441–446.

Fellay, J. et al. (2007). A whole-genome association study of major determinants for host control of HIV-1. *Science* 317(5840), 944–947.

Ferwerda, B. et al. (2008). Functional consequences of toll-like receptor 4 polymorphisms. *Mol Med* 14(5–6), 346–352.

Guerfali, F.Z. et al. (2008). Simultaneous gene expression profiling in human macrophages infected with Leishmania major parasites using SAGE. *BMC Genomics* 9, 238.

Hartman, A.L. et al. (2008). Whole-genome expression profiling reveals that inhibition of host innate immune response pathways by Ebola virus can be reversed by a single amino acid change in the VP35 protein. *J Virol* 82(11), 5348–5358.

Hawn, T.R. et al. (2006). A polymorphism in Toll-interleukin 1 receptor domain containing adaptor protein is associated with susceptibility to meningeal tuberculosis. *J Infect Dis* 194(8), 1127–1134.

Katze, M.G. et al. (2008). Innate immune modulation by RNA viruses: Emerging insights from functional genomics. *Nat Rev Immunol* 8(8), 644–654.

Khor, C.C. et al. (2007). A Mal functional variant is associated with protection against invasive pneumococcal disease, bacteremia, malaria and tuberculosis. *Nat Genet* 39(4), 523–528.

Liu, N. et al. (2008). Proteomics analysis of differential expression of cellular proteins in response to avian H9N2 virus infection in human cells. *Proteomics* 8(9), 1851–1858.

Ma, X. et al. (2007). Full-exon resequencing reveals toll-like receptor variants contribute to human susceptibility to tuberculosis disease. *PLoS ONE* 2(12), e1318.

Misch, E.A. and Hawn, T.R. (2008). Toll-like receptor polymorphisms and susceptibility to human disease. *Clin Sci* 114(5), 347–360.

Ng, M.W. et al. (2007). The association of RANTES polymorphism with severe acute respiratory syndrome in Hong Kong and Beijing Chinese. *BMC Infect Dis* 7, 50.

Oh, D.Y. et al. (2008). A functional toll-like receptor 8 variant is associated with HIV disease restriction. *J Infect Dis* 198(5), 701–709.

Proud, D. et al. (2008). Gene expression profiles during in vivo human rhinovirus infection: Insights into the host response. *Am J Respir Crit Care Med* 178(9), 962–968.

Shui, W. et al. (2009). Quantitative proteomic profiling of host–pathogen interactions: The macrophage response to *Mycobacterium tuberculosis* lipids. *J Proteome Res* 8(1), 282–289.

Soriano-Sarabia, N. et al. (2008). Influence of the Toll-like receptor 9 1635A/G polymorphism on the CD4 count, HIV viral load, and clinical progression. *J Acquir Immune Defic Syndr* 49(2), 128–135.

Tan, S.L. et al. (2007). Systems biology and the host response to viral infection. *Nat Biotechnol* 25(12), 1383–1389.

Wong, D. et al. (2007). Genome-wide investigation reveals pathogen-specific and shared signatures in the response of *Caenorhabditis elegans* to infection. *Genome Biol* 8(9), R194.

Yang, X. et al. (2008). Infectious disease in the genomic era. *Annu Rev Genomics Hum Genet* 9, 21–48.

Zhang, H. et al. (2005). Association between mannose-binding lectin gene polymorphisms and susceptibility to severe acute respiratory syndrome coronavirus infection. *J Infect Dis* 192(8), 1355–1361.

ACKNOWLEDGEMENT

I would like to thank Stephanie Palmer for help with preparation and proof-reading of the manuscript.

REFERENCES

Aderem, A. and Smith, K.D. (2004). A systems approach to dissecting immunity and inflammation. *Semin Immunol* 16(1), 55–67.

Alcais, A., Mira, M. et al. (2005). Genetic dissection of immunity in leprosy. *Curr Opin Immunol* 17(1), 44–48.

Arbour, N.C., Lorenz, E. et al. (2000). TLR4 mutations are associated with endotoxin hyporesponsiveness in humans. *Nat Genet* 25(2), 187–191.

Barton, C.H., Biggs, T.E. et al. (1999). Nramp1: A link between intracellular iron transport and innate resistance to intracellular pathogens. *J Leukoc Biol* 66(5), 757–762.

Beck, J.A., Lloyd, S. et al. (2000). Genealogies of mouse inbred strains. *Nat Genet* 24(1), 23–25.

Bennett, S.T., Barnes, C. et al. (2005). Toward the 1,000 dollars human genome. *Pharmacogenomics* 6(4), 373–382.

Beutler, B. and Rietschel, E.T. (2003). Innate immune sensing and its roots: The story of endotoxin. *Nat Rev Immunol* 3(2), 169–176.

Bochud, P.Y., Hawn, T.R. et al. (2003). Cutting edge: A Toll-like receptor 2 polymorphism that is associated with lepromatous leprosy is unable to mediate mycobacterial signaling. *J Immunol* 170(7), 3451–3454.

Boldrick, J.C., Alizadeh, A.A. et al. (2002). Stereotyped and specific gene expression programs in human innate immune responses to bacteria. *Proc Natl Acad Sci USA* 99(2), 972–977.

Boyce, J.D., Cullen, P.A. et al. (2004). Genomic-scale analysis of bacterial gene and protein expression in the host. *Emerg Infect Dis* 10(8), 1357–1362.

Bryant, P.A., Venter, D. et al. (2004). Chips with everything: DNA microarrays in infectious diseases. *Lancet Infect Dis* 4(2), 100–111.

Bucheton, B., Abel, L. et al. (2003). Genetic control of visceral l eishmaniasis in a Sudanese population: Candidate gene testing indicates a linkage to the NRAMP1 region. *Genes Immun* 4(2), 104–109.

Chan, V.L. (2003). Bacterial genomes and infectious diseases. *Pediatr Res* 54(1), 1–7.

Clementi, M. and Di Gianantonio, E. (2006). Genetic susceptibility to infectious diseases. *Reprod Toxicol* 21(4), 345–349.

Cook, D.N., Pisetsky, D.S. et al. (2004). Toll-like receptors in the pathogenesis of human disease. *Nat Immunol* 5(10), 975–979.

Cook, D.N., Wang, S. et al. (2004). Genetic regulation of endotoxin-induced airway disease. *Genomics* 83(6), 961–969.

Dean, M., Carrington, M. et al. (1996). Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. *Science* 273(5283), 1856–1862.

Delahaye, N.F., Coltel, N. et al. (2006). Gene-expression profiling discriminates between cerebral malaria (CM)-susceptible mice and CM-resistant mice. *J Infect Dis* 193(2), 312–321.

Dharmana, E., Joosten, I. et al. (2002). HLA-DRB1*12 is associated with protection against complicated typhoid fever, independent of tumour necrosis factor alpha. *Eur J Immunogenet* 29(4), 297–300.

Domachowske, J.B., Bonville, C.A. et al. (2001). Glucocorticoid administration accelerates mortality of pneumovirus-infected mice. *J Infect Dis* 184(12), 1518–1523.

Dorrell, N., Hinchliffe, S.J. et al. (2005). Comparative phylogenomics of pathogenic bacteria by microarray analysis. *Curr Opin Microbiol* 8(5), 620–626.

Dragic, T., Litwin, V. et al. (1996). HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature* 381(6584), 667–673.

Dunstan, S.J., Stephens, H.A. et al. (2001). Genes of the class II and class III major histocompatibility complex are associated with typhoid fever in Vietnam. *J Infect Dis* 183(2), 261–268.

Feezor, R.J., Cheng, A. et al. (2005). Functional genomics and gene expression profiling in sepsis: Beyond class prediction. *Clin Infect Dis* 41(Suppl 7), S427–sS435.

Feterowski, C., Emmanuilidis, K. et al. (2003). Effects of functional Toll-like receptor-4 mutations on the immune response to human and experimental sepsis. *Immunology* 109(3), 426–431.

Fitzgerald, J.R. and Musser, J.M. (2001). Evolutionary genomics of pathogenic bacteria. *Trends Microbiol* 9(11), 547–553.

Foti, M., Granucci, F. et al. (2006). Dendritic cells in pathogen recognition and induction of immune responses: A functional genomics approach. *J Leukoc Biol* 79(5), 913–916.

Frodsham, A.J. and Hill, A.V. (2004). Genetics of infectious diseases. *Hum Mol Genet* 13(Spec No 2), R187–R194.

Gao, J.J., Diesl, V. et al. (2002). Regulation of gene expression in mouse macrophages stimulated with bacterial CpG-DNA and lipopolysaccharide. *J Leukoc Biol* 72(6), 1234–1245.

Ghodke, Y., Joshi, K. et al. (2005). HLA and disease. *Eur J Epidemiol* 20(6), 475–488.

Govoni, G., Vidal, S. et al. (1996). The Bcg/Ity/Lsh locus: genetic transfer of resistance to infections in C57BL/6J mice transgenic for the Nramp1 Gly169 allele. *Infect Immun* 64(8), 2923–2929.

Hawn, T.R., Verbon, A. et al. (2003). A common dominant TLR5 stop codon polymorphism abolishes flagellin signaling and is associated with susceptibility to legionnaires' disease. *J Exp Med* 198(10), 1563–1572.

Hedeler, C., Paton, N.W. et al. (2006). A classification of tasks for the systematic study of immune response using functional genomics data. *Parasitology* 132(Pt 2), 157–167.

Hendel, H., Caillat-Zucman, S. et al. (1999). New class I and II HLA alleles strongly associated with opposite patterns of progression to AIDS. *J Immunol* 162(11), 6942–6946.

Hill, A.V. (1996). Genetic susceptibility to malaria and other infectious diseases: From the MHC to the whole genome. *Parasitology* 112(Suppl), S75–S84.

Hill, A.V. (1999). The immunogenetics of resistance to malaria. *Proc Assoc Am Physicians* 111(4), 272–277.

Huang, Q., Liu, D. et al. (2001). The plasticity of dendritic cell responses to pathogens and their components. *Science* 294(5543), 870–875.

Jenner, R.G. and Young, R.A. (2005). Insights into host responses against pathogens from transcriptional profiling. *Nat Rev Microbiol* 3(4), 281–294.

Johnston, C., Jiang, W. et al. (2001). Identification of genes involved in the host response to neurovirulent alphavirus infection. *J Virol* 75(21), 10431–10445.

Kang, T.J., Lee, S.B. et al. (2002). A polymorphism in the toll-like receptor 2 is associated with IL-12 production from monocyte in lepromatous leprosy. *Cytokine* 20(2), 56–62.

Kato-Maeda, M., Gao, Q. et al. (2001). Microarray analysis of pathogens and their interaction with hosts. *Cell Microbiol* 3(11), 713–719.

Korth, M.J., Kash, J.C. et al. (2005). Virus infection and the interferon response: A global view through functional genomics. *Methods Mol Med* 116, 37–55.

Korth, M.J. and Katze, M.G. (2002). Unlocking the mysteries of virus-host interactions: Does functional genomics hold the key?. *Ann NY Acad Sci* 975, 160–168.

Kusuhara, K., Yamamoto, K. et al. (2007). Association of IL12RB1 polymorphisms with susceptibility to and severity of tuberculosis in Japanese: A gene-based association analysis of 21 candidate genes. *Int J Immunogenet* 34(1), 35–44.

Li, H.T., Zhang, T.T. et al. (2006). SLC11A1 (formerly NRAMP1) gene polymorphisms and tuberculosis susceptibility: A meta-analysis. *Int J Tuberc Lung Dis* 10(1), 3–12.

Li, S., Jiao, H. et al. (2007). Human leukocyte antigen class I and class II allele frequencies and HIV-1 infection associations in a Chinese cohort. *J Acquir Immune Defic Syndr* 44(2), 121–131.

Lipoldova, M. and Demant, P. (2006). Genetic susceptibility to infectious disease: Lessons from mouse models of leishmaniasis. *Nat Rev Genet* 7(4), 294–305.

Little, A.M. and Parham, P. (1999). Polymorphism and evolution of HLA class I and II genes and molecules. *Rev Immunogenet* 1(1), 105–123.

Lombard, Z., Dalton, D.L. et al. (2006). Association of HLA-DR, -DQ, and vitamin D receptor alleles and haplotypes with tuberculosis in the Venda of South Africa. *Hum Immunol* 67(8), 643–654.

Lund, J.M., Alexopoulou, L. et al. (2004). Recognition of single-stranded RNA viruses by Toll-like receptor 7. *Proc Natl Acad Sci USA* 101(15), 5598–5603.

MacFarlane, R., Bhattacharya, D. et al. (2005). Genomic DNA microarrays for Entamoeba histolytica: Applications for use in expression profiling and strain genotyping. *Exp Parasitol* 110(3), 196–202.

Magierowska, M., Theodorou, I. et al. (1999). Combined genotypes of CCR5, CCR2, SDF1, and HLA genes can predict the long-term nonprogressor status in human immunodeficiency virus-1-infected individuals. *Blood* 93(3), 936–941.

Mak, T.W., Penninger, J.M. et al. (2001). Knockout mice: A paradigm shift in modern immunology. *Nat Rev Immunol* 1(1), 11–19.

Marquet, S., Abel, L. et al. (1996). Genetic localization of a locus controlling the intensity of infection by *Schistosoma mansoni* on chromosome 5q31-q33. *Nat Genet* 14(2), 181–184.

McGuire, W., Hill, A.V. et al. (1994). Variation in the TNF-alpha promoter region associated with susceptibility to cerebral malaria. *Nature* 371(6497), 508–510.

Mehra, N.K., Rajalingam, R. et al. (1995). Variants of HLA-DR2/DR51 group haplotypes and susceptibility to tuberculoid leprosy and pulmonary tuberculosis in Asian Indians. *Int J Lepr Other Mycobact Dis* 63(2), 241–248.

Merx, S., Zimmer, W. et al. (2006). Characterization and functional investigation of single nucleotide polymorphisms (SNPs) in the human TLR5 gene. *Hum Mutat* 27(3), 293.

Migot-Nabias, F., Luty, A.J. et al. (2001). HLA alleles in relation to specific immunity to liver stage antigen-1 from plasmodium falciparum in Gabon. *Genes Immun* 2(1), 4–10.

Mira, M.T., Alcais, A. et al. (2004). Susceptibility to leprosy is associated with PARK2 and PACRG. *Nature* 427(6975), 636–640.

Mohamed, H.S., Ibrahim, M.E. et al. (2003). Genetic susceptibility to visceral leishmaniasis in The Sudan: Linkage and association with IL4 and IFNGR1. *Genes Immun* 4(5), 351–355.

Mohamed, H.S., Ibrahim, M.E. et al. (2004). SLC11A1 (formerly NRAMP1) and susceptibility to visceral leishmaniasis in The Sudan. *Eur J Hum Genet* 12(1), 66–74.

Morens, D.M., Folkers, G.K. et al. (2004). The challenge of emerging and re-emerging infectious diseases. *Nature* 430(6996), 242–249.

Mosmann, T.R., Cherwinski, H. et al. (2005). Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins. 1986. *J Immunol* 175(1), 5–14.

Nau, G.J., Richmond, J.F. et al. (2002). Human macrophage activation programs induced by bacterial pathogens. *Proc Natl Acad Sci USA* 99(3), 1503–1508.

Ng, A., Bursteinas, B. et al. (2006). pSTIING: A "systems" approach towards integrating signalling pathways, interaction and transcriptional regulatory networks in inflammation and cancer. *Nucleic Acids Res* 34(Database issue), D527–D534.

O'Brien, S.J. and Moore, J.P. (2000). The effect of genetic variation in chemokines and their receptors on HIV transmission and progression to AIDS. *Immunol Rev* 177, 99–111.

Papathanasiou, P. and Goodnow, C.C. (2005). Connecting mammalian genome with phenome by ENU mouse mutagenesis: Gene combinations specifying the immune system. *Annu Rev Genet* 39, 241–262.

Patrinos, G.P., Kollia, P. et al. (2005). Molecular diagnosis of inherited disorders: Lessons from hemoglobinopathies. *Hum Mutat* 26(5), 399–412.

Picard, C., Puel, A. et al. (2003). Pyogenic bacterial infections in humans with IRAK-4 deficiency. *Science* 299(5615), 2076–2079.

Quackenbush, J. (2002). Microarray data normalization and transformation. *Nat Genet* 32(Suppl), 496–501.

Ricciardi-Castagnoli, P. (2005). Genomic signatures of immune and inflammatory responses. *J Pediatr Gastroenterol Nutr* 40(Suppl 1), S45–S46.

Ricciardi-Castagnoli, P. and Granucci, F. (2002). Opinion: Interpretation of the complexity of innate immune responses by functional genomics. *Nat Rev Immunol* 2(11), 881–889.

Rogner, U.C. and Avner, P. (2003). Congenic mice: Cutting tools for complex immune disorders. *Nat Rev Immunol* 3(3), 243–252.

Roy, S., McGuire, W. et al. (1997). Tumor necrosis factor promoter polymorphism and susceptibility to lepromatous leprosy. *J Infect Dis* 176(2), 530–532.

Samson, M., Libert, F. et al. (1996). Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* 382(6593), 722–725.

Schoolnik, G.K. (2002). Functional and comparative genomics of pathogenic bacteria. *Curr Opin Microbiol* 5(1), 20–26.

Schroder, N.W. and Schumann, R.R. (2005). Single nucleotide polymorphisms of Toll-like receptors and susceptibility to infectious disease. *Lancet Infect Dis* 5(3), 156–164.

Sebastiani, G., Olien, L. et al. (1998). Mapping of genetic modulators of natural resistance to infection with *Salmonella typhimurium* in wild-derived mice. *Genomics* 47(2), 180–186.

Selvaraj, P., Swaminathan, S. et al. (2006). Association of human leukocyte antigen-A11 with resistance and B40 and DR2 with susceptibility to HIV-1 infection in south India. *J Acquir Immune Defic Syndr* 43(4), 497–499.

Serre, D. and Hudson, T.J. (2006). Resources for genetic variation studies. *Annu Rev Genomics Hum Genet*. 7, 443–457.

Shaw, M.A., Donaldson, I.J. et al. (2001). Association and linkage of leprosy phenotypes with HLA class II and tumour necrosis factor genes. *Genes Immun* 2(4), 196–204.

Sherlock, G. (2000). Analysis of large-scale gene expression data. *Curr Opin Immunol* 12(2), 201–205.

Smith, K.D. and Bolouri, H. (2005). Dissecting innate immune responses with the tools of systems biology. *Curr Opin Immunol* 17(1), 49–54.

Strunk, T. and Burgner, D. (2006). Genetic susceptibility to neonatal infection. *Curr Opin Infect Dis* 19(3), 259–263.

Ubalee, R., Suzuki, F. et al. (2001). Strong association of a tumor necrosis factor-alpha promoter allele with cerebral malaria in Myanmar. *Tissue Antigens* 58(6), 407–410.

Weiss, K.M. and Terwilliger, J.D. (2000). How many diseases does it take to map a gene with SNPs?. *Nat Genet* 26(2), 151–157.

Wells, C.A., Ravasi, T. et al. (2005). Inflammation suppressor genes: Please switch out all the lights. *J Leukoc Biol* 78(1), 9–13.

Young, K., Frodsham, A. et al. (2005). Inverse associations of human leukocyte antigen and malaria parasite types in two West African populations. *Infect Immun* 73(2), 953–955.

Zhang, R. and Zhang, C.T. (2006). The impact of comparative genomics on infectious disease research. *Microbes Infect* 8(6), 1613–1622.

RECOMMENDED RESOURCES

Journals

Jenner, R.G., Young, R.A. (2005). Insights into host responses against pathogens from transcriptional profiling. *Nat Rev Microbiol* 3(4), 281–294. This review describes a meta-analysis of expression profiling data from the literature of cells infected with different pathogens or treated with pathogenic components.

Lipoldova, M., Demant, P. (2006). Genetic susceptibility to infectious disease: Lessons from mouse models of leishmaniasis. *Nat Rev Genet* 7(4), 294–305. This paper reviews the literature on the genetic susceptibility to infection with *Leishmania* and compares susceptibility loci to those identified for other infections.

Cook, D.N., Pisetsky, D.S., Schwartz, D.A. (2004). Toll-like receptors in the pathogenesis of human disease. *Nat Immunol* 5(10), 975–979. This review describes the role of Toll-like receptors in detection of pathogens and summarizes their involvement in infectious disease susceptibility.

Websites

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene

UniGene is an Organized View of the Transcriptome. Each UniGene entry is a set of transcript sequences that appear to come from the same transcription locus (gene or expressed pseudogene), together with information on protein similarities, gene expression, cDNA clone reagents and genomic location.

http://www.ncbi.nih.gov/entrez/query.fcgi?db=OMIM

Online Mendelian Inheritance in Man is a database catalog of human genes and genetic disorders.

http://www.genome.jp/kegg/ Kyoto Encyclopedia of Genes and Genomes.

http://pstiing.licr.org/

pSTIING (Protein, Signaling, Transcriptional Interactions and Inflammation Networks Gateway) is a publicly accessible knowledge-base about protein–protein, protein–lipid, protein–small molecules, ligand–receptor interactions, receptor–cell type information, transcriptional regulatory and signal transduction modules relevant to inflammation, cell migration and tumorigenesis.

http://www.genmapp.org/

Gene Map Annotator and Pathway Profiler is a computer application designed to visualize gene expression data on maps representing biological pathways and groupings of genes.

http://www.ensembl.org/index.html

Ensembl is a joint project between EMBL-EBI and the Sanger Institute to develop a software system which produces and maintains automatic annotation on selected eukaryotic genomes.

http://www.informatics.jax.org/

MGD includes information on mouse genetic markers, molecular clones (probes, primers and YACs), phenotypes, sequences, comparative mapping data, graphical displays of linkage, cytogenetic and physical maps, experimental mapping data, as well as strain distribution patterns for recombinant inbred strains (RIs) and cross haplotypes.

http://www.geneontology.org/

The Gene Ontology project provides a controlled vocabulary to describe gene and gene product attributes in any organism.

http://www.ncbi.nlm.nih.gov/geo/

Gene Expression Omnibus is a gene expression/molecular abundance repository supporting MIAME compliant microarray data submissions, and a curated, online resource for gene expression data browsing, query and retrieval.

http://www.genome-www5.stanford.edu/

The Stanford MicroArray Database stores raw and normalized data from microarray experiments, and provides data retrieval, analysis and visualization.

http://www.expression.microslu.washington.edu/expression/index.html

Public Microarray Data Download Site powered by Expres-sion Array Manager.

http://www.ncbi.nlm.nih.gov/projects/SNP/

The National Center for Biotechnology Information has established the Single Nucleotide Polymorphism (dbSNP) database to serve as a central repository for both single base nucleotide substitutions and short deletion and insertion polymorphisms.

CHAPTER

Host Genomics and Bacterial Infections

Melissa D. Johnson and Mihai Netea

INTRODUCTION

Bacterial pathogens have been known to cause infectious diseases in humans for centuries. Many of these pathogens are commensal organisms that are ubiquitous in the environment or colonize tissues within the host. Despite near constant contact with bacteria, only few of the exposed individuals actually develop clinical signs and symptoms associated with infection. The interplay between host and pathogen is complex, and infection may depend on genetically determined factors such as host immunity or virulence of the pathogen. Despite the long history of infectious diseases, genetic investigations of bacteria and host immunity have only recently been advanced to improve our understanding of the complex host–pathogen relationship. This area of study has experienced great advances since information from the Human Genome Project became available. Recent technologies such as genetic screening and expression analysis may help to better define the role of key features predisposing to infection, transition of a commensal microorganism into a pathogen, or response to infection once the pathogen has invaded normally sterile body tissue. Such advancements have been used to identify new therapeutic modalities, targeting overexpressed or deficient host immune factors, as well as components of pathogens that are critical for the development of resistance or survival. Collectively "omic" approaches to the study of microbial infections have been termed "infectomics" (Huang et al., 2007). Approaches may be centered on the pathogen, the host, or chemical gene or protein targets in an organism for potential therapeutic benefit. In this chapter, we will focus on host genomics, highlighting examples of studies aimed at unraveling the complex nature of immune response to infection. In the future, better understanding of genomics will hopefully enable us to better prevent and/or treat infections.

GENOMICS AND THE STUDY OF BACTERIAL INFECTIONS

Bacterial Genomics

Functional genomics of bacterial pathogens has been the focus of basic science research for some time. Scientists have long wondered what differentiates a commensal organism from an invading one, how some organisms develop resistance to various antibiotic treatments, and how organisms can so rapidly express the Darwinian extreme of "survival of the fittest". At the heart of this is genetic makeup and expression within bacteria or fungi. On recent review, complete genomes have been mapped for more than several hundred species of bacteria, including *Streptococcus pneumoniae*, *Staphylococcus aureus*, *Pseudomonas aeruginosa*, *Escherichia coli*, and *Neisseria meningitidis* (Celestino et al., 2004; J. Craig Venter Institute, 2006). These data are valuable in developing new targets for antibiotics, new diagnostic methods, and potential vaccination strategies as well as furthering our understanding of bacterial resistance mechanisms, bacterial pathogenicity and virulence, bacterial niche adaptation, and

Copyright © 2010, Elsevier Inc.
All rights reserved.

contribution of bacteria to certain chronic disease states (Fritz and Raczniak, 2002; Monaghan and Barrett, 2006). A discussion of microbial genomics is outside the scope of this text and will thus not be further considered. Nonetheless, many of the methods and procedures discussed in this text are applicable to the field of microbial genomics.

Host Genomics

The study of the association between the role of genes in susceptibility or outcome of infections is still in its infancy. Recent technological advancements have resulted in a surge of studies in this area. These advances include the availability of a large array of candidate genes that can be studied for association, as well as more affordable and fast technology to perform genome-wide screens to identify new genes that may have a role (Hill, 1998).

Briefly, host defense to infectious pathogens depends upon both innate and acquired immunity (Emonts et al., 2003). Phagocytosis is an important first line of defense for some pathogens, particularly those that enter via mucosal surfaces. Cell types such as neutrophils, macrophages, dendritic cells, and monocytes are important components of response to invading pathogens, and CD4+ expressing T lymphocytes differentiate into TH1, TH2, or TH17 cells that modulate the inflammatory process. Monocytes express a number of molecules that are important for antigen presentation and cell signaling, including FC receptors for IgG, complement receptors, and MHC Class II molecules. Complement is a critical part of the innate immune response. Three pathways are activated by either antibody–antigen and C-reactive protein (classical pathway), the interaction of C3 with factors B and D resulting in C3b (alternative pathway) or mannose-binding lectin (MBL) (innate activation pathway). Proteins important for cellular response to infection are depicted in Figure 58.1 (Jenner and Young, 2005).

Pathogen recognition receptors (PRRs) are cell membrane receptors that recognize structures from pathogenic microorganisms called pathogen-associated molecular patterns (PAMPs). Four major classes of PRRs have been described to date: Toll-like receptors (TLRs), NOD-like receptors (NLRs), C-type lectin receptors (CLRs) and RigI-helicases: PRR members of the first three families are involved in the recognition of bacteria, and are involved in the triggering and activation of the innate host defense (Cambi and Figdor, 2005; Kawai and Akira, 2005; Murray, 2005). MBL is part of the collectin group of C-type lectins, which recognizes sugars, differentiating between foreign

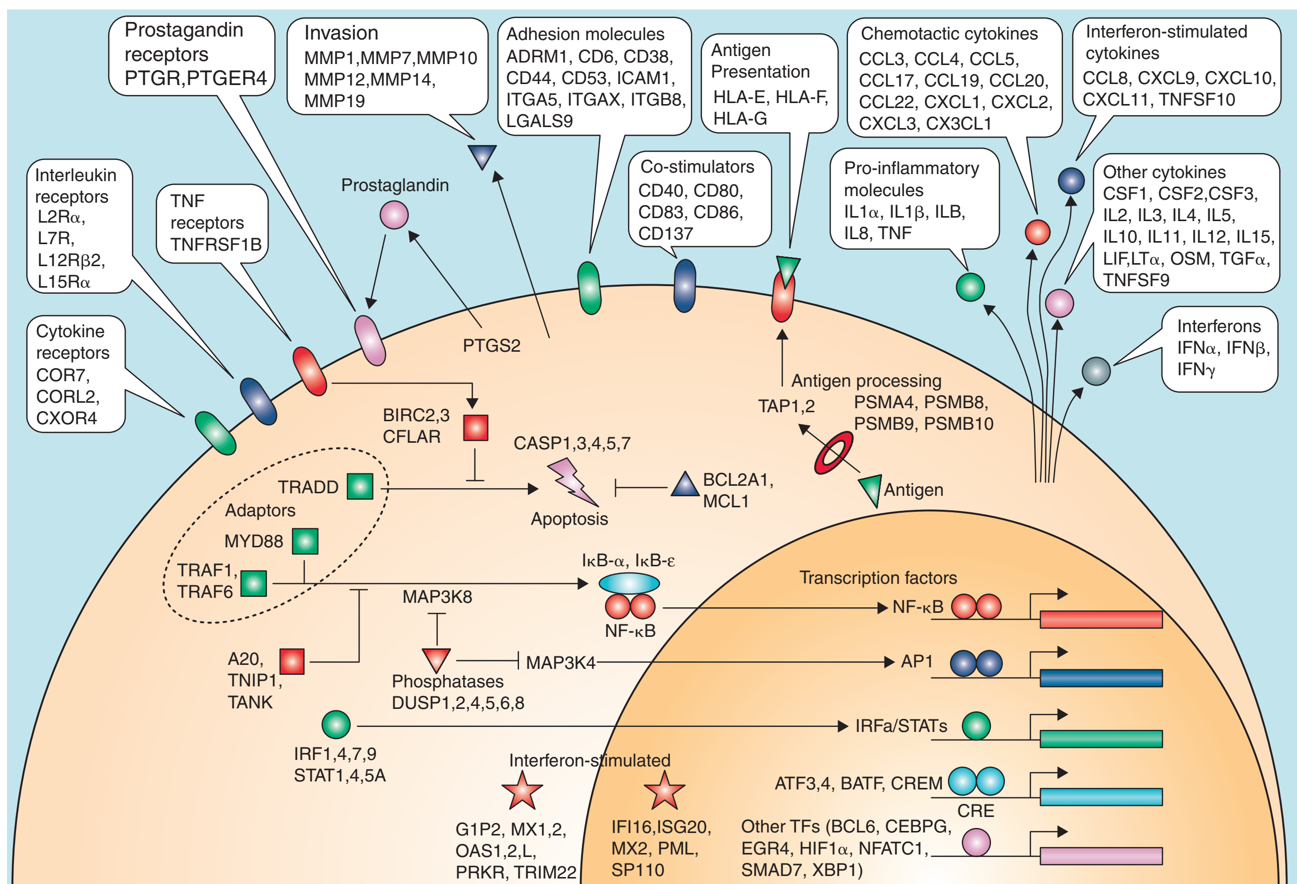


Figure 58.1 Cellular responses to infection (Jenner and Young, 2005). Gene products of the common host response and their functions inside and outside the cell. Arrows represent activatory connections between proteins in signaling pathways; lines ending in bars represent inhibitory connections. This figure does not indicate that these proteins are present in all cells at the same time.

TABLE 58.1 TLR family ligands

TLR	Ligand	Entity commonly bearing ligand
TLR1	Triacyl lipopeptides	Numerous bacteria, mycobacteria
	Soluble factors	*Neisseria* species
TLR2	Lipoprotein/lipopeptides	Numerous pathogens
	Peptidoglycan	Gram-positive bacteria
	Lipoteichoic acid	Gram-positive bacteria
	Lipoarabinomannan	Mycobacteria
	A phenol-soluble modulin	*Staphylococcus* species
	Glycoinositolphospholipids	*Trypanosoma cruzi*
	Glycolipids	*Treponema maltophilum*
	Porins	*Neisseria* spp., *Shigella* spp., *Haemophilus influenzae*
	Zymosan	Fungi
	Atypical LPS	*Leptospira interrogans, Porphyromonas gingivalis*
	HSP70	Host
TLR3	Double-stranded RNA	Viruses
TLR4	LPS	Gram-negative bacteria
	Taxol	Plants
	Fusion protein	Respiratory syncytial virus
	Envelope proteins	MMTV
	HSP60	*Chlamydia pneumoniae*
	F-protein	Cytomegalovirus
	HSP60, HSP70, Type III repeat extra domain A of fibronectin, oligosaccharides of hyaluronic acid, polysaccharide fragments of heparin sulfate, fibrinogen	Host
TLR5	Flagellin	Numerous bacteria (i.e., *Legionella*)
TLR6	Diacyl lipopeptides	*Mycoplasma* species
TLR7	Imidazoquinoline	Synthetic compounds
	Loxoribine	
	Bropirimine	
	Guanine nucleoside analogues	
TLR8	R848/resiquimod	Single-stranded RNA
TLR9	CpG rich motifs of DNA	Bacteria
	Chromatin: IgG complexes	Host
TLR10	Unknown	Unknown
TLR11	Profilin	*T. gondii*
	Not determined	Uropathogenic bacteria
TLR12	Unknown	Unknown
TLR13	Unknown	Unknown

Based on Akira et al., 2006; Brikos and O'Neill, 2008; Dieffenbach and Tramont, 2005.

and self, and subsequently activates complement upon binding to foreign molecules. More than 10 TLRs have been identified, and of these TLR1, TLR2, TLR4, TLR5, TLR6, and TLR9 have proposed roles in recognition of certain bacteria (Table 58.1). Binding of a ligand to a TLR triggers two signaling pathways which differentiate on the basis of MyD88 involvement (Albiger et al., 2007a). MyD88 and another adaptor protein, TIR-associated protein (TIRAP or Mal) lead to induction of a pro-inflammatory cytokine response. With the exception of TLR3, MyD88-dependent pathways are activated by all TLRs. The MyD88-independent pathways are triggered by TLR3 and TLR4; TRIF (TIR domain containing adaptor protein inducing IFN-β) and TRAM (TRIF-related adaptor molecules) activate type I interferons. TRAM is only required in TLR-4 mediated signaling (Turvey and Hawn, 2006). These pathways are depicted in Figure 58.2. NLRs are related molecules that are more specific in their recognition of bacterial organisms in cytoplasm (Elson et al., 2007; Hill, 2006). CLRs are receptors for polysaccharide components such as glucans or mannans, especially in fungi, but they have also been involved in recognition of mycobacteria. CD14 is another pattern recognition receptor and may be important in recognition of both Gram-positive and Gram-negative bacteria. Additional mediators that aid in recruitment, activation or suppression of immune cells include cytokines such as IL-10, IL-6, interferon gamma, TNF, IL-12, and IL-4. Alterations in production or expression of these components

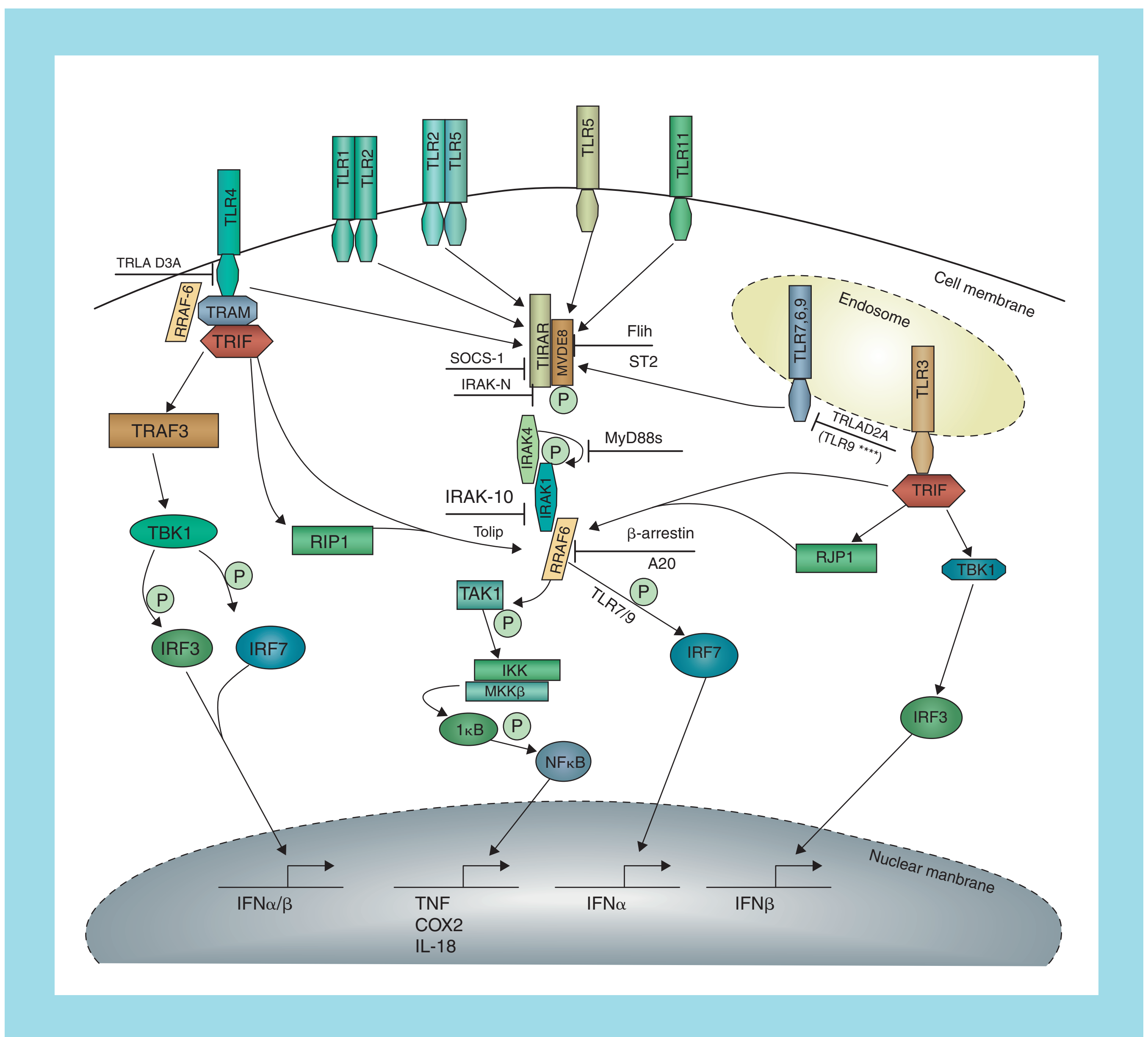


Figure 58.2 TLR signaling (Albiger, 2007a).

can lead to changes in susceptibility to an infecting pathogen or host response once actually infected.

Based on investigations of various pathogens, it seems that there is not one standard panel of factors which predisposes to all infections. Rather, immunity to a given pathogen may depend on the nature of the infecting organism, site of entry, its growth characteristics and/or morphology, and its biochemical composition. Thus, it seems logical that in studies to date, varying host genetic factors have been implicated in either susceptibility to or prognosis with different infectious pathogens. Given that there have been numerous studies in host susceptibility to infections, this chapter will focus on the most common and well-studied Gram-positive and Gram-negative bacterial infections as well as mycobacterial infections. We will compare and contrast the available data for genetic associations and infection with selected pathogens, and highlight potential areas for future exploration.

Methods of Study

Much has been written regarding the different approaches to study host genomics as it relates to infectious diseases, and the limitations of certain study designs are well-recognized. Most infectious diseases can be classified as complex, with multifactorial contributions from environment, biological factors, and inheritance. Susceptibility to an infectious disease may follow a Mendelian pattern of inheritance, but the majority of infectious diseases are due to polygenic factors. It may be difficult to distinguish between these two origins, because other genes can influence expression of a Mendelian trait and expression of a polygenic trait may be heavily influenced by a single gene. While both linkage and association are important, association studies are more commonly employed in studies of genetics and infectious diseases susceptibility (Bochud et al., 2007). While powerful, linkage studies of infectious diseases are relatively uncommon for a number of reasons. First, it is challenging to identify and recruit affected sibling pairs or family members to establish well-characterized pedigrees during the time frame that the study is conducted. The rapid mortality associated with many infectious diseases also makes it difficult to identify affected persons and recruit them for study participation. In addition to these issues, most infectious diseases are expected to have a small contribution from individual genetic regions which would not necessarily be apparent given the limited sample size and power of such studies (Burgner et al., 2006). Thus, case–control association studies including unrelated affected versus non-affected persons are more common than linkage and association studies which might more closely explore genetics among family members. Association studies are not without problems, however. The biggest problem with these study designs is power, and almost all published data for genetic association and infectious diseases are based on underpowered studies. One group has suggested that at least 1500 case–control diads or case–parent triads are necessary for a sufficiently powered association study using a candidate-gene approach. A genome-wide association study could be sufficiently powered with 1000 cases and 1000 controls; however, studies of this kind are subject to a high false discovery rate. New technologies are making such studies more feasible, both from a technical and financial perspective. Given the issues with power and the potential for spurious findings, association studies should be validated in subsequent studies using independent cohorts (see Chapter 4).

In addition to candidate-gene and genome-wide analyses, newer technologies are used to profile host-gene expression in response to different pathogens of interest (Campbell and Ghazal, 2004; Hossain et al., 2006). This approach can create a wealth of information about the functional and dynamic response to infection. For example, in a recent study murine liver cells were studied in response to challenge with *E. coli* or *S. aureus*. Seventeen genes were differentially expressed in response to these two different pathogens, including GBP-2, C-type lectin, adenylate cyclase, and genes involved in vesicle trafficking (Yu et al., 2004). The data suggested that sepsis due to these two pathogens may involve a common late host response, but initial responses to each of these bacteria were quite different. Studies such as this can yield a wealth of new information about pathways that might be important in susceptibility or response to infection. However, since they analyze so many data points, the statistical analysis must be carefully performed to minimize spurious findings. Since expression is a dynamic phenomenon, timing of the sampling and processing procedures used in the laboratory for these specimens are critical elements for this kind of research. Also, many such studies use only one particular cell line and study response of that line to challenge with the pathogen of interest *in vitro*. For many of these infections, multiple cell lines are involved in defense. Thus, incubation with only one kind of cell may not yield a complete picture of what is involved in susceptibility and response to infection *in vivo*. For these reasons, although studies such as these produce a tremendous amount of information, they should be considered hypothesis generating and generally require further validation (Modlin and Bloom, 2001).

HOST GENOMICS AND GRAM-POSITIVE, GRAM-NEGATIVE AND MYCOBACTERIAL INFECTIONS

Gram-Positive Organisms

A few Gram-positive infections have been investigated in human studies for genetic associations and immunity (Table 58.2). Importantly, these organisms lack lipopolysaccharides (LPS), so other mechanisms are responsible for host recognition and defense against these pathogens. In Gram positives, lipoteichoic acid (LTA) is a prominent player in triggering host recognition of the bacteria. LTA and peptidoglycan in Gram positives have been noted to be recognized by TLR2. Gene expression studies have suggested that the initial response to Gram-positive pathogens involves recognition by TLR2, which leads to a NF-κB mediated response with release of cytokines including IL-1 and TNF as well as IL-6 and IL-8. Subsequently, a later TLR-independent response arises in response to the pathogen or its

TABLE 58.2 Examples of host genetic factors implicated in association with select bacterial pathogens in the *in vitro*, animal model, or human studies

Disease	Factors
Gram-positive organisms	
S. pneumoniae	TLR1, TLR2, TLR4, TLR6, TLR9, MyD88, CD14, LBP, NOD1, NOD2, IκB, NFκB-1 A, NFκB-1E, MBL, FC-γRIIA
S. aureus	TLR1, TLR2, TLR6, MyD88, CD14, CD36, NOD2, TNF-α R1, peptidoglycan-recognition proteins
Listeria monocytogenes	TLR2, MyD88, NOD2
Gram-negative organisms	
Gram-negative sepsis/infections	TLR1, TLR2, TLR4, LPB, CD14, MyD88
Legionella	TLR2, TLR4, TLR5, MyD88, Naip5
Neisseria meningitidis	TLR4, MBL, properidin, ACE, FC-γR, TNFα, TNFβ, IL10, IL6, IL1, IL1-R antagonist
Mycobacteria	
M. tuberculosis	TLR1, TLR2, SLC11A1(NRAMP1), Class I and II HLA, MBL2, SP-A, SP-D, P2X7, DC-SIGN, DRB1, HLA-DR, INF-γ, VDR, IL12, IL-12R, IL-10, IL-1R antagonist, TNF-α, TNF-αR1, STAT-1
M. leprae	TLR1, TLR2, TAP-2, HSPA 1 A, DC-SIGN, TNF, VDR, MHC region 6p21, 10p13, HLA-DR, PARK2/PACRG

Based on Hirschhorn et al., 2002; Mira et al., 2004; Mira, 2006; Roy et al., 1997; Roy et al., 1999.

metabolites inside the host cell. These responses appear to be quite pathogen-specific, and differ among various microorganisms. Studies investigating host genomics and *Streptococcus pneumoniae*, *Staphylococcus aureus*, and *Listeria monocytogenes* have been completed and help further define the role of candidate genes for each of these infections.

Of the Gram-positive pathogens, *S. pneumoniae* has been the most studied in regards to host immunity and susceptibility to infection. *S. pneumoniae* is a common cause of pneumonia, bacteremia, meningitis, and otitis media. TLR1, TLR2, TLR4, TLR6, and TLR9 have been associated with recognition and immune response to the pneumococcus (Albiger et al., 2007b; Echchannaoui et al., 2002; Echchannaoui et al., 2005; Knapp et al., 2004; Koedel et al., 2003; Malley et al., 2003; Schroder et al., 2003b; Srivastava et al., 2005; Yoshimura et al., 1999). Much of this evidence comes from animal models of pneumococcal infection. Knockout mice deficient in TLR2 appear to be more susceptible to *S. pneumoniae* meningitis. Reduced survival has also been reported among TLR2 knockout mice with intraperitoneal *S. pneumoniae* infection (Khan et al., 2005). However, in another study, knockout mice deficient in TLR2 did not appear to be at increased risk of pneumococcal pneumonia following intranasal infection. TLR4 has been suggested to have protective effects through its recognition of bacterial pneumolysin, which can potentially reduce colonization and subsequent infection. These benefits may also extend to pneumonia, as evidenced by experimental models in mice with pneumococcal pneumonia. Knockout mice deficient in TLR4 had higher colony counts, although this did not result in increased mortality in this model (Branger et al., 2004). In a model of intravenous infection, TLR4 deficiency did not result in reduced survival or increased colony counts in infected mice (Benton et al., 1997), further suggesting that the role of TLR4 on immunity to *S. pneumoniae* may be limited to the airways. A recent study in human, however, did not find any significant associations between mutations in TLR2 or TLR4 and invasive pneumococcal infection (Moens et al., 2007). In this cohort, neither Caucasian Belgians with invasive pneumococcal disease nor healthy controls were homozygous for mutations in TLR2 R753Q, TLR2 P631H, or TLR2 R579H. Heterozygotes were identified in TLR2 R753Q (6.1% cases, 5% controls, $p=0.94$) and TLR2 P631H (5.1% cases, 8.9% controls, $p=0.34$), but were not significantly different between cases and controls. Similarly, there were no differences in either homozygous or heterozygous mutations in TLR4 D299G between cases and controls. Clinical outcomes were not different among those who were heterozygous or homozygous for TLR2 and TLR4 mutations at these sites, compared to those that had homozygous wild-type genotypes. In this cohort, cases and controls were not matched for gender or age, and the small number of subjects with homozygous mutations in TLR2 and TLR4 limited the power of the study. Additional studies in a larger, more well-defined population with carefully selected controls may be necessary to establish the role of TLR2 and TLR4 mutations and susceptibility to pneumococcal infection.

Other factors such as myeloid differentiation factor 88 (MyD88) and CD14 have also been shown to increase risk of

S. pneumoniae infection in animal models (Albiger et al., 2005). MyD88 is important in the cell signaling cascade triggered by TLR or IL-1 family receptors, while CD14 serves as a co-receptor for LPS. Other proteins may also play a role in recognition and response to pneumococcus, including LPS-binding protein (LBP) and the nucleotide-binding oligomerization domains (Nod) Nod1 and Nod2 (Opitz et al., 2004). The role of these proteins in defense against pneumococcus, particularly in the context of other signaling pathways, has yet to be fully elucidated in animal models and humans.

Recently, mutations in the IκB family of receptors has been associated with invasive pneumococcal disease among hospitalized patients in the United Kingdom (Chapman et al., 2007a). This study included a large cohort of Caucasian UK residents with invasive pneumococcal disease who were identified as part of two other studies (Maskell et al., 2005; Roy et al., 2002) and one control group identified as part of a previous study as well as an independently collected cohort of 370 healthy adult volunteers from the United Kingdom. SNPs were identified in the first cohort of infected patients (including bacteremia, pneumonia, and meningitis) and controls, and then confirmed in the second set of Gram-positive infection (limited to empyema) cases and controls. The investigators identified 43 SNPs in three IκB genes for analysis. Of these, six SNPs in NF-κB-IA or NF-κB-IE appeared to be associated with pneumococcal infection in the first analysis. In the second group of those with Gram-positive empyema (not limited to pneumococcus), the investigators did not find any association between the three most common SNPs and thoracic empyema. When this analysis was limited to the small number of subjects with pneumococcal empyema (42 subjects), mutations in NF-κB-IA (rs3138053 and rs2233406) appeared to be protective, and were similar to that found in the first study cohort with invasive pneumococcal infection. Mutations in NF-κB-IE (rs529948) were significant in the first invasive pneumococcal cohort, but did not appear so in the group of pneumococcal empyema subjects. These data are considered preliminary and should be validated in subsequent studies.

MBL has been suggested to play a role in host defense against *S. pneumoniae.* This protein can phagocytose some bacteria by binding oligosaccharides on the bacterial cell surface (Bocchino et al., 2005). Homozygous mutations in the MBL gene on chromosome 10 have been associated with an increase in invasive pneumococcal infection (Garred et al., 2003; Kronborg et al., 2002; Moens et al., 2006a). In a large study of 225 white adults and children from the United Kingdom with invasive pneumococcal disease (blood, cerebrospinal fluid, joint fluid, etc.) and 353 white adult blood and transplant donors, homozygous mutations for MBL at codons 52, 54, or 57 occurred in 28 infected persons (12%) versus 18 controls (5%) (OR 2.59, 95% CI: 1.39–4.83). This finding was validated in a second cohort of 108 additional cases of invasive pneumococcal infection and 679 healthy neonates from the United Kingdom matched for ethnicity and geographical area of residence. Homozygotes among cases and controls were 10% and 5%, respectively ($p = 0.046$). Frequency of heterozygotes as well as a functional promoter polymorphism at position -221 in both cohorts were similar, and not different between cases and controls. In addition, homozygotes did not have significantly different sites of infection, lengths of stay, disease severity, or survival from those heterozygous for these mutations. The results of this study have been confirmed in two smaller studies. One limitation of this particular study is that cases and controls were not matched for age, despite the fact that the infected cohorts included both children and adults with a wide age range. Age was more closely matched in one of the subsequent studies. In addition, those with concomitant illnesses which would contribute to the risk of invasive pneumococcal disease were not excluded from some of the early studies. Such patients were excluded from the later confirmatory study. Vaccination status of study subjects with pneumococcal vaccine was not described by the investigators of these studies, and could potentially confound the study results. Given the time frame of the study, vaccination among children was probably rare, but from the time period of 1993 to 2003 uptake of pneumococcal vaccination was more than 20% among those 65 years of age and older (Noakes et al., 2006). However, the efficacy of pneumococcal vaccination in those with MBL deficiency is unknown.

In a more recent study using samples from the same cohort of UK residents with invasive pneumococcal infection and non-infected controls, polymorphisms in the FCN2 gene were not found to be associated with this infection (Chapman et al., 2007b). FNC2 encodes for *L*-ficolin, which has been shown to bind Gram-positive organisms, and activates the lectin-complement pathway as well as working directly as an opsonin.

Other studies in humans have investigated mutations in Fcγ-receptor IIA and pneumococcal disease, with conflicting results (Moens et al., 2006b). Fcγ-receptor IIA is thought to play an important role in phagocytosis of bacteria once they have been bound by IgG2. In very small studies, patients who were homozygous for Fcγ-receptor IIA polymorphisms at amino acid 131 appeared to be at increased risk of invasive pneumococcal infection (i.e., bacteremia) than controls (Yee et al., 1997; Yee et al., 2000; Yuan et al., 2003). In a more recent study of 55 Belgians with invasive pneumococcal disease and 100 gender and geographical matched non-infected controls of a wide age distribution, there was no difference in prevalence of Fcγ-receptor IIA genotypes between cases and controls (Fcγ-RIIA–R/R131 genotype 21.8% versus 31% in cases and controls, respectively ($p = 0.047$)). In general, all of these studies were limited by very small sample sizes.

Although *S. aureus* also has LTA in its cell wall, lipoproteins and peptidoglycans may be even more important in its recognition by host TLR2 (Hashimoto et al., 2006). For *S. aureus*, knockout mice deficient in TLR2 experienced greater rates of mortality compared to wild-type mice with *S. aureus* infection. Similarly, mice deficient in MyD88, a downstream adaptor molecule involved in cell signaling, had substantially reduced survival which was even more pronounced than the TLR2 deficient mice. Colony counts in the organs of the Myd88 deficient mice indicated that these mutants were even more susceptible to *S. aureus* infection than the TLR2 deficient mice (Takeuchi et al., 2000). A study in 91 white French patients who had participated

in a study of the role of TNF-alpha in septic shock found that 2 of 22 patients (9%) with Gram-positive septic shock had a mutation in TLR2 (Arg753Gln) while this mutation was not identified among 69 subjects with other causes of septic shock. Both subjects were heterozygous for Arg753Gln mutation, and had shock due to *S. aureus* infection. Allelic frequency of this mutation among non-infected controls was 3% (Lorenz et al., 2000). This finding has yet to be confirmed from other studies to date (Moore et al., 2004). Other important elements of host recognition and response to *S. aureus* may include TLR1, TLR6, CD14, CD36, NOD2, TNF-alpha receptor 1 (TNFR1), and peptidoglycan recognition proteins (Fournier and Philpott, 2005). The specific role and association of these molecules with *S. aureus* infection has yet to be established or studied in humans.

Listeria monocytogenes may also be recognized by TLR2, and in animal models knockout mice deficient in TLR2 and MyD88 were more susceptible to this infection (Seki et al., 2002; Torres et al., 2004). In addition, NOD2 may play a role in defense against *Listeria*, as evidenced by knockout mice who were more susceptible to oral *Listeria* infection (Kobayashi et al., 2005). A number of other factors may be important to defense against *L. monocytogenes* and have been explored in animal models, but all of these findings have yet been validated in human studies of *Listeria* infection (Pamer, 2004).

Gram-Negative Organisms

Enterobacteriaceae and Other Gram-Negative Rods

LPS in the cell wall of Gram-negative bacteria are responsible, once recognized by host-cell receptors, for triggering a cascade of events leading to adhesion of neutrophils to endothelial cells, clotting, and activation of secondary inflammatory mediators such as interleukins and leukotrienes. LPS is recognized by TLR4, LBP, and CD14. Some Gram-negative organisms also contain membrane proteins in the cytoplasm or outer cell membrane that are recognized by host-cell receptors such as TLR2 and TLR1. Several studies in mice have demonstrated that MyD88 signaling may be critical for the early host response to *Ps. aeruginosa* lung infections (Power et al., 2004; Skerrett et al., 2004).

TLR4 may be important for a number of Gram-negative pathogens; however, results in human studies have yielded conflicting results. In one *in vitro* study using human cell lines, TLR4 but not TLR2 was the major factor mediating host-cell activation in response to *Salmonella minnesota*- and *E. coli*-derived LPS (Tapping et al., 1993). This was recently validated in a study where TLR4 ligands seemed to predominate for Enterobacteriaceae. In contrast, responses to *Ps. aeruginosa* were dependent on both TLR4 and TLR2.

Several small studies in humans have explored the role of TLR4 in Gram-negative infections. Among Europeans, TLR4 Asp299Gly and Thr399Ile SNPs have been found to cosegregate in 98% of individuals, while in African populations both an Asp299Gly and an Asp299Gly/Thr399Ile haplotype is present (Schroder and Schumann, 2005). In a case–control study of 91 Caucasian patients with septic shock and 73 healthy non-infected blood donors in France, a TLR4 Asp299Gly mutation occurred exclusively among those with septic shock (5/91 cases (5.5%) versus 0/73 controls (0%)) (Lorenz et al., 2002). There was a similar rate of dual mutations at the 299 and 399 position (i.e., TLR4 Asp299Gly/Thr399Ile) in both cases and controls, 6/91 (6.6%) and 8/73 (11%) respectively. Among those with septic shock, solo mutations in 299 were present exclusively in those with Gram-negative infections ($n = 4$) or polymicrobial infections ($n = 1$). Survival and SAPS II scores of these subjects were similar to those with wild-type or dual mutation genotypes at the 299 and 399 position. Only one subject in this study was homozygous for the 299 mutation, and she died of rapidly progressive urosepsis and *E. coli* bacteremia. Based on these findings, the authors concluded that risk of Gram-negative septic shock may be higher among those with TLR4 299 mutations. Additional studies in a larger study population with more carefully selected controls should be performed to confirm these findings.

In another study of 80 Spanish hospital inpatients with either acute ($n = 24$) or chronic ($n = 56$) osteomyelitis due to Gram-negative ($n = 23$) or Gram-positive ($n = 57$) organisms and 155 healthy blood donors matched by age and gender, polymorphisms in TLR4 Asp299Gly appeared to be associated with Gram-negative infections (Montes et al., 2006). In this cohort, three cases were homozygotes for polymorphisms at this site (Asp299Gly G/G) and two of these had Gram-negative infections. No controls were found to have homozygous polymorphisms in TLR4 Asp299Gly ($p = 0.038$). Allelic frequencies did not suggest that patients with osteomyelitis had higher frequency of mutant carrier alleles for this site than controls ($p = 0.08$). Similar results were found for TLR4 Thr399Ile. Polymorphisms in TLR2 (Arg753Gln) were not significantly different between cases and controls. Those with TLR4 Asp299Gly heterozygous or homozygous polymorphisms had higher rates of Gram-negative infections (60% versus 21.5% respectively, $p = 0.0086$), higher incidence of hematogenous osteomyelitis (40% versus 11% respectively, $p = 0.013$) and chronic osteomyelitis more frequently (93.3% versus 64.6% respectively, $p = 0.031$) than those who did not have these mutant alleles. Functional significance of these polymorphisms in TLR4 Asp299Gly were explored, and the investigators observed reduced apoptosis of neutrophils after incubation with LPS among those with osteomyelitis compared to non-infected controls, and among those with mutant alleles. In addition, phosphorylation of IκB after incubation with LPS was reduced among those homozygous for the Asp299Gly mutant alleles. Furthermore, neutrophils from those with these mutant alleles did not demonstrate increases in IL-6 secretion, as was observed in non-infected controls and those homozygous for the wild-type alleles. As with previous studies, there was a very low number of subjects who were homozygous for the TLR4 Asp299Gly mutation in this study, which limits its ability to make powerful statistical comparisons. Also, the authors did not apparently control statistically for the large number of comparisons used in this study. This may bias the study findings, especially considering its small sample size and rather heterogenous population.

Thus, the Asp299Gly/Thr399Ile haplotype of TLR4 does not conclusively appear to increase risk of Gram-negative

infections. While several studies have implicated TLR4 haplotypes containing only the Asp299Gly (or Thr399Ile) SNPs, the small number of patients with these haplotypes in each of these studies limits our ability to make conclusions about these haplotypes and their role in Gram-negative infections.

Legionella

A few studies have investigated immunogenetic factors specific to *Legionella* spp. infections. *Legionella* is a flagellated Gram-negative rod, and *in vitro* TLR2, TLR4, and TLR5 have been shown to recognize this pathogen (Akamine et al., 2005; Girard et al., 2003; Hawn et al., 2003; Hawn et al., 2005; Kikuchi et al., 2004). In murine models TLR2 and MyD88 have been shown to be important (Hawn et al., 2006b). However, studies in animals have produced conflicting evidence regarding the role of TLR4 and *Legionella*. TLR4 polymorphisms have been suggested to play a protective role in a human cohort exposed to *Legionella* at a flower show in the Netherlands. In a study of 108 subjects who contracted Legionnaire's disease following this exposure and 508 controls that were exposed but not clinically ill, Legionella pneumonia was significantly less common among those with the Asp299Gly SNP. In this study, only heterozygotes for Asp299Gly were found, and Thr399Ile cosegregated with Asp299Gly in this European population. As the authors note, it is not clear why TLR4 polymorphisms are associated with a protective effect against *Legionella*, especially since these polymorphisms have been associated with no effect or increased risk of infection due to other Gram-negative pathogens. This phenomenon may be due to the unique LPS structure of *Legionella* that results in differential cell recognition and signaling. However, additional studies are needed to validate these findings.

Recently Naip5, which is important for intracellular signaling and the apoptosis mechanism of macrophages, has been suggested to play a role in the intracellular pathogenicity of *Legionella* (Wright et al., 2003). Naip5 can activate macrophages via TLR5 and MyD88-independent mechanisms (Delbridge and O'Riordan, 2007). Mice with polymorphisms in Naip5 have been shown to be at increased susceptibility to *Legionella* infection (Diez et al., 2003). No human studies to date have investigated Naip5 and Legionnaire's Disease.

Neisseria meningitidis

Neisseria meningitidis is an encapsulated Gram-negative diplococcus that is a common cause of meningitis and sepsis. A study of siblings in the United Kingdom suggested that genetic factors in the host accounted for approximately 33% of risk of meningococcal disease among siblings of those with meningococcal infection (Haralambous et al., 2003). Multiple mechanisms appear to be involved with recognition and host immunity to meningococcus. Innate and adaptive immunity mechanisms have been demonstrated in response to meningococcal challenge, as well as complement activation and subsequent shock/coagulopathy. Studies of genetics and meningococcal disease have suggested that polymorphisms in key molecules can result in increased susceptibility to this infection or more severe phenotypes.

Inherited complement deficiencies have been reported in association with meningococcal disease, but these are relatively rare in the general population and probably not the primary reason for susceptibility to this infection.

Among candidate genes, polymorphisms in TLR4 have not been strongly associated with either increased susceptibility to or severity of infection (Read et al., 2001; Smirnova et al., 2003). A study of UK residents with meningococcal infection found no association between the TLR4 polymorphism Asp299Gly and meningococcal infection. Although most of the subjects with meningococcal infection ($n = 1047$) were <1 year of age, the investigators used a control population consisting of 879 healthy adult blood donors. However, the investigators also stratified the infected cohort by age group and serotype and found no associations between TLR4 Asp299Gly polymorphism among these strata and fatality. A recent study suggested that polymorphisms in TLR4 Asp299Gly and The399Ile were associated with increased susceptibility to meningococcal disease among European Caucasian children less than 12 months of age. There was no significant difference in the rate of meningococcal disease among those 3 years of age and older who were heterozygotes or homozygotes for the Asp299Gly and The399Ile SNPs compared to healthy controls, however (Faber et al., 2006). The authors suggest these mutations may be particularly relevant among infants who have an immature immune system.

Polymorphisms in TLR2 have not appeared to increase susceptibility to meningococcal disease, but have not been extensively studied.

Studies of MBL polymorphisms have yielded more substantial associations for meningococcal infection (Hibberd et al., 1999). Heterozygotes and homozygotes for variant alleles of MBL have been shown to have substantially lower circulating MBL concentrations. Several reports have suggested that those with variant alleles are at increased risk for meningococcal infection (Bax et al., 1999; Salimans et al., 2004). A recent study also suggested that combined polymorphisms in properidin and MBL results in an increase risk of *N. meningitidis* meningitis (Bathum et al., 2006). Three properidin deficiency variant types have been identified and have been associated with increase severity of meningococcal infection and clinical outcome. However, additional genetic factors may influence mortality and severity of disease among individuals with properidin deficiencies (Densen et al., 1987; Fijen et al., 1999; Spath et al., 1999; Westberg et al., 1995).

Angiotensin-converting enzyme (ACE) has pro-inflammatory properties and in one study, the DD genotype that imparts higher circulating levels of ACE was associated with increased severity of meningococcal disease (Harding et al., 2002).

Other elements of acquired immunity may be important for susceptibility to and severity of meningococcal disease. For example, mutations in FCγ receptors have been associated with meningococcal sepsis and were more frequent in those with meningococcal disease than controls in some studies (Bredius et al., 1994; Platonov et al., 1998; van der Pol et al., 2001). However, these results were not consistent across investigations (Domingo et al., 2002; Fijen et al., 2000).

Polymorphisms in cytokines such as TNFα, TNFβ, IL10, IL6, IL1, and IL1 receptor antagonist may be important but have produced inconsistent results in human studies of meningococcal disease.

Mycobacteria

There are several indicators that TB infection might be associated with genetic susceptibility (Bellamy, 2005). These include a potential role of innate immunity of the disease, whereby only 30–40% of close contacts of an infected individual contract TB infection (National Center for HIV SATPC, 2005). In addition, numerous animal models and human investigations have demonstrated the potential role of interferon-γ, TNF-α, reactive nitrogen intermediates, and other immunomodulators as well as CD4+ and CD8+ T cells in controlling the disease (Stead, 1992).

Numerous reports provide supportive data for genetic susceptibility to tuberculosis. Some of the most convincing data come from a tragic accident involving BCG (Bacilli Calmette-Guerin) vaccine in Germany to 249 infants in 1926 (Dubos and Dubos, 1952). Unfortunately, one strain of live *M. tuberculosis* was administered to all of these children, resulting in subsequent death of 76 babies within 1 year. Since only one strain was involved, this incident suggested that there was a differential response to tuberculosis infection in this population. Other factors that could have impacted response to infection were rather limited in this population due to age and disease exposure. Other evidence comes from selective pressure within a relatively homogenous population of Qu'Appelle Indians who experienced an annual death rate of 10% of their population from tuberculosis infection in 1890. This rate declined to less than 0.2% after two generations, suggesting a selective advantage among families with the ability to fight *M. tuberculosis* (Motulsky, 1960). Further evidence for the role of genetics in TB comes from a study of more than 25,000 residents of racially integrated nursing home in Arkansas (Stead et al., 1990). Development of a new TB infection (as evidenced by skin test conversion with 60 days of a negative test) while in the nursing home was approximately twice as likely among blacks and whites, with 13.8 versus 7.2% experiencing new infections, respectively (relative risk 1.9, 95% CI: 1.7–2.1). Even when the source patient was white, blacks became infected more frequently. However, once infected, individuals from both races appeared to be at similar risk of progressing to clinical infection. Finally, other studies in monozygotic twins compared to dizygotic twins in older studies found TB concordance rates twice as high among monozygotic twins, further suggesting a genetic role in susceptibility to disease.

Subsequent investigations in the genomic era have furthered our understanding of genetic factors and susceptibility to tuberculosis. *In vitro* studies and animal models of infection have been established to help describe the role of innate immunity against mycobacteria. Studies in humans have included analysis of individual patients with disastrous consequences following vaccination with other mycobacterial species, candidate-gene approaches using case–control designs, and genome-wide scans using families. The populations included in these various studies of genetic association and linkages have mostly included African (Gambian or South African) and Asian (Japanese, Chinese, Koreans, Taiwanese, and Hong Kong) populations. Only a few have included Europeans, Indians, or Mexicans (Fernando and Britton, 2006).

Based on these studies, a number of genes have been identified that may have a role in increased or decreased risk of mycobacterial diseases (Table 58.2). These include: interferon-γ, IL-10 and IL-12 receptor, TNF-α receptor 1, STAT-1, vitamin D receptor (VDR), solute carrier 11a1 protein (SLC11A1) (formerly known as natural resistance-associated macrophage protein (NRAMP1)), Class I and Class II HLA, mannose-binding lectin (MBL2), pulmonary surfactant proteins SP-A and SP-D, the purinergic receptor P2X7, TLR2, IL-1 receptor antagonist, and TNF-α (Soborg et al., 2007; Stein et al., 2007). To date, associations with heat shock protein 1A (HSPA1A) and transporter associated with antigen processing-2 (TAP-2) have only been reported in association with *M. leprae* infections. In addition to loci, microsatellite markers have been identified on the X chromosome, as well as chromosome 15 which suggest association with TB in Gambians (Bellamy et al., 2000). Genome-wide scan approaches in *M. leprae* infections identified other microsatellite markers on chromosome 6, 10, and 20 that may be important.

A study in Turkish patients with tuberculosis and non-infected controls suggested that the Arg753Gln polymorphism of TLR2 was associated with increased susceptibility to TB. Homozygotes for the AA genotype were more frequent among those with TB (OR 6.04 (95% CI: 2.01–20.08)), and this remained significant when the analysis was restricted to those with pulmonary TB. Heterozygotes were also more frequent among those with TB, although this finding was less substantial (OR 1.6 (95% CI: 1.01–2.55)) (Ogus et al., 2004). A recent study also suggested that TLR2 genotype 597CC were associated with susceptibility to tuberculous meningitis in Vietnamese subjects (Thuong et al., 2007). This association was not found for pulmonary tuberculosis in this case–control study. Another analysis of this cohort suggested that a SNP in the TLR1 receptor domain (C558T) was associated with tuberculosis, and more strongly with tuberculous meningitis (Hawn et al., 2006a).

In those with lepromatous leprosy, the Arg677Trp mutation of TLR2 was found in 10 of 45 subjects, but was absent in those with tuberculoid leprosy and healthy controls (Kang and Chae, 2001). This mutation was later shown to eliminate activation of NF-κB mediated by TLR2, in response to challenge with *M. leprae* (Bochud et al., 2003). Subsequent studies have not replicated the findings of the Arg677Trp mutation in other cohorts, however, and additional studies are necessary to establish the role of this TLR2 mutation (Alcais et al., 2005; Sanchez et al., 2004; Schroder et al., 2003a).

Another study has suggested that the isoleucine to serine mutation at position 602 on TLR1 was protective against leprosy among Turkish Caucasians (Johnson et al., 2007). TLR1 has been suggested to be an important co-receptor for TLR2. The 602S mutation was quite common among Caucasians and substantially more frequent than in those of African or East Asian descent. The authors suggest that these differences may

contribute to the increased and unexplained risk of tuberculosis among African Americans in previous studies.

Several recent studies have also investigated variation in the C-type lectin DC-SIGN and mycobacterial diseases such as tuberculosis and leprosy. Numerous studies indicate that variation in repeats in the neck region of DC-SIGN had no association with susceptibility to tuberculosis infection among South African Coloureds, Colombians, and Tunisians (Barreiro et al., 2007; Ben-Ali et al., 2007; Gomez et al., 2006). Variations in the neck region of *L*-SIGN (or DC-SIGNR) were also not associated with susceptibility to tuberculosis among South Africans. Promoter polymorphisms in DC-SIGN have been implicated in tuberculosis susceptibility among South Africans in one study (Barreiro et al., 2006a) but not other studies of Tunisians and West Africans (Olesen et al., 2007). The correlation among South Africans may reflect Eurasian descent of the population, in contrast to the other populations studied.

DC-SIGN variations were also recently studied in association with leprosy in a cohort of 194 infected Pakistanis and 78 non-infected controls (Barreiro et al., 2006b). 109 patients had lepromatous disease, while 85 had tuberculoid leprosy. In a cold binding assay, DC-SIGN expressing cells showed substantial binding to *M. leprae,* and the proportion was similar to binding of *M. tuberculosis*. However, there was no association between SNPs in DC-SIGN and susceptibility to leprosy. Since the number of cases/controls was relatively small, the study was potentially limited in its power to detect these differences. Future studies may be helpful to determine if DC-SIGN has a role in susceptibility to leprosy.

2009 UPDATE

The promise of genomic approaches for the identification and diagnosis of bacterial infections continues to be explored with fervor. Several recent reviews discuss the many approaches that can be taken to the design and conduction of genomic studies (Farber et al., 2008, Polpitiya et al., 2009; Yang et al., 2008). Additional methods such as transcriptomic (RNA) and proteomic (protein) analyses may provide further insight as novel diagnostics are applied to infectious diseases (Polpitiya et al., 2009). Clinical studies demonstrating these concepts are still somewhat limited, which reflects the substantial challenges that still face investigators in this area. These challenges include (1) establishing communication and uniform methods among researchers from various backgrounds and with disparate skill sets; (2) having both the experience and the necessary technical proficiency to produce reliable data; and (3) developing and implementing novel applications to facilitate data management and analysis (Polpitiya et al., 2009).

One recent microarray study performed using neutrophilic RNA from septic Australian ICU patients identified a unique expression profile for sepsis (Tang et al. 2008). This method did not distinguish between gram-positive and gram-negative bacterial sepsis using a rather large gene expression analysis of 18,664 genes involved in inflammation, immune regulation and mitochondrial function, but did distinguish infected (n = 55) from non-infected (n = 17) patients. The authors suggest that the lack of discrimination between gram-positive and gram-negative sepsis in this investigation and others (Nau et al., 2002; Boldrick et al. 2002) may be due to a common general response of gene expression when the host is challenged with these pathogens.

Genetic associations and immunity to bacterial infections continue to be explored as well. A recent case series of nine children with invasive pyrogenic bacterial infections provided human evidence for the role of Myd88 deficiency in susceptibility to infection (von Bernuth et al. 2008). Myd88 deficiency in these children appeared to be the result of an autosomal recessive trait, which resulted in loss of Myd88 function. However, unlike phenotypes observed with this deficiency in mouse models of infection, these children were not universally vulnerable to all pathogens and retained immunity to most bacteria, viruses, and parasites. These patients appeared to be particularly susceptible to bacteria such as *S. pneumoniae*, *S. aureus*, and *Ps. aeruginosa*. These data suggest that the role of Myd88-dependent signaling in TLRs may be especially important among infants, before other adaptive or TLR-independent innate immune mechanisms fully develop.

Another recent paper suggested an independent association between donor polymorphisms in mannose-binding lectin 2 (MBL2) and the rate of clinically significant infections among orthotopic liver transplant recipients following transplantation (Worthley et al. 2009). In this study, donor MBL2 polymorphisms were associated with low MBL serum concentrations in transplant recipients, who experienced a variety of post-transplant infections such as pneumonia, bacteremia and peritonitis due to numerous pathogens, including cytomegalovirus, gram-positive bacteria such as *S. aureus* and *Enterococcus spp.* and gram-negative bacteria such as *Ps. aeruginosa*, *K. pneumoniae* and *Enterobacter cloacae*. When controlling for other clinical factors such as CMV infection, donor/recipient CMV serostatus, organ rejection, age, and MELD score, donor MBL2 genotype (hazard ratio = 2.8, p = 0.002) and the receipt of CMV prophylaxis (hazard ratio = 2.6, p = 0.005) were independent predictors of clinically significant infection in the post-transplant period. These results confirm the findings of a smaller study that was previously performed in liver transplant recipients (Bouwman et al. 2005), but remain to be validated in other populations.

An association between NOD2 and TLR4 polymorphisms and acquisition of bacteremia was recently demonstrated in a study of 774 medical intensive care unit patients of Western European descent (Henckaerts et al., 2009). Data

regarding the incidence of gram-positive versus gram-negative bacteremia were not presented, but the authors state that this association persisted whether patients had gram-positive bacteremia, gram-negative bacteremia, or both. The association remained significant even after adjusting for other known risk factors for acquisition of bacteremia in the ICU. There was a higher rate of death among patients with both at least one NOD2 variant and a TL4 mutation, but not among those with variants in either gene alone. Those with both NOD2 and TL4 variants had higher APACHE II scores on admission to the ICU, which could have impacted mortality rates among those patients. Variants of other candidate genes such as TLR2, IL4, IL5, IL6, IL10, TNF-alpha, NOD1, and MBL2 were not found to be associated with acquisition of bacteremia in this cohort. However, those with variants of MASP2, a mediator in the MBL pathway of the complement system, also appeared to have a higher rate of death.

Variants in TLR1 were implicated in gram-positive bacteremia in a study of sepsis in the ICU from one cohort of 1183 Canadian medical/surgical ICU patients and a separate cohort of 237 patients with severe sepsis in the United States (Wurfel et al. 2008). In both cohorts, homozygotes for the TLR1 variant -7202 GG were found to have a higher rate of gram positive organisms, but their overall rate of infection from all causes was similar to others with septic shock. These subjects were found to have higher mortality and higher rates of organ failure.

Together these data provide additional insights in identifying markers of infection and patients at risk for infection. However, more studies are needed to further develop and apply these concepts in the clinical arena.

2009 UPDATE REFERENCES

Boldrick, J.C., Alizadeh, A.A., Diehn, M. et al. (2002). Stereotyped and specific gene expression programs in human innate immune responses to bacteria. *Proc Natl Acad Sci U S A* 99, 972–977.

Bouwman, L.H., Roos, A., Terpstra, O.T. et al. (2005). Mannose binding lectin gene polymorphisms confer a major risk for severe infections after liver transplantation. *Gastroenterology* 129, 408–414.

Farber, C.R. and Lusis, A.J. (2008). Integrating global gene expression analysis and genetics. *Adv Genet* 60, 571–601.

Henchaerts, L., Nielsen, K.R., Steffensen, R. et al. (2009). Polymorphisms in innate immunity genes predispose to bacteremia and death in the medical intensive care unit. *Crit Care Med* 37(1), 192–201.

Nau, G.J., Richmond, J.F., Schlesinger, A., Jennings, E.G., Lander, E.S. and Young, R.A. (2002). Human macrophage activation programs induced by bacterial pathogens *Proc Natl Acad Sci USA* 99, 1503–1508.

Polpitiya, A.D., McDunn, J.E., Burykin, A., Ghosh, B.K. and Cobb, J.P. (2009). Using systems biology to simplify complex disease: Immune cartography. *Crit Care Med* 37(1 Suppl), S16–S21.

Tang, B.M., McLean, A.S., Dawes, I.W., Huang, S.J., Cowley, M.J. and Lin, R.C. (2008). Gene-expression profiling of gram-positive and gram-negative sepsis in critically ill patients. *Crit Care Med* 36(4), 1125–1128.

von Bernuth, H., Picard, C., Jin, Z. et al. (2008). Pyogenic bacterial infections in humans with MyD88 deficiency. *Science* 321(5889), 691–696.

Worthley, D.L., Johnson, D.F., Eisen, D.P. et al. (2009). Donor mannose-binding lectin deficiency increases the likelihood of clinically significant infection after liver transplantation. *Clin Infect Dis* 48, 410–417.

Wurfel, M.M., Gordon, A.C., Holden, T.D. et al. (2008). Toll-like receptor 1 polymorphisms affect innate immune responses and outcomes in sepsis. *Am J Respir Crit Care Med* 178, 710–720.

Yang, X., Yang, H., Zhou, G. and Zhao, G.P. (2008). Infectious disease in the genomic era. *Annu Rev Genomics Hum Genet* 9, 21–48.

REFERENCES

Akamine, M., Higa, F., Arakaki, N., Kawakami, K., Takeda, K., Akira, S. and Saito, A. (2005). Differential roles of Toll-like receptors 2 and 4 in *in vitro* responses of macrophages to *Legionella pneumophila*. *Infect Immun* 73, 352–361.

Akira, S., Uematsu, S. and Takeuchi, O. (2006). Pathogen recognition and innate immunity. *Cell* 124, 783–801.

Albiger, B., Sandgren, A., Katsuragi, H., Meyer-Hoffert, U., Beiter, K., Wartha, F., Hornef, M., Normark, S. and Normark, B.H. (2005). Myeloid differentiation factor 88-dependent signalling controls bacterial growth during colonization and systemic pneumococcal disease in mice. *Cell Microbiol* 7, 1603–1615.

Albiger, B., Dahlberg, S., Henriques-Normark, B. and Normark, S. (2007a). Role of the innate immune system in host defence against bacterial infections: Focus on the Toll-like receptors. *J Intern Med* 261, 511–528.

Albiger, B., Dahlberg, S., Sandgren, A., Wartha, F., Beiter, K., Katsuragi, H., Akira, S., Normark, S. and Henriques-Normark, B. (2007b). Toll-like receptor 9 acts at an early stage in host defence against pneumococcal infection. *Cell Microbiol* 9, 633–644.

Alcais, A., Mira, M., Casanova, J.L., Schurr, E. and Abel, L. (2005). Genetic dissection of immunity in leprosy. *Curr Opin Immunol* 17, 44–48.

Barreiro, L.B., Neyrolles, O., Babb, C.L., Tailleux, L., Quach, H., McElreavey, K., Helden, P.D., Hoal, E.G., Gicquel, B. and Quintana-Murci, L. (2006a). Promoter variation in the DC-SIGN-encoding gene CD209 is associated with tuberculosis. *PLoS Med* 3, e20.

Barreiro, L.B., Quach, H., Krahenbuhl, J., Khaliq, S., Mohyuddin, A., Mehdi, S.Q., Gicquel, B., Neyrolles, O. and Quintana-Murci, L. (2006b). DC-SIGN interacts with *Mycobacterium leprae* but sequence variation in this lectin is not associated with leprosy in the Pakistani population. *Hum Immunol* 67, 102–107.

Barreiro, L.B., Neyrolles, O., Babb, C.L., van Helden, P.D., Gicquel, B., Hoal, E.G. and Quintana-Murci, L. (2007). Length variation of

DC-SIGN and L-SIGN neck-region has no impact on tuberculosis susceptibility. *Hum Immunol* 68, 106–112.

Bathum, L., Hansen, H., Teisner, B., Koch, C., Garred, P., Rasmussen, K. and Wang, P. (2006). Association between combined properdin and mannose-binding lectin deficiency and infection with *Neisseria meningitidis*. *Mol Immunol* 43, 473–479.

Bax, W.A., Cluysenaer, O.J., Bartelink, A.K., Aerts, P.C., Ezekowitz, R.A. and van, D.H. (1999). Association of familial deficiency of mannose-binding lectin and meningococcal disease. *Lancet* 354, 1094–1095.

Bellamy, R. (2005). Genetic susceptibility to tuberculosis. *Clin Chest Med* 26, 233–246.

Bellamy, R., Beyers, N., McAdam, K.P., Ruwende, C., Gie, R., Samaai, P., Bester, D., Meyer, M., Corrah, T. and Collin, M.etal. (2000). Genetic susceptibility to tuberculosis in Africans: A genome-wide scan. *Proc Natl Acad Sci USA* 97, 8005–8009.

Ben-Ali, M., Barreiro, L.B., Chabbou, A., Haltiti, R., Braham, E., Neyrolles, O., Dellagi, K., Gicquel, B., Quintana-Murci, L. and Barbouche, M.R. (2007). Promoter and neck region length variation of DC-SIGN is not associated with susceptibility to tuberculosis in Tunisian patients. *Hum Immunol* 68, 908–912.

Benton, K.A., Paton, J.C. and Briles, D.E. (1997). The hemolytic and complement-activating properties of pneumolysin do not contribute individually to virulence in a pneumococcal bacteremia model. *Microb Pathog* 23, 201–209.

Bocchino, M., Marruchella, A. and Saltini, C. (2005). Immunogenetics of severe respiratory infections: Models for the development of new therapeutic strategies. *Respiration* 72, 449–457.

Bochud, P.Y., Hawn, T.R. and Aderem, A. (2003). Cutting edge: A Toll-like receptor 2 polymorphism that is associated with lepromatous leprosy is unable to mediate mycobacterial signaling. *J Immunol* 170, 3451–3454.

Bochud, P.Y., Bochud, M., Telenti, A. and Calandra, T. (2007). Innate immunogenetics: A tool for exploring new frontiers of host defence. *Lancet Infect Dis* 7, 531–542.

Branger, J., Knapp, S., Weijer, S., Leemans, J.C., Pater, J.M., Speelman, P., Florquin, S. and van der, P.T. (2004). Role of Toll-like receptor 4 in Gram-positive and Gram-negative pneumonia in mice. *Infect Immun* 72, 788–794.

Bredius, R.G., Derkx, B.H., Fijen, C.A., de Wit, T.P., de, H.M., Weening, R.S., van de Winkel, J.G. and Out, T.A. (1994). Fc gamma receptor IIa (CD32) polymorphism in fulminant meningococcal septic shock in children. *J Infect Dis* 170, 848–853.

Brikos, C. and O'Neill, L.A. (2008). Signalling of toll-like receptors. In *Toll-Like Receptors (TLRs) and Innate Immunity* (S. Bauer and G. Hartmann, eds), Springer, Heidelberg, Berlin, pp. 21–50.

Burgner, D., Jamieson, S.E. and Blackwell, J.M. (2006). Genetic susceptibility to infectious diseases: Big is beautiful, but will bigger be even better?. *Lancet Infect Dis* 6, 653–663.

Cambi, A. and Figdor, C.G. (2005). Levels of complexity in pathogen recognition by C-type lectins. *Curr Opin Immunol* 17, 345–351.

Campbell, C.J. and Ghazal, P. (2004). Molecular signatures for diagnosis of infection: Application of microarray technology. *J Appl Microbio* 96, 18–23.

Celestino, P.B., de Carvalho, L.R., de Freitas, L.M., Dorella, F.A., Martins, N.F., Pacheco, L.G., Miyoshi, A. and Azevedo, V. (2004). Update of microbial genome programs for bacteria and archaea. *Genet Mol Res* 3, 421–431.

Chapman, S.J., Khor, C.C., Vannberg, F.O., Frodsham, A., Walley, A., Maskell, N.A., Davies, C.W., Segal, S., Moore, C.E., Gillespie, S.H. et al. (2007a). IkappaB genetic polymorphisms and invasive pneumococcal disease. *Am J Resp Crit Care Med* 176, 181–187.

Chapman, S.J., Vannberg, F.O., Khor, C.C., Segal, S., Moore, C.E., Knox, K., Day, N.P., Davies, R.J.O., Crook, D.W. and Hill, A.V.S. (2007b). Functional polymorphisms in the FCN2 gene are not associated with invasive pneumococcal disease. *Mol Immunol* 44, 3267–3270.

Delbridge, L.M. and O'Riordan, M.X. (2007). Innate recognition of intracellular bacteria. *Curr Opin Immunol* 19, 10–16.

Densen, P., Weiler, J.M., Griffiss, J.M. and Hoffmann, L.G. (1987). Familial properdin deficiency and fatal meningococcemia. Correction of the bactericidal defect by vaccination. *New Engl J Med* 316, 922–926.

Dieffenbach, C.W. and Tramont, E.C. (2005). Innate (general or non-specific) host defense mechanisms. In *Principles and Practice of Infectious Diseases* (G.L. Mandell, J.E. Bennett and R. Dolin, eds), Elsevier Churchill Livingstone, Philadelphia, pp. 31–38.

Diez, E., Lee, S.H., Gauthier, S., Yaraghi, Z., Tremblay, M., Vidal, S. and Gros, P. (2003). Birc1e is the gene within the Lgn1 locus associated with resistance to *Legionella pneumophila*. *Nat Genet* 33, 55–60.

Domingo, P., Muniz-Diaz, E., Baraldes, M.A., Arilla, M., Barquet, N., Pericas, R., Juarez, C., Madoz, P. and Vazquez, G. (2002). Associations between Fc gamma receptor IIA polymorphisms and the risk and prognosis of meningococcal disease. *Am J Med* 112, 19–25.

Dubos, R. and Dubos, J. (1952). *The White Plague: Tuberculosis, Man and Society*. Little, Brown, and Co., Boston.

Echchannaoui, H., Frei, K., Schnell, C., Leib, S.L., Zimmerli, W. and Landmann, R. (2002). Toll-like receptor 2-deficient mice are highly susceptible to *Streptococcus pneumoniae* meningitis because of reduced bacterial clearing and enhanced inflammation. *J Infect Dis* 186, 798–806.

Echchannaoui, H., Bachmann, P., Letiembre, M., Espinosa, M. and Landmann, R. (2005). Regulation of *Streptococcus pneumoniae* distribution by Toll-like receptor 2 *in vivo*. *Immunobiology* 210, 229–236.

Elson, G., Dunn-Siegrist, I., Daubeuf, B. and Pugin, J. (2007). Contribution of Toll-like receptors to the innate immune response to Gram-negative and Gram-positive bacteria. *Blood* 109, 1574–1583.

Emonts, M., Hazelzet, J.A., de Groot, R. and Hermans, P.W. (2003). Host genetic determinants of *Neisseria meningitidis* infections. *Lancet Infect Dis* 3, 565–577.

Faber, J., Meyer, C.U., Gemmer, C., Russo, A., Finn, A., Murdoch, C., Zenz, W., Mannhalter, C., Zabel, B.U., Schmitt, H.J. et al. (2006). Human toll-like receptor 4 mutations are associated with susceptibility to invasive meningococcal disease in infancy. *Pediatr Infect Dis J* 25, 80–81.

Fernando, S.L. and Britton, W.J. (2006). Genetic susceptibility to mycobacterial disease in humans. *Immunol Cell Biol* 84, 125–137.

Fijen, C.A., Kuijper, E.J., te Bulte, M.T., Daha, M.R. and Dankert, J. (1999). Assessment of complement deficiency in patients with meningococcal disease in The Netherlands. *Clin Infect Dis* 28, 98–105.

Fijen, C.A., Bredius, R.G., Kuijper, E.J., Out, T.A., de, H.M., De Wit, A.P., Daha, M.R. and de Winkel, J.G. (2000). The role of Fcgamma receptor polymorphisms and C3 in the immune defence against *Neisseria meningitidis* in complement-deficient individuals. *Clin Exp Immunol* 120, 338–345.

Fournier, B. and Philpott, D.J. (2005). Recognition of *Staphylococcus aureus* by the innate immune system. *Clin Microbiol Rev* 18, 521–540.

Fritz, B. and Raczniak, G.A. (2002). Bacterial genomics: Potential for antimicrobial drug discovery. *Biodrugs* 16, 331–337.

Garred, P., Strom, J., Quist, L., Taaning, E. and Madsen, H.O. (2003). Association of mannose-binding lectin polymorphisms with sepsis and fatal outcome, in patients with systemic inflammatory response syndrome. *J Infect Dis* 188, 1394–1403.

Girard, R., Pedron, T., Uematsu, S., Balloy, V., Chignard, M., Akira, S. and Chaby, R. (2003). Lipopolysaccharides from Legionella and Rhizobium stimulate mouse bone marrow granulocytes via Toll-like receptor 2. *J Cell Sci* 116, 293–302.

Gomez, L.M., Anaya, J.M., Sierra-Filardi, E., Cadena, J., Corbi, A. and Martin, J. (2006). Analysis of DC-SIGN (CD209) functional variants in patients with tuberculosis. *Hum Immunol* 67, 808–811.

Haralambous, E., Weiss, H.A., Radalowicz, A., Hibberd, M.L., Booy, R. and Levin, M. (2003). Sibling familial risk ratio of meningococcal disease in UK Caucasians. *Epidemiol Infect* 130, 413–418.

Harding, D., Baines, P.B., Brull, D., Vassiliou, V., Ellis, I., Hart, A., Thomson, A.P., Humphries, S.E. and Montgomery, H.E. (2002). Severity of meningococcal disease in children and the angiotensin-converting enzyme insertion/deletion polymorphism. *Am J Resp Crit Care Med* 165, 1103–1106.

Hashimoto, M., Tawaratsumida, K., Kariya, H., Kiyohara, A., Suda, Y., Krikae, F., Kirikae, T. and Gotz, F. (2006). Not lipoteichoic acid but lipoproteins appear to be the dominant immunobiologically active compounds in *Staphylococcus aureus*. *J Immunol* 177, 3162–3169.

Hawn, T.R., Verbon, A., Lettinga, K.D., Zhao, L.P., Li, S.S., Laws, R.J., Skerrett, S.J., Beutler, B., Schroeder, L., Nachman, A. et al. (2003). A common dominant TLR5 stop codon polymorphism abolishes flagellin signaling and is associated with susceptibility to Legionnaires' disease. *J Exper Med* 198, 1563–1572.

Hawn, T.R., Verbon, A., Janer, M., Zhao, L.P., Beutler, B. and Aderem, A. (2005). Toll-like receptor 4 polymorphisms are associated with resistance to Legionnaires' disease. *Proc Natl Acad Sci USA* 102, 2487–2489.

Hawn, T.R., Dunstan, S.J., Thwaites, G.E., Simmons, C.P., Thuong, N.T., Lan, N.T., Quy, H.T., Chau, T.T., Hieu, N.T., Rodrigues, S. et al. (2006a). A polymorphism in Toll-interleukin 1 receptor domain containing adaptor protein is associated with susceptibility to meningeal tuberculosis. *J Infect Dis* 194, 1127–1134.

Hawn, T.R., Smith, K.D., Aderem, A. and Skerrett, S.J. (2006b). Myeloid differentiation primary response gene (88)- and toll-like receptor 2-deficient mice are susceptible to infection with aerosolized *Legionella pneumophila*. *J Infect Dis* 193, 1693–1702.

Hibberd, M.L., Sumiya, M., Summerfield, J.A., Booy, R. and Levin, M. (1999). Association of variants of the gene for mannose-binding lectin with susceptibility to meningococcal disease. Meningococcal Research Group. *Lancet* 353, 1049–1053.

Hill, A.V. (1998). The immunogenetics of human infectious diseases. *Annu Rev Immunol* 16, 593–617.

Hill, A.V. (2006). Aspects of genetic susceptibility to human infectious diseases. *Annu Rev Genet* 40, 469–486.

Hirschhorn, J.N., Lohmueller, K., Byrne, E. and Hirschhorn, K. (2002). A comprehensive review of genetic association studies. *Genet Med* 4, 45–61.

Hossain, H., Tchatalbachev, S. and Chakraborty, T. (2006). Host gene expression profiling in pathogen-host interactions. *Curr Opin Immunol* 18, 422–429.

Huang, S., Li, B. and Cao, H. (2007). Infectomics: Study of microbial infections using omic approaches. *Encyclopedia of Life Sciences*. John Wiley & Sons, Ltd., Chichester.

J. Craig Venter Institute (2006). Comprehensive microbial resource. http://cmr.tigr.org/tigr-scripts/CMR/shared/Genomes.cgi

Jenner, R.G. and Young, R.A. (2005). Insights into host responses against pathogens from transcriptional profiling. *Nat Rev Microbiol* 3, 281–294.

Johnson, C.M., Lyle, E.A., Omueti, K.O., Stepensky, V.A., Yegin, O., Alpsoy, E., Hamann, L., Schumann, R.R. and Tapping, R.I. (2007). Cutting edge: A common polymorphism impairs cell surface trafficking and functional responses of TLR1 but protects against leprosy. *J Immunol* 178, 7520–7524.

Kang, T.J. and Chae, G.T. (2001). Detection of Toll-like receptor 2 (TLR2) mutation in the lepromatous leprosy patients. *FEMS Immunol Med Microbiol* 31, 53–58.

Kawai, T. and Akira, S. (2005). Pathogen recognition with Toll-like receptors. *Curr Opin Immunol* 17, 338–344.

Khan, A.Q., Chen, Q., Wu, Z.Q., Paton, J.C. and Snapper, C.M. (2005). Both innate immunity and type 1 humoral immunity to *Streptococcus pneumoniae* are mediated by MyD88 but differ in their relative levels of dependence on toll-like receptor 2. *Infect Immun* 73, 298–307.

Kikuchi, T., Kobayashi, T., Gomi, K., Suzuki, T., Tokue, Y., Watanabe, A. and Nukiwa, T. (2004). Dendritic cells pulsed with live and dead *Legionella pneumophila* elicit distinct immune responses. *J Immunol* 172, 1727–1734.

Knapp, S., Wieland, C.W., van't Veer, C., Takeuchi, O., Akira, S., Florquin, S. and van der Poll, T. (2004). Toll-like receptor 2 plays a role in the early inflammatory response to murine pneumococcal pneumonia but does not contribute to antibacterial defense. *J Immunol* 172, 3132–3138.

Kobayashi, K.S., Chamaillard, M., Ogura, Y., Henegariu, O., Inohara, N., Nunez, G. and Flavell, R.A. (2005). Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. *Science* 307, 731–734.

Koedel, U., Angele, B., Rupprecht, T., Wagner, H., Roggenkamp, A., Pfister, H.W. and Kirschning, C.J. (2003). Toll-like receptor 2 participates in mediation of immune response in experimental pneumococcal meningitis. *J Immunol* 170, 438–444.

Kronborg, G., Weis, N., Madsen, H.O., Pedersen, S.S., Wejse, C., Nielsen, H., Skinhoj, P. and Garred, P. (2002). Variant mannose-binding lectin alleles are not associated with susceptibility to or outcome of invasive pneumococcal infection in randomly included patients. *J Infect Dis* 185, 1517–1520.

Lorenz, E., Mira, J.P., Cornish, K.L., Arbour, N.C. and Schwartz, D.A. (2000). A novel polymorphism in the toll-like receptor 2 gene and its potential association with staphylococcal infection. *Infect Immun* 68, 6398–6401.

Lorenz, E., Mira, J.P., Frees, K.L. and Schwartz, D.A. (2002). Relevance of mutations in the TLR4 receptor in patients with Gram-negative septic shock. *Arch Inter Med* 162, 1028–1032.

Malley, R., Henneke, P., Morse, S.C., Cieslewicz, M.J., Lipsitch, M., Thompson, C.M., Kurt-Jones, E., Paton, J.C., Wessels, M.R. and Golenbock, D.T. (2003). Recognition of pneumolysin by Toll-like receptor 4 confers resistance to pneumococcal infection. *Proc Natl Acad Sci USA* 100, 1966–1971.

Maskell, N.A., Davies, C.W., Nunn, A.J., Hedley, E.L., Gleeson, F.V., Miller, R., Gabe, R., Rees, G.L., Peto, T.E., Woodhead, M.A. et al. (2005). U.K. Controlled trial of intrapleural streptokinase for pleural infection. *New Engl J Med* 352, 865–874.

Mira, M.T. (2006). Genetic host resistance and susceptibility to leprosy. *Microbes Infect* 8, 1124–1131.

Mira, M.T., Alcais, A., Nguyen, V.T., Moraes, M.O., Di, F.C., Vu, H.T., Mai, C.P., Nguyen, T.H., Nguyen, N.B., Pham, X.K. et al. (2004). Susceptibility to leprosy is associated with PARK2 and PACRG. *Nature* 427, 636–640.

Modlin, R.L. and Bloom, B.R. (2001). Immunology. Chip shots – Will functional genomics get functional?. *Science* 294, 799–801.

Moens, L., Van Hoeyveld, E., Peetermans, W.E., De Boeck, C., Verhaegen, J. and Bossuyt, X. (2006a). Mannose-binding lectin genotype and invasive pneumococcal infection. *Hum Immunol* 67, 605–611.

Moens, L., Van Hoeyveld, E., Verhaegen, J., De Boeck, K., Peetermans, W.E. and Bossuyt, X. (2006b). Fc[gamma]-receptor IIA genotype and invasive pneumococcal infection. *Clin Immunol* 118, 20–23.

Moens, L., Verhaegen, J., Pierik, M., Vermeire, S., De Boeck, K., Peetermans, W.E. and Bossuyt, X. (2007). Toll-like receptor 2 and Toll-like receptor 4 polymorphisms in invasive pneumococcal disease. *Microbes Infect* 9, 15–20.

Monaghan, R.L. and Barrett, J.F. (2006). Antibacterial drug discovery – Then, now and the genomics future. *Biochem Pharmacol* 71, 901–909.

Montes, A.H., Asensi, V., Alvarez, V., Valle, E., Ocana, M.G., Meana, A., Carton, J.A., Paz, J., Fierer, J. and Celada, A. (2006). The Toll-like receptor 4 (Asp299Gly) polymorphism is a risk factor for Gram-negative and haematogenous osteomyelitis. *Clin Exp Immunol* 143, 404–413.

Moore, C.E., Segal, S., Berendt, A.R., Hill, A.V.S. and Day, N.P.J. (2004). Lack of association between Toll-like receptor 2 polymorphisms and susceptibility to severe disease caused by *Staphylococcus aureus*. *Clin Diagn Lab Immunol* 11, 1194–1197.

Motulsky, A.G. (1960). Metabolic polymorphisms and the role of infectious diseases in human evolution. *Hum Biol* 32, 28–62.

Murray, P.J. (2005). NOD proteins: An intracellular pathogen-recognition system or signal transduction modifiers?. *Curr Opin Immunol* 17, 352–358.

National Center for HIV SATPC (2005). Controlling tuberculosis in the United States. *MMWR Recommendations and Reports* 54, 1–81.

Noakes, K., Pebody, R.G., Gungabissoon, U., Stowe, J. and Miller, E. (2006). Pneumococcal polysaccharide vaccine uptake in England, 1989–2003, prior to the introduction of a vaccination programme for older adults. *J Public Health* 28, 242–247.

Ogus, A.C., Yoldas, B., Ozdemir, T., Uguz, A., Olcen, S., Keser, I., Coskun, M., Cilli, A. and Yegin, O. (2004). The Arg753GLn polymorphism of the human toll-like receptor 2 gene in tuberculosis disease. *Eur Respir J* 23, 219–223.

Olesen, R., Wejse, C., Velez, D.R., Bisseye, C., Sodemann, M., Aaby, P., Rabna, P., Worwui, A., Chapman, H., Diatta, M. et al. (2007). DC-SIGN (CD209), pentraxin 3 and vitamin D receptor gene variants associate with pulmonary tuberculosis risk in West Africans. *Genes Immun* 8, 456–467.

Opitz, B., Puschel, A., Schmeck, B., Hocke, A.C., Rosseau, S., Hammerschmidt, S., Schumann, R.R., Suttorp, N. and Hippenstiel, S. (2004). Nucleotide-binding oligomerization domain proteins are innate immune receptors for internalized *Streptococcus pneumoniae*. *J Biol Chem* 279, 36426–36432.

Pamer, E.G. (2004). Immune responses to *Listeria monocytogenes*. *Nat Rev Immunol* 4, 812–823.

Platonov, A.E., Shipulin, G.A., Vershinina, I.V., Dankert, J., van de Winkel, J.G. and Kuijper, E.J. (1998). Association of human Fc gamma RIIa (CD32) polymorphism with susceptibility to and severity of meningococcal disease. *Clin Infect Dis* 27, 746–750.

Power, M.R., Peng, Y., Maydanski, E., Marshall, J.S. and Lin, T.J. (2004). The development of early host response to *Pseudomonas aeruginosa* lung infection is critically dependent on myeloid differentiation factor 88 in mice. *J Biol Chem* 279, 49315–49322.

Read, R.C., Pullin, J., Gregory, S., Borrow, R., Kaczmarski, E.B., di Giovine, F.S., Dower, S.K., Cannings, C. and Wilson, A.G. (2001). A functional polymorphism of toll-like receptor 4 is not associated with likelihood or severity of meningococcal disease. *J Infect Dis* 184, 640–642.

Roy, S., McGuire, W., Mascie-Taylor, C.G., Saha, B., Hazra, S.K., Hill, A.V. and Kwiatkowski, D. (1997). Tumor necrosis factor promoter polymorphism and susceptibility to lepromatous leprosy. *J Infect Dis* 176, 530–532.

Roy, S., Frodsham, A., Saha, B., Hazra, S.K., Mascie-Taylor, C.G. and Hill, A.V. (1999). Association of vitamin D receptor genotype with leprosy type. *J Infect Dis* 179, 187–191.

Roy, S., Knox, K., Segal, S., Griffiths, D., Moore, C.E., Welsh, K.I., Smarason, A., Day, N.P., McPheat, W.L., Crook, D.W. et al. (2002). MBL genotype and risk of invasive pneumococcal disease: A case-control study. *Lancet* 359, 1569–1573.

Salimans, M.M., Bax, W.A., Stegeman, F., van, D.M., Bartelink, A.K. and van, D.H. (2004). Association between familial deficiency of mannose-binding lectin and mutations in the corresponding gene and promoter region. *Clin Diagn Lab Immunol* 11, 806–807.

Sanchez, E., Orozco, G., Lopez-Nevot, M.A., Jimenez-Alonso, J. and Martin, J. (2004). Polymorphisms of toll-like receptor 2 and 4 genes in rheumatoid arthritis and systemic lupus erythematosus. *Tissue Antigens* 63, 54–57.

Schroder, N.W. and Schumann, R.R. (2005). Single nucleotide polymorphisms of Toll-like receptors and susceptibility to infectious disease. *Lancet Infect Dis* 5, 156–164.

Schroder, N.W., Hermann, C., Hamann, L., Gobel, U.B., Hartung, T. and Schumann, R.R. (2003a). High frequency of polymorphism Arg753Gln of the Toll-like receptor-2 gene detected by a novel allele-specific PCR. *J Mol Med* 81, 368–372.

Schroder, N.W., Morath, S., Alexander, C., Hamann, L., Hartung, T., Zahringer, U., Gobel, U.B., Weber, J.R. and Schumann, R.R. (2003b). Lipoteichoic acid (LTA) of *Streptococcus pneumoniae* and *Staphylococcus aureus* activates immune cells via Toll-like receptor (TLR)-2, lipopolysaccharide-binding protein (LBP), and CD14, whereas TLR-4 and MD-2 are not involved. *J Biol Chem* 278, 15587–15594.

Seki, E., Tsutsui, H., Tsuji, N.M., Hayashi, N., Adachi, K., Nakano, H., Futatsugi-Yumikura, S., Takeuchi, O., Hoshino, K., Akira, S. et al. (2002). Critical roles of Myeloid Differentiation Factor 88-dependent proinflammatory cytokine release in early phase clearance of *Listeria monocytogenes* in mice. *J Immunol* 169, 3863–3868.

Skerrett, S.J., Liggitt, H.D., Hajjar, A.M. and Wilson, C.B. (2004). Cutting edge: Myeloid differentiation factor 88 is essential for pulmonary host defense against *Pseudomonas aeruginosa* but not *Staphylococcus aureus*. *J Immunol* 172, 3377–3381.

Smirnova, I., Mann, N., Dols, A., Derkx, H.H., Hibberd, M.L., Levin, M. and Beutler, B. (2003). Assay of locus-specific genetic load implicates rare Toll-like receptor 4 mutations in meningococcal susceptibility. *Proc Natl Acad Sci USA* 100, 6075–6080.

Soborg, C., Andersen, A.B., Range, N., Malenganisho, W., Friis, H., Magnussen, P., Temu, M.M., Changalucha, J., Madsen, H.O. and Garred, P. (2007). Influence of candidate susceptibility genes on tuberculosis in a high endemic region. *Mol Immunol* 44, 2213–2220.

Spath, P.J., Sjoholm, A.G., Fredrikson, G.N., Misiano, G., Scherz, R., Schaad, U.B., Uhring-Lambert, B., Hauptmann, G., Westberg, J., Uhlen, M. et al. (1999). Properdin deficiency in a large Swiss family: Identification of a stop codon in the properdin gene, and association of meningococcal disease with lack of the IgG2 allotype marker G2m(n). *Clin Exp Immunol* 118, 278–284.

Srivastava, A., Henneke, P., Visintin, A., Morse, S.C., Martin, V., Watkins, C., Paton, J.C., Wessels, M.R., Golenbock, D.T. and Malley, R. (2005). The apoptotic response to pneumolysin is Toll-like receptor 4 dependent and protects against pneumococcal disease. *Infect Immun* 73, 6479–6487.

Stead, W.W. (1992). Genetics and resistance to tuberculosis. Could resistance be enhanced by genetic engineering?. *Ann Intern Med* 116, 937–941.

Stead, W.W., Senner, J.W., Reddick, W.T. and Lofgren, J.P. (1990). Racial differences in susceptibility to infection by *Mycobacterium tuberculosis*. *New Engl J Med* 322, 422–427.

Stein, C.M., Zalwango, S., Chiunda, A.B., Millard, C., Leontiev, D. V., Horvath, A.L., Cartier, K.C., Chervenak, K., Boom, W.H., Elston, R.C. et al. (2007). Linkage and association analysis of candidate genes for TB and TNFalpha cytokine expression: Evidence for association with IFNGR1, IL-10, and TNF receptor 1 genes. *Hum Genet* 121, 663–673.

Takeuchi, O., Hoshino, K. and Akira, S. (2000). Cutting edge: TLR2-deficient and MyD88-deficient mice are highly susceptible to *Staphylococcus aureus* infection. *J Immunol* 165, 5392–5396.

Tapping, R.I., Syroid, D.E., Bilan, P.T. and Capone, J.P. (1993). The 5′ flanking sequence negatively modulates the *in vivo* expression and *in vitro* transcription of a human tRNA gene. *Nucleic Acids Res* 21, 4476–4482.

Thuong, N.T., Hawn, T.R., Thwaites, G.E., Chau, T.T., Lan, N.T., Quy, H.T., Hieu, N.T., Aderem, A., Hien, T.T., Farrar, J.J. et al. (2007). A polymorphism in human TLR2 is associated with increased susceptibility to tuberculous meningitis. *Genes Immun* 8, 422–428.

Torres, D., Barrier, M., Bihl, F., Quesniaux, V.J.F., Maillet, I., Akira, S., Ryffel, B. and Erard, F. (2004). Toll-like receptor 2 is required for optimal control of *Listeria monocytogenes* infection. *Infect Immun* 72, 2131–2139.

Turvey, S.E. and Hawn, T.R. (2006). Towards subtlety: Understanding the role of Toll-like receptor signaling in susceptibility to human infections. *Clin Immunol* 120, 1–9.

van der Pol, W.L., Huizinga, T.W., Vidarsson, G., van der Linden, M.W., Jansen, M.D., Keijsers, V., de Straat, F.G., Westerdaal, N.A., de Winkel, J.G. and Westendorp, R.G. (2001). Relevance of Fcgamma receptor and interleukin-10 polymorphisms for meningococcal disease. *J Infect Dis* 184, 1548–1555.

Westberg, J., Fredrikson, G.N., Truedsson, L., Sjoholm, A.G. and Uhlen, M. (1995). Sequence-based analysis of properdin deficiency: Identification of point mutations in two phenotypic forms of an X-linked immunodeficiency. *Genomics* 29, 1–8.

Wright, E.K., Goodart, S.A., Growney, J.D., Hadinoto, V., Endrizzi, M.G., Long, E.M., Sadigh, K., Abney, A.L., Bernstein-Hanley, I. and Dietrich, W.F. (2003). Naip5 affects host susceptibility to the intracellular pathogen *Legionella pneumophila*. *Curr Biol* 13, 27–36.

Yee, A.M., Ng, S.C., Sobel, R.E. and Salmon, J.E. (1997). Fc gammaRIIA polymorphism as a risk factor for invasive pneumococcal infections in systemic lupus erythematosus. *Arthritis Rheum* 40, 1180–1182.

Yee, A.M., Phan, H.M., Zuniga, R., Salmon, J.E. and Musher, D.M. (2000). Association between FcgammaRIIa-R131 allotype and bacteremic pneumococcal pneumonia. *Clin Infect Dis* 30, 25–28.

Yoshimura, A., Lien, E., Ingalls, R.R., Tuomanen, E., Dziarski, R. and Golenbock, D. (1999). Cutting edge: Recognition of Gram-positive bacterial cell wall components by the innate immune system occurs via Toll-like receptor 2. *J Immunol* 163, 1–5.

Yu, S.L., Chen, H.W., Yang, P.C., Peck, K., Tsai, M.H., Chen, J.J. and Lin, F.Y. (2004). Differential gene expression in Gram-negative and Gram-positive sepsis. *Am J Resp Crit Care Med* 169, 1135–1143.

Yuan, F.F., Wong, M., Pererva, N., Keating, J., Davis, A.R., Bryant, J.A. and Sullivan, J.S. (2003). FcgammaRIIA polymorphisms in *Streptococcus pneumoniae* infection. *Immunol Cell Biol* 81(3), 192–195.

CHAPTER

Genomics in the Evaluation and Management of Sepsis

Christopher W. Woods, Robert J. Feezor and Stephen F. Kingsmore

INTRODUCTION

An 18-year-old female college freshman who lives in a dormitory with 50 other women presents to her local acute care center with 12 h of fever and headache. She is referred to the emergency department and develops a petechial rash while in transit. In the emergency department she becomes hypotensive and lethargic. She is admitted to the intensive care unit (ICU). Blood and cerebrospinal fluid cultures taken at the time of admission grow *Neisseria meningitidis*, serogroup C. She dies 2 days later with multi-organ failure despite aggressive attempts at resuscitation and support. No other students in her dormitory become ill after prophylaxis is provided. Her devastated parents wish to know if her two siblings are at high risk of bad outcome with a similar infection.

A 64-year-old hospitalized, African American veteran is admitted to the ICU after developing fever and hypotension subsequent to aortobifemoral bypass graft procedure. Blood cultures grow enteric Gram-negative rods. His fever persists and he develops disseminated intravascular coagulopathy, acute renal failure, acute lung injury, and succumbs to his illness after 3 weeks. His physicians considered drotrecogin alfa, but chose not to give it in the setting of recent surgery.

A 52-year-old woman with a several month history of fever with negative evaluation including blood cultures presents with acute renal failure and severe aortic regurgitation. A transthoracic echocardiogram reveals vegetations on her aortic valve leaflet. Immunohistochemistry at the time of valve replacement confirms *Bartonella henselae*. The patient raises feral cats. Her primary care physician is frustrated that he was unable to diagnose the infection earlier.

Each of these cases represents extremes of the heterogeneity of the sepsis spectrum and, yet, each highlights the potential benefits of the genomic revolution in the way we identify and care for patients presenting with the sepsis syndrome. In the past decade new techniques have been developed to investigate the causes of complex diseases. In this chapter, we will review the relevant literature and look toward the future.

Sepsis is a common, heterogeneous clinical entity that is defined by the physiological changes known collectively as the systemic inflammatory response syndrome (SIRS) that occur in response to a presumed infectious etiology (Table 59.1) (Bone et al., 1989, 1992; Sands et al., 1997). The insulting agent may be bacterial, viral, fungal, or parasitic with more than 80% of cases originating from a pulmonary, genitourinary, or abdominal source (Opal and Cohen, 1999). However, sepsis is not a single disease, but rather a heterogeneous syndrome that is expressed through the interaction of networks of biochemical mediators and inflammatory cascades. Clinical expression is variable and its severity is influenced by the infectious etiology, the genetic background of the patient, comorbid conditions, the time to clinical intervention, and the supportive care provided by the physician.

Patients with coincident acute organ dysfunction are considered to have severe sepsis. Patients with sepsis who fail to maintain their blood pressure despite adequate hydration

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.

TABLE 59.1	Criteria for systemic inflammatory response syndrome and sepsis
SIRS	• Temp >38 C or <36 C • Heart rate >90 beats per minute • White Blood Cell count >12,000 cell/cc, <4000 cells/cc, or >10% band forms • Respiratory rate >20 breaths per minute or $PaCO_2 < 32$ torr
Sepsis	SIRS criteria and evidence of infection, or: • White cells in normally sterile body fluid • Perforated viscus • Radiographic evidence of pneumonia • Syndrome associated with high risk of infection
Severe Sepsis	Sepsis criteria and evidence of organ dysfunction including: • Cardiovascular: SBP ≤ 90 mmHg, MAP ≤ 70 mmHg for at least 1 hour despite adequate volume resuscitation or the use of vasopressors • Renal: urine output <0.5 mL/kg body weight/hour for 1 hour despite adequate volume resuscitation • Pulmonary: PaO_2/FiO2 ≤250 if other organ dysfunction present or ≤200 if the lung is the only dysfunctional organ • Hematologic: platelet count <80,000/cc or decreased by 50% within 3 days • Metabolic: pH ≤7.3 or base deficit >5.0 mmol/L and plasma lactate >1.5 x upper limit of normal

are considered to have septic shock. Severe sepsis is a major cause of morbidity and mortality with an annual incidence of 50–100 cases per 100,000 persons in several industrialized nations (Martin et al., 2003). In the United States, there are approximately 750,000 new cases of severe sepsis annually with an economic impact approaching $17 billion (Angus and Wax, 2001). Despite an enormous investment in critical care resources, 20–50% of patients with sepsis died; it is the third leading cause of infectious death and tenth leading cause of death overall.

The incidence of severe sepsis is increasing by approximately 9% annually. Between 1979 and 2000, the incidence increased from 82.7/100,000 to 240.4/100,000 (Martin, 2003). This increase is multifactorial, resulting from increased awareness and documentation of sepsis, "graying" of the population, greater use of invasive procedures for the diagnosis and monitoring of critically ill patients, emergence of antibiotic-resistant organisms, and increasing prevalence of immunocompromised patients (e.g., malignancy, AIDS, transplant recipients, diabetes mellitus, alcoholism, and malnutrition) (Parrillo et al., 1990). Furthermore, the relative contribution of etiological organisms has changed substantially over time. Gram-positive organisms superseded Gram-negative organisms in predominance in 1987, and fungal sepsis has increased by more than 200% since 1980 (Martin, 2003).

Genomic medicine appears well situated to assist in the identification of individuals at substantial risk for certain infections, to stratify subsets of individuals who are likely to progress to adverse outcomes or most benefit from therapeutic intervention, and to facilitate the rapid identification of etiological organisms. However, the complex physiology and epidemiology of sepsis, together with the diversity of physicians confronted by this illness, have slowed progress. Much of the literature focuses on the tremendous achievements in describing the genomes of various bacterial, viral, and fungal pathogens; however, this chapter focuses on the human genome and its response to the infectious perturbations that result in sepsis.

GENETIC POLYMORPHISMS ASSOCIATED WITH SEPSIS

Microbiological infections and their complications, such as sepsis and severe sepsis, occur at the interface of host genes, microbial genes, and the environment. While exposure to a microbial agent is necessary, it is not sufficient to cause sepsis in the host. Regardless of etiology, sepsis stimulates the host's immune, inflammatory, and coagulation responses (Figure 59.1). Although the general direction of activation is similar between persons, there are notable differences in response to infection that may have important clinical implications (Bellamy and Hill, 1998; Burgner and Levin, 2003; Choi et al., 2001). Therefore, sepsis progression depends on the relative weight of host defense versus microbe virulence. Several recent studies have shown evidence that host and pathogen molecular interactions drive adaptive evolution of the immune system – a host:pathogen genetic arms race (Hughes and Nei, 1988; Sackton et al., 2007; Schlenke and Begun, 2003). Accordingly, it is useful to evaluate how the genotype of patients determines their individual susceptibility and response to infection.

Family and twin studies have demonstrated familial or twin aggregation of some, but not all sepsis outcomes and clinical presentations on a common genetic basis (Härtel et al., 2007; Jepson et al., 1995). There are several dramatic examples of lack

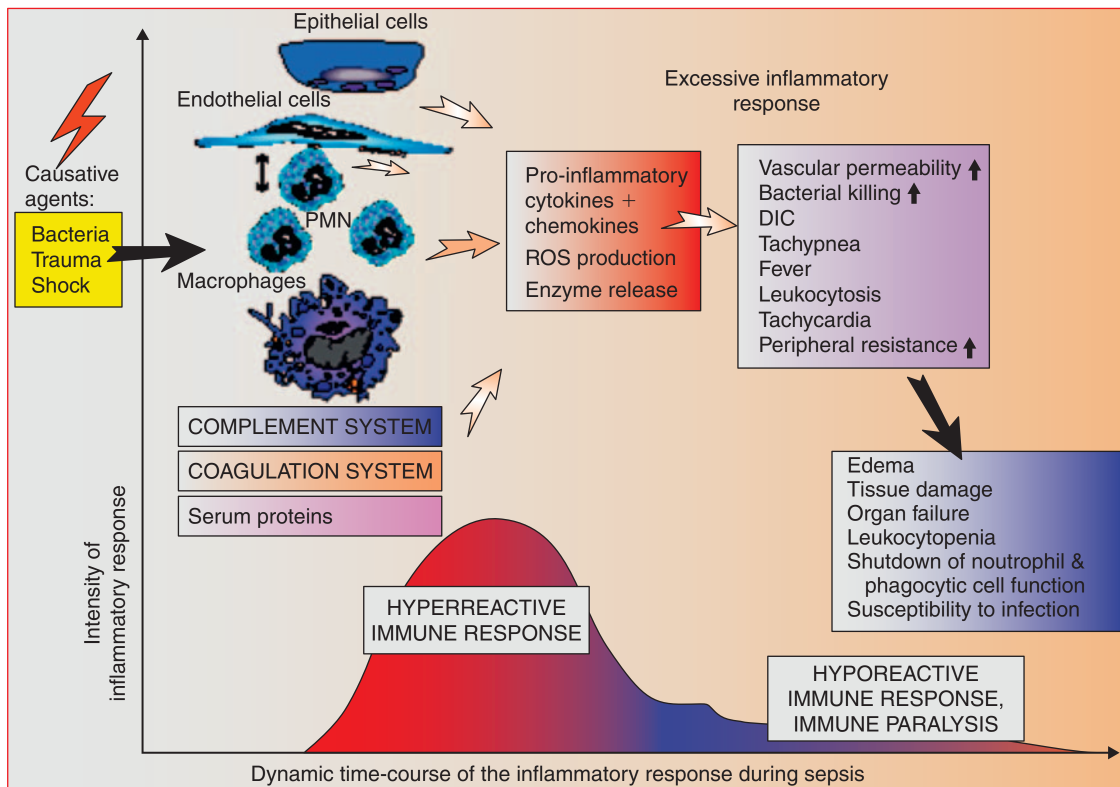


Figure 59 .1 Time course and intensity of host immune inflammatory and coagulation responses during sepsis (reproduced with permission from Riedemann, N.C., Guo, P.-F. and Ward, P.A. (2003) "Novel Strategies for the treatment of sepsis". *Nat. Med.* 9: 517–524). Microbiologic antigens and factors cause activation of the innate and cognate immune system, inflammatory response and coagulation system in sepsis, leading to production of proinflammatory cytokines, activation of the alternate pathway of complement and upregulation of adhesion and signaling molecules on white blood cells and blood vessel endothelium. ROS: Reactive oxygen species. DIC: Disseminated intracascular coagulation.

of host defense factors resulting from Mendelian mutations. For example, autosomal dominant inheritance in meningococcal meningitis susceptibility in a few large families due to a defective properdin gene in the alternative complement pathway (Densen et al., 1987). Mutations in ATP-sensitive K^+ channels expressed in murine endothelial cells and coronary artery smooth muscle may cause profound susceptibility to mouse cytomegalovirus (Croker et al., 2007). Other examples of susceptibility alleles are variants in immune response genes, including the MHC, CCR5 in HIV infection, and Toll receptor 2 in Mycobacteria and non-immune response genes such as Nramp in Mycobacteria (Drennan et al., 2004; Liu et al., 1996; Samson et al., 1996; Skamene et al., 1998). Linkage-based methods and candidate gene-based association studies are being extensively applied to microbial infection outcomes and sepsis but, to date, have met with limited success in the identification of risk alleles for host gene:pathogen interaction (Botstein and Risch, 2003; Freimer and Sabatti, 2004; Hill, 2006). In sepsis, most studies have investigated candidate genes that are involved in pathogen detection (e.g., toll-like receptors, TLRs), the inflammatory response (e.g., tumor necrosis factor, TNF-α) or coagulation (e.g., plasminogen activator inhibitor, PAI). Several reviews of the associations between candidate gene polymorphisms and the risk and outcome of sepsis have been published (Arcaroli et al., 2005; Lin and Albertson, 2004; Majetschak et al., 2002; Mira et al., 1999). However, ultimately, complex diseases like sepsis represent the cumulative effect of many minor susceptibility alleles and/or a small number of alleles of large effect. Several of those alleles of potential large effect are discussed briefly in this chapter and in more detail in adjoining chapters in this volume.

Limitations of Sepsis Gene-Association Studies

The analysis of single nucleotide polymorphisms (SNPs) among various susceptibility genes, such as those coding for pro- and anti-inflammatory mediators, in the response to sepsis has the potential to nominate prediction tools that will permit the clinician to more precisely determine the type of therapy a particular patient should receive. In fact, early identification of a genotypic risk may even help guide the introduction of prophylactic therapy. However, unraveling the genetic variation in sepsis is complicated and close attention must be paid to study design issues, including the selection of an appropriate study population and sample size and understanding gene–environment interactions.

A lack of replication among studies provides considerable concern in the interpretation of results. For example, an initial study may identify an allele with large estimated genetic effects, but subsequent studies fail to corroborate the results (Goring et al., 2001; Hirschhorn et al., 2002; Ioannidis et al., 2001; Lander and Kruglyak, 1995). Biological explanations of inconsistent results include unacknowledged confounding heterogeneity, such as poorly defined phenotypes, heterogeneous genetic sources for the phenotype (genocopies), population diversity (ethnic ancestry), population-specific linkage disequilibrium (LD), heterogeneous genetic and epigenetic backgrounds, or heterogeneous environmental influences (phenocopies). Analytic reasons for problems with reproducibility include failure to control the rate of false discoveries, lack of power, model misspecification, and heterogeneous bias in estimated effects among studies (Cardon and Bell, 2001; Cardon and Palmer, 2003; Redden and Allison, 2003; Sillanpaa and Auranen, 2004). Among these, the most frequent source of non-replication has been lack of power due to the limited number of individuals genotyped and phenotyped (Lohmueller et al., 2003; Risch, 2000).

Pathogen Recognition/Signaling

Toll-Like Receptors

There has been substantial progress in defining variants of several genes and pathways implicated in infectious disease susceptibility; however, none have caused more excitement than investigation of genes affecting innate immunity, particularly newly discovered pattern recognition receptors and their associated signaling pathways. The TLRs have captured the attention of most investigators in this field.

Innate immunity, recognition of invading microorganisms, is mediated by a set of soluble and membrane receptors that recognize conserved, pathogen-associated molecular patterns (PAMPs) shared among each class of infectious agents but absent in higher eukaryotes. Having a cytoplasmic domain that bears homology to the interleukin (IL)-1 receptor, TLRs have been conserved throughout diverse life forms including plants, insects, and mammals. Thus far, 10 human TLRs have been identified that play a role in sensing pathogens. Activation of TLRs stimulates macrophages resulting in the elaboration of proinflammatory cytokines and antimicrobial molecules such as nitric oxide and defensins. Concurrently, stimulated dendritic cells migrate to lymph nodes and over-express antigen MHC and costimulatory molecules (CD80/CD86). Therefore, TLRs are an essential link between innate and adaptive immunity.

The discovery and characterization of TLRs as a key pattern recognition molecules for pathogens (or PAMPs, Beutler and Reitschel, 2003) and initiators of the innate immune responses (Akira et al., 2001; Creagh and O'Neill, 2006) stimulated many investigators to further characterize the relevance of polymorphisms in these receptors and susceptibility to infectious disease. Sepsis was an early target for characterization of the relevance of polymorphisms in TLRs. The TLR4 protein activated NF-kappa B and increased expression of the pro-inflammatory cytokines, IL-1, IL-6, and IL-8 in cultured human cells. and TLR4 deficient mice were resistant to systemic endotoxin exposure, but remained susceptible to Gram-negative infections (Oureshi et al., 1999). Asp229Gly and Thr399Ile are common TLR4 missense mutations that affect the extracellular domain and some reports suggest an association with varied human responses to inhaled endotoxin (Arbour et al., 2000). A number of small studies associated the Asp299Gly variant of the *TLR4* gene with increased susceptibility to either Gram-negative infections (Agnese et al., 2002; Lorenz et al., 2002) or SIRS (Child et al., 2003). However, an evaluation of an intravenous lipopolysaccharide (LPS) challenge did not reveal an association with TLR4 mutations (Calvano et al., 2006). In addition, Smirnova et al. found an excess of rare TLR4 coding changes in meningococcal cases compared with controls (Smirnova et al., 2003). However, the study did not support an association with the functional Asp299Gly or Thr399Ile polymorphisms of *TLR4*. The absence of an association was also supported by a much larger case–control study (Read et al., 2001). Together, these studies of TLR4 polymorphisms suggest that individuals with the 299/399 polymorphisms may have an aberrant response to certain, but not all, Gram-negative bacterial diseases, resulting in an increased susceptibility to infection and severity of disease.

TLR2 and TLR5 have also been targets of interest. TLR2 is most notable for detecting a wide repertoire of pathogens, including Gram-positive or Gram-negative bacteria, mycobacteria, fungi, viruses, and parasites. Largely this results from an ability to recognize ligands as a heterodimer with TLR1 or TLR6 (Lorenz, 2006). Of particular importance, TLR2 has been linked to the recognition of Gram-positive bacteria, which are now the leading cause of sepsis (Martin, 2003). TLR2 accomplishes this via response to peptidoglycan, lipoteichoic acid, and a variety of macromolecules in Gram-positive bacteria such as *S. aureus*. The relationship between TLR2 polymorphisms (primarily Arg753Gln) and *S. aureus* infection has been examined in several studies with contradictory results (Lorenz et al., 2000; Moore et al., 2004). Therefore, additional investigations are warranted to confirm this association.

TLR5, which recognizes bacterial flagellin from both Gram-positive and Gram-negative bacteria, activates NF-kappa B and the release of pro-inflammatory cytokines in response to this bacterial antigen (Hayashi et al., 2001). A stop codon polymorphism (Arg392TER) has been identified in the TLR5 gene and is associated with an increased susceptibility to *Legionella pneumophila* (Hawn et al., 2003). suggesting that this allele may increase susceptibility to pneumonia associated with flagellated organisms.

NOD-Like and RIG-Like Receptors

TLRs are involved in the recognition of all types of pathogens regardless of location. In contrast, NOD-like receptors (NLRs) primarily recognize cytoplasmic bacterial pathogens (Inohara and Nunez, 2003). Many NLRs have caspase recruitment domains (CARD). These domains are involved in the assembly of protein complexes that promote apoptosis. These CARD proteins may also participate in NF-kappa B signaling pathways. Hill et al. have studied a truncation variant of the human

CARD8 gene and found that African children homozygous for this inactivating mutation are susceptible to non-typhoidal Salmonella bacteremia (Hill, 2006).

A third family of pattern recognition receptors described more recently is the DExD/H Box RNA helicase family of RIG-I-like receptor (RLR) genes, which includes MDA5 as well as RIG-1 (Creagh et al., 2006). These receptors appear to differentially recognize double-stranded RNA from various viruses (Kato et al., 2006). An amino acid change in MDA5 has been associated with susceptibility to type 1 diabetes lending support to a viral etiology for this disease (Smyth et al., 2006).

Mannose-Binding Lectin

Mannose-binding lectin (MBL) is an acute-phase protein that can opsonize many bacterial and fungal pathogens and activate complement (Kuhlman et al., 1989). Approximately one-third of most human populations are heterozygous for one of the variants in the coding region of exon 1 that lead to lower MBL concentrations. Individuals with 2 copies of low MBL haplotypes have a higher risk of pneumococcal sepsis (Kronborg and Garred, 2002; Roy et al., 2002), and there is less well-replicated evidence for susceptibility to other bacterial pathogens and candidiasis.

Although multiple studies indicate that MBL plays a role in mitigating certain pathogens, the high rates of haplotypes that specify low MBL levels in certain ethnic groups suggest that a relative lack of MBL might be beneficial to the host under other circumstances. In particular, studies of the MBL-MASP pathway suggest that it may play a role in reducing reperfusion injury in the heart and resulting in relative protection against the complications of myocardial infarction (Walsh et al., 2005).

CD 14

CD14 functions as an anchor protein and is a ubiquitous pattern recognition receptor, specific for LPS and other ligands. Performing as a co-receptor for TLR4, it is shed by monocytes to facilitate LPS signaling for all other cells. Two polymorphisms at the promoter region of the CD14 gene have been widely explored. The -159 T allele has been associated with increased prevalence of Gram-negative infections and sepsis, but not with shock or survival (Gibot et al., 2002; Sutherland et al., 2005). However, other studies have yielded contradictory results. More recently, the -260 T allele has been linked to increased survival in a Brazilian sepsis population (D'Avila et al., 2006).

Intracellular Signaling Molecules

The IL1 receptor-associated kinase-4 (IRAK4) is an intracellular kinase that transduces intracellular signals conveyed by the TLRs and IL1 receptors. Patients with IRAK4 deficiency are at increased risk for invasive bacterial disease (Von Bernuth et al., 2005). Homozygous individuals for either of the two known variant alleles appear to have completely inhibited expression of IRAK4, thereby inhibiting activation of NF-kappa B and NF-kappa B-dependent pro-inflammatory mediators (Lasker et al., 2005).

Cytokine Polymorphisms

Genetic variation in the pro-inflammatory cytokines TNF-α, TNF-β, IL-6, IL-8 and macrophage migration inhibitory factor (MIF) and the anti-inflammatory cytokines IL-10 and IL-1 RA are the most extensively studied in relation to sepsis. Patients predisposed to a balanced anti-, pro-inflammatory response appear to have better chances for survival (Figure 59.1). Defects in regulation occur when the balance is shifted and allows for unmitigated inflammation causing tissue and organ damage or the inability to extirpate invading pathogens.

Pro-inflammatory Cytokines

TNF-α is a primary mediator in sepsis and an initial trigger of the immune response. Most studies have investigated the significance of the -308 A allele of the promoter region and have yielded mixed results on outcome of sepsis and no association with the expression of TNF-α (Mira et al., 1999; Stüber et al., 1996). Potential associations with polymorphisms at −376 and position −238 are also not convincing. Bayley et al. reviewed the functional impact of genomic variations within the TNF- α locus and showed that most of the variation had no impact on the expression of TNF-α or had contradictory results (Bayley et al., 2004).

TNF-β binds to the same receptor as TNF-α and a number of polymorphisms have also been described. In particular, the TNF-β 250 SNP has been the target of multiple investigations. However, LD with heat shock protein 70 alleles confounds interpretation and more advanced models will be necessary before inferences of TNF-β variability in sepsis can be made.

Both IL-1α and IL-1β engage the same receptor and are potent pro-inflammatory cytokines released by macrophages involved in the systemic inflammatory response and are capable of inducing the symptoms of septic shock and organ failure in animal models (Leon et al., 1992). Despite the finding that a homozygous TaqI genotype (−511) correlates with expression of IL-1β, no association with incidence or outcome of sepsis has been determined (Ma et al., 2002; Pociot et al., 1992).

IL-6 plays a role in lymphocyte stimulation and its levels consistently associate with severity and mortality in sepsis (van der Poll and van Deventer, 1999). Allelic variation has been described in the promoter region (-174 G/C). A German study demonstrated improved survival with the -174 G/C SNP, but not incidence of sepsis (Schluter et al., 2002). It is possible that variation at the IL-6 gene may contribute to the outcome of sepsis, but variable association of specific SNPs with IL-6 levels and the potential for LD with several alleles obscure the clinical significance of these results.

Similar studies for the identification of polymorphisms in IL-8 and MIF promoter regions and genes are ongoing.

Anti-inflammatory Cytokines

IL-10 is an integral part of the body's anti-inflammatory processes and is involved in the suppression of innate and adaptive immune responses. Its expression is stimulated by inflammation,

and it is a potent inhibitor of non-specific inflammatory response. In sepsis, high IL-10 levels have been consistently associated with disease severity (Wunder et al., 2004). This cytokine has several promoter variants with SNPs at −592, −819, and −1082. Several studies investigating the association of the promoter polymorphisms and sepsis severity, and survival found a higher frequency of the −1082 allele in patients with sepsis (particularly pneumococcal) and community-acquired pneumonia, but more well-designed studies are needed.

The IL-1 receptor antagonist (IL1RN) inhibits the pro-inflammatory actions of IL-1 by binding to their receptor. Polymorphisms in IL1RN may be promising candidates for genetic associations, but should be explored in relation to the rest of IL-1 peptides.

Coagulation

Activation of the coagulation cascade is a key event in the pathogenesis of sepsis. As a result, abnormalities in the antithrombin III, protein C, and tissue-factor inhibitor pathways have been implicated in the pathogenesis of sepsis. Recombinant human activated protein C (APC) is the only pharmacotherapy shown to be effective in the treatment of a subset of patients with severe sepsis (Bernard et al., 2001). Protein C −1641 AA genotype is associated with decreased survival, more organ dysfunction, and more systemic inflammation in patients having severe sepsis (Walley and Russell, 2007). APC not only inhibits factors Va, Xa, and PAI-1, but also neutrophil adherence, chemotaxis, and cytokine release. In animal models of septic shock genetic deficiency of tissue factor reduces mortality (Texereau et al., 2004). In contrast, genetic deficiencies of the anticoagulants thrombomodulin, antithrombin II, and protein C increase mortality. Several SNPs and other genetic polymorphisms have been described in genes of hemostatic factors, including thrombin fibrinogen, factor V, PAI-1, protein C, and endothelial protein C receptor.

Factor V and the Protein C Pathway

Factor V influences protein C activation by promoting thrombin generation. Three independent SNPs of factor V have been described that all make factor Va partially resistant to inactivation by APC, resulting in a pro-thrombotic state (Nicolaes and Dahlback, 2003). The frequency of this polymorphism in certain populations suggests that it confers some evolutionary advantage. In a large study of children with meningococcal disease, factor V Leiden heterozygosity was not only associated with increased incidence of purpura fulminans, but also a trend toward reduced mortality (Kondaveeti et al., 1999). In a substudy of PROWESS, factor V Leiden carrier status was associated with lower 28-day mortality and with less vasopressor use (Bernard et al., 2001). However, it did not determine responsiveness to recombinant human APC infusion. Biomarker studies have demonstrated, however, that plasma protein C levels do predict usefulness of recombinant human APC infusion (Shorr et al., 2006).

Plasminogen Activator Inhibitor-1 and the Fibrinolytic System

PAI-1 not only promotes clot stability, extension, and resistance to lysis, but also acts as an acute-phase reactant (Hoekstra et al., 2004). Increased PAI-1 is a risk factor for cardiovascular disease, but may confer a survival benefit in meningococcal sepsis (Kornelisse et al., 1996). A common PAI-1 promoter SNP is characterized by 4G and 5G alleles. 4G homozygosity is not only associated with an increase in PAI-1 transcription and higher rates of myocardial infarction, but also a higher risk for mortality and vascular complications among patients with sepsis (Haralambous et al., 2003; Hermans and Hazelzet, 2005; Zhan et al., 2005). Genetic polymorphisms of other constituents of the fibrin formation and degradation pathways also deserve further study.

Future Investigations

To address the complexity of the sepsis response and to predict its outcome, multiple surrogate markers that reflect the nature and severity of the inflammatory response, the status of the coagulation and fibrinolysis systems, and the magnitude of organ injury are likely to be more effective in identifying patients at risk of an adverse outcome and those who may benefit from interventional therapies. Such strategies will inherently require the use of multiplex approaches that have sufficiently high throughput to be cost effective (Kingsmore, 2006). An example of the use of multivariable models for prediction of sepsis is shown in Figure 59.2. A combination of measurements of five serum or plasma

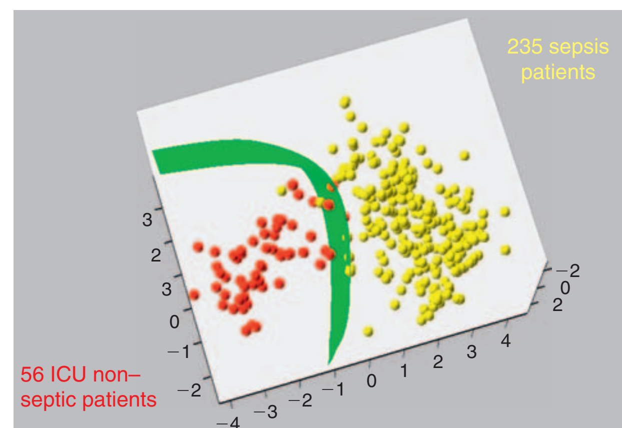


Figure 59 .2 Multiple serum biomarkers effectively distinguish patients with sepsis from ill, non-septic patients (Kingsmore et al., 2005). Individual host protein biomarkers, such as IL6, demonstrate good diagnostic sensitivity for sepsis, but poor specificity. Principal component analysis with a quadratic decision surface (green curved surface) is shown for values of five serum protein biomarkers measured in 235 patients with sepsis (yellow spheres) and 56 patients without sepsis (red spheres), but with severe illness necessitating ICU admission. Multiplexed serum biomarkers exhibited a sensitivity of 100% and specificity of 99% (positive predictive value of 96%).

cytokines showed high sensitivity and specificity in distinguishing between acutely ill, septic and non-septic patients, whereas individual cytokine measurements lacked specificity. Therefore, isolated, candidate gene-by-gene, or protein-by-protein approaches, may be rendered obsolete by these constraints.

MOLECULAR SIGNATURES AND SEPSIS

The recognition and treatment of sepsis have posed a daunting challenge for clinicians despite the improvements in technology and the principles of critical care medicine. In addition to a broad array of infectious etiologies, many of the clinical signs of overwhelming sepsis such as fever, tachycardia, and leukocytosis are non-specific and can also be seen with systemic inflammation that is induced by non-infectious causes like mesenteric ischemia-reperfusion, pulmonary embolus, pancreatitis, or trauma. Regardless, empirical broad spectrum antibiotic use is the rule as a component of early goal-directed therapy (Rivers et al., 2001).

The characterization of the human genome has provided a mechanism to study infection at the molecular level. Whereas former techniques have focused on obtaining fluid or tissue from a body compartment and incubating the sample to determine the cause of infection (if any), the study of genomics provides another means. The discovery of surface receptors that recognize microbial invasion and transmit signals to the immune system has led scientists to try to characterize the molecular and genetic changes in the host organism as it is invaded by exogenous infections. From as little as 100 ng of total cellular RNA isolated from whole blood or any tissue, the relative expression of all known genes, expressed sequence tags, and open-reading frames can be evaluated simultaneously. More importantly, the global interaction among these thousands of genes can be studied without any specific selection bias. By studying global expression, one can devise methods of class prediction based on global gene expression activity and this can be done without prior knowledge of the function of any of the genes.

Functional genomics relates these technologies to clinical medicine. Although most investigations have focused on the changes in one or relatively few genes in response to a disease or treatment, these newer technologies are exploring the changes in gene expression of the entire genome. This has resulted in functional genomics offering two unique perspectives on sepsis biology: the ability to use "patterns" of gene, protein or metabolite expression to class predict or classify tissue responses (i.e., to develop a "signature" or "fingerprint" for a specific tissue, pathogen, or outcome) and to explore the underlying biological changes that occur in health and disease while not being limited to a subset of selected genes, proteins, or metabolites.

Similarly, during the past year, a novel genetic approach – genome-wide association (GWA) – has demonstrated its potential to identify common genetic variants associated with complex diseases such as diabetes, inflammatory bowel disease, and cancer (Kingsmore et al., 2008). GWA studies seek statistically significant associations between a disease phenotype and genotypes of hundreds of thousands of common single nucleotide variants distributed throughout the genome in hundreds or thousands of affected individuals and matched controls. GWA studies are anticipated to have broad impact on drug discovery and development by providing molecular understanding of common diseases and tools for molecular stratification of patients. As yet, no GWA studies of sepsis have been published, although several are underway.

Challenges Associated with Applying Genomic Science to Sepsis

Data acquisition from microarray analyses is not problematic; however, quality control, data analysis, and information extraction remain particularly challenging. Traditional statistical and bioinformational approaches are not appropriate for the simultaneous analyses of the large number of analytes on most microarrays. In addition to using probabilities to define confidence intervals among groups, false discovery rates are often applied (Steinmetz and Davis, 2004; Storey and Tibshirani, 2003). Equally problematic are the issues of how such vast quantities of data are analyzed or presented in a manner that can be readily understood. Heat maps, dendrograms, cluster, and principal component analyses, and network analyses are generally used to extract meaningful information (Eisen et al., 1998).

As methods for accruing and analyzing data are improved, investigators can turn to other important issues to maximize the utility of genomic methodologies. The correlation between transcript abundance patterns in the peripheral blood and those within local cells at the site of infection is unclear. Clearly, the boundaries are not distinct. Furthermore, human clinical samples other than blood, cells, and tissue from the primary site of infection might be useful for genomic analysis.

Pathogen Signatures

Recognition of microbial invasion is one of the hallmarks of innate immunity. Antigen-presenting cells, as well as cells of epithelial and endothelial origin, express surface receptors that can recognize PAMPs. With the recent description of TLRs, considerable effort has been directed at examining the genome-wide expression response to different microbial pathogens. Nau et al. found that human macrophages have a profound genetic perturbation that is demonstrated following infection by any of a number of exogenous bacteria (Nau et al., 2002). The shared responses alter the baseline expression of genes that encode for other cell surface receptors, signal transduction proteins, and transcription factors. Other investigators have examined the genomic responses to exogenous microbial stimuli using varying populations of cells. Calvano et al. examined the genome-wide response of leukocytes from healthy volunteers inoculated with endotoxin. Out of 44,000 genes studied, there were 3714 genes whose expression signal changed significantly with *in vivo* stimulation (Calvano et al., 2005). This proved, on a genomic level, that the systemic perturbations of infection were more vast than clinically evident.

Even with proof of the vast alterations in gene expression that infection can induce, genomic technology has been used to discriminate among types of infections. Huang et al. not only confirmed a large shared response among dendritic cells exposed to *E. coli*, *Candida albicans*, and influenza virus, but they also showed a differential response based on the pathogen (Huang et al., 2001). Feezor et al. stimulated whole blood from human volunteers with *E. coli* LPS or heat-killed *Staphylococcus aureus* (SAC) and found that there were 758 distinct genes whose expression level differentiated between the Gram-negative and Gram-positive exposures (Feezor et al., 2003). These investigators demonstrated that the families of genes affected by different pathogenic stimuli were strikingly unique (Figure 59.3). For instance, Gram-negative infection tended to alter the expression of genes involved in global immune response, signal transduction, and plasma membrane function while Gram-positive infection altered the expression of genes that are believed to control ribosomal proteins and cell cycle regulation.

Chung et al. sought to characterize a genomic profile that would indeed distinguish between sepsis and presumably sterile systemic inflammation (Chung et al., 2006). In parallel analyses of murine and human models, they compared these two cohorts. Genomic data from spleens of septic and injured patients were used to create a septic profile, the accuracy of which was 67.1%; in the murine model, the accuracy of the genomic predictor profile was as high as 96%.

Most recently, Ramilo and colleagues have demonstrated distinctive gene expression patterns in peripheral blood leukocytes from patients with confirmed systemic bacterial (*E. coli, S. aureus, S. pneumoniae*) and viral (influenza A) infections (Ramilo et al., 2007). Furthermore, distinctive gene expression patterns were observed in patients with respiratory infections of different etiologies demonstrating the utility of blood in the evaluation of infections of different organ systems. Similar results were documented in a cohort of military trainees with febrile respiratory illness (Thach et al., 2005).

These studies are the first to show that molecular characterization can be performed to distinguish between the infected and non-infected states, even in the face of very similar clinical presentations. These data also serve to highlight the grossly insensitive criteria used in clinical medicine to determine infection or sepsis, and the power of genomic science to provide a molecular characterization of a clinical dilemma. On a macrosystem level, the whole body response to either stimuli has been well established: fever, leukocytosis, tachycardia, with the potential of devolving into multi-system organ failure. Functional genomics is a better test of the cause of the patient's global dysfunction and may one day be used to guide initial therapy or to tailor ongoing therapy.

Progression Signatures

The need for rapid, accurate identification of disease progression in sepsis has increased dramatically with the upcoming availability of several novel treatment regimens. While novel sepsis therapies are improving sepsis outcomes, they are creating new patient management and diagnostic challenges for physicians. In 2001, the Food and Drug Administration (FDA) approved APC (Xigris, Eli Lilly and Company, Indianapolis, IN) for treatment of patients with severe sepsis (SS) and Acute Physiology and Chronic Health Evaluation II score of ≥25 — one of the first examples of inclusion of a biomarker in the indication for a therapy. In the pivotal phase III trial of APC in SS (PROWESS), 28-day mortality was decreased by 6% (Bernard et al., 2001). The greatest reduction in mortality (13%) and cost effectiveness was observed in the most seriously ill patients (those with APACHE II score ≥25). APC benefit was also observed in patients with pneumonia. In contrast, APC exhibited modest survival benefit and cost-ineffectiveness in patients with APACHE II score <25. However, APC therapy is associated with a 1–2% incidence of major bleeding. For these reasons, appropriate use of APC is most likely to occur following development of an objective, accurate, rapid diagnostic test for progression to severe sepsis. Diminution of protein C activity in citrated plasma was recently shown to be a surrogate marker for APC efficacy in patients with severe sepsis (Shorr et al., 2006).

Prompt and accurate diagnosis is especially important for effective sepsis management and contributes significantly to positive outcomes (Rivers et al., 2001). Rates of progression of severe sepsis to organ failure, septic shock and death heterogeneous are largely independent of the specific underlying infectious disease process. For example, case-fatality rates in those with culture-negative severe sepsis are similar to those with positive cultures (Rangel-Frausto et al., 1995). Case-fatality rates in sepsis are, however, critically dependent upon disease staging. Current differentiation of SIRS, sepsis, severe sepsis, and septic shock relies exclusively on clinical assessment (Balk, 2000). Thus, a key unmet diagnostic need is for rapid, quantitative, objective determination of the stage of sepsis development and likelihood of progression to severe.

Several currently available clinical indices do provide quantitative assessment of staging and severity of sepsis, but with limitations. They include sequential organ failure assessment (SOFA) score, APACHE II score, and blood lactate level (Pacelli et al., 1996). APACHE II score, although quantitative, is largely subjective, complex, cumbersome, and has a relatively narrow dynamic range. Blood lactate levels are quantitative, but limited by false negative normal values in elderly patients and by confounding comorbidities such as liver failure, diabetes mellitus, and the use of certain medications. The clinical picture in sepsis patients is highly dynamic, and assessment of such indices tends to be subjective and to occur with insufficient frequency. Furthermore, sepsis is very heterogeneous in terms of pathogen, source of infection, associated comorbidity, course and complications. Delayed or errant diagnosis of severe sepsis may result in failure of timely treatment. Thus, an early, rapid, objective diagnostic of severe sepsis would be a significant adjunct to these clinical indices and would significantly advance patient management.

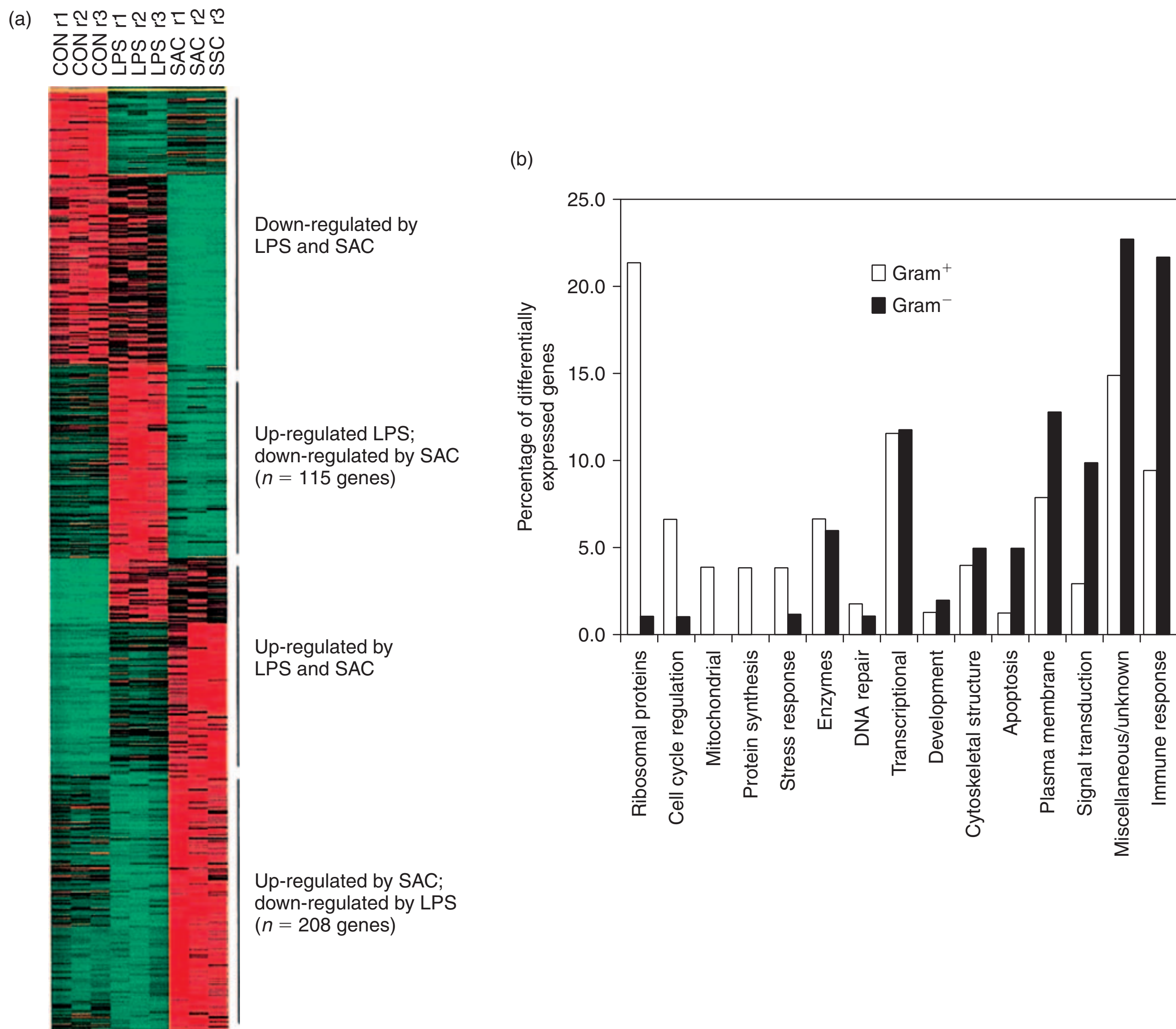


Figure 59.3 Heat map and gene ontologies of leukocytes when ex vivo whole blood is stimulated with either LPS or heat-killed *S. aureus*. (a) K-means cluster analysis on 780 genes whose expression significantly changed in response to ex vivo stimulation. Patterns of gene expression could be classified into bins on the basis of the similar and disparate responses to Gram-negative and Gram-positive pathogens. CON, control. (b) Differences in gene expression between the two stimuli based on the ontologies for the 780 genes. Gram-positive (Gram+) stimulus (heat-killed *S. aureus*) preferentially altered the expression of ribosomal and mitochondrial proteins and cell cycle proteins, whereas the Gram-negative (Gram) stimulus altered the expression of genes involved in signal transduction and the immune response. Figure is from Feezor et al. (2003).

Elevations of IL-6, IL-8, IGFBP1, IL-2sR and MIF in sepsis have been widely documented. Furthermore, levels of IL-6, IL-8 and IL-2sR have variously shown correlation with sepsis severity, progression to septic shock, organ failure, and mortality. However, many other conditions elevate blood levels of these analytes, and diagnostic sensitivity and specificity of individual analyte levels has been insufficient for diagnosis of sepsis or severe sepsis (Delogu et al., 1995; Martin et al., 1994; Selberg et al., 2000).

Using an antibody microarray on specimens from patients from the PROWESS study (Bernard et al., 2001; Perlee et al., 2004), including sepsis non-survivors, 133 analytes were measured in 139 matched serum and plasma samples (between day 0 and 28) from 12 severe sepsis patients (7 survivors and 5 who died) and 8 normal individuals. Sixty-three candidate sepsis biomarkers were identified (35 in serum and 48 in plasma); each exhibited greater than twofold change between severe sepsis patients and controls on day zero. The top biomarker was IL-6

(26-fold increase) (Kingsmore et al., 2005). Multivariable models were more effective than single analytes in prediction of sepsis (Figure 59.2). A combination of measurements of five serum or plasma cytokines showed high sensitivity and specificity in distinguishing between acutely ill, septic, and non-septic patients, whereas individual cytokine measurements lacked specificity.

In an additional study, biomarkers predictive of sepsis death were identified by at least a twofold difference in average level between sepsis survivors and non-survivors in the days immediately preceding death (days 0–4). A total of 10 candidate sepsis mortality biomarkers were identified. The top biomarker was follistatin (15-fold difference).

DNA microarray-based gene transcript profiling of the responses of primates to infection has begun to yield new insights into host–pathogen interactions; however, this approach remains plagued by challenges and complexities that have yet to be adequately addressed. The rapidly changing nature over time of acute infectious diseases in a host, and the genetic diversity of microbial pathogens present unique problems for the design and interpretation of functional-genomic studies in this field. In addition, there are more common problems related to heterogeneity within clinical samples, the complex non-standardized confounding variables associated with human subjects and the complexities posed by the analysis and validation of highly parallel data. Various approaches have been developed to address each of these issues, but there are still significant limitations that need to be overcome. The resolution of these problems should lead to a better understanding of the dialog between the host and pathogen.

2009 UPDATE

As noted throughout this chapter, the genome revolution holds great promise for both scientific and therapeutic advances related to sepsis. Although we anxiously await genomic breakthroughs in the ability to identify patients at high risk for certain infections, to stratify those at risk for progression, or to rapidly determine etiology, investigators are still early in developing the cohorts and technological innovations needed for validation of initial findings. Regardless, a number of tempting observations have been recently reported.

Due to its heterogeneity, the absence of a clear genetic "cause" of sepsis is not surprising. Though we continually add to our nuanced understanding of the impact of a growing array of SNPs, the genetic risk is likely polygenic with a strong environmental component affecting likelihood of exposure to certain pathogens. Among the best-understood pattern recognition pathway genes, the Toll-like receptors remain the focus of attention. The previously described association with sepsis susceptibility for polymorphisms in innate immunity genes was further evaluated in a large cohort ($n = 774$) intensive care unit patients. The study analyzed genes involved in infection, inflammation, and coagulation. The combination of at least one variant in both *NOD2* (which synergizes with *TLR2*) and *TLR4* that occurs in 5% of the population was associated with bacteremia. However, the variants did not affect the type of infection (e.g., gram negative or positive) or the source of infection (pulmonary, catheter, etc.). Furthermore, the presence of variants of *MASP2* gene (9% of population) or the *NOD2/TLR4* combination both predispose to death (Henckaerts et al., 2009). In addition to the longstanding interest in *NOD2* and *TLR4*, *TLR9* has recently been described as having a dual role as a potent proinflammatory stimulator as well as bactericidal activity related to bacterial clearance (Sjolinder et al., 2008). Continued advances in the analysis of gene expression profiling via microarrays and proteomics not only reveal how pathogens affect the host's gene expression patterns but also a wealth of candidate genes for genetic susceptibility studies. To more accurately identify novel genes with different clinical outcomes, genome wide expression analyses are now being combined with genetic studies (Thuong et al., 2008).

Gene expression changes can indicate both general infection and disease-specific responses. Investigators are increasingly interested in the dynamics of gene expression over the course of infection, beginning with the presymptomatic period in high risk patients, through recovery or death. McDunn and colleagues recently defined the concept of the "Riboleukogram" (McDunn et al., 2008). The authors focused on ventilator associated pneumonia (VAP) and, in a mouse model, were able to develop RNA profiles capable of distinguishing the severity of the insult and the type of infectious agent. Using principal components analysis on gene expression data from a group of mechanically ventilated patients, they examined the behavior of the human orthologs to the genes identified in the mouse study and although there was a conserved peripheral leukocyte transcriptional response, the information did not appear to be more useful than current clinical criteria. However, by explicitly accounting for variance over time, a set of 85 genes could identify a subset of patients whose expression levels changed before the clinical diagnosis of VAP. Although numerous mouse studies suggest distinct signatures for different bacterial pathogens, a definitive clinical study demonstrating that capacity in humans remains elusive (Tang et al., 2008). The fact that these studies are typically small and use different analytic techniques highlights the importance of study design and methods.

Transcriptional profiles may also be expected to aid prognosis and response to treatment as is now the case in oncology. However, most studies in sepsis and infectious diseases have just begun to classify patients with clinical phenotypes that may benefit from genomic directed therapy.

2009 UPDATE REFERENCES

Henckaerts, L., Nielsen, K.R., Steffensen, R. et al. (2009). Polymorphisms in innate immunity genes predispose to bacteremia and death in the medical intensive care unit. *Crit Care Med* 37, 192–201.

McDunn, J.E., Husain, K.D., Polpitiya, A.D. et al. (2008). Plasticity of the systemic inflammatory response to acute infection during critical illness: Development of the riboleukogram. *PLoS ONE* 3(2), e1564.

Sjolinder, H., Mogensen, T.H., Kilian, M., Jonsson, A.B. and Paludan, S.R. (2008). Important role for Toll-like receptor 9 in host defense against mengococccal sepsis. *Infect Immun* 76, 54211–5428.

Tang, B.M.P., McLean, A.S., Dawes, I.W. et al. (2008). Gene-expression profiling of gram-positive and gram-negative sepsis in critically ill patients. *Crit Care Med* 36, 1125–1128.

Thuong, N.T., Dunstan, S.J., Chau, T.T. et al. (2008). Identification of tuberculosis susceptibility genes with human macrophage gene expression profiles. *PLoS Pathog* 4, 1–13.

REFERENCES

Agnese, D.M., Calvano, J.E., Hamh, S.J. et al. (2002). Human toll-like receptor 4 mutations but not CD14 polymorphisms are associated with increased risk of gram-negative infections. *J Infect Dis* 186, 1522–1525.

Akira, S., Takedai, K. and Kaisho, T. (2001). Toll-like receptors: Critical proteins liniking innate and acquired immunity. *Nat Review Immunol* 2, 675–680.

Angus, D.C. and Wax, R.S. (2001). Epidemiology of sepsis: an update. *Crit Care Med* 29, S109–S116.

Arbour, N.C., Lorenz, E., Schutte, B.C. et al. (2000). TLR4 mutations are associated with endotoxin hyporesponsiveness in humans. *Nat Genet* 25, 187–191.

Arcaroli, J., Fessler, M.B. and Abraham, E. (2005). Genetic polymorphisms and sepsis. *Shock* 24, 300–312.

Balk, R.A. (2000). Severe sepsis and septic shock. Definitions, epidemiology, and clinical manifestations. *Crit Care Clin* 16, 179–192.

Bayley, J.P., Ottenhoff, T.H. and Verweij, C.L. (2004). Is there a future for TNF promoter polymorphisms? *Genes Immun* 5, 315–329.

Bellamy, R. and Hill, A.V. (1998). Genetic susceptibility to mycobacteria and other infectious pathogens in humans. *Curr Opinion Immunol* 10, 483–487.

Bernard, G.R., Vincent, J.L., Laterre, P.F. et al. (2001). Efficacy and safety of recombinant human activated protein C for severe sepsis. *N Engl J Med* 344, 699–709.

Beutler, B. and Reitschel, E.T. (2003). Innate immune sensing and its roots: the story of endotoxin. *Nat Review Immunol* 3, 169–176.

Bone, R.C., Balk, R.A., Cerra, F.B. et al. (1992). American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference: Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. *Chest* 01, 1644–1655.

Bone, R.C., Fisher, C.J., Jr., Clemmer, T.P. et al. (1989). Sepsis syndrome: A valid clinical entity. *Crit Care Med* 17, 389–393.

Botstein, D. and Risch, N. (2003). Discovering genotypes underlying human phenotypes: Past successes for Mendelian disease, future approaches for complex disease. *Nat Genet* 33, 228–237.

Burgner, D. and Levin, M. (2003). Genetic susceptibility to infectious disease. *Pediatr Infect Dis J* 22, 1–6.

Calvano, J.E., Bowers, D.J., Coyle, S.M. et al. (2006). Response to systemic endotoxemia among humans bearing polymorphisms of the toll-like receptor 4 (TLR4). *Clin Immunol* 121, 186–190.

Calvano, S.E., Xiao, W., Richards, D.R. et al. (2005). A network-based analysis of systemic inflammation in humans. *Nature* 437, 1032–1037.

Cardon, L.R. and Bell, J.I. (2001). Association study designs for complex diseases. *Nat Rev Genet* 2, 91–99.

Cardon, L.R. and Palmer, L.J. (2003). Population stratification and spurious allelic association. *Lancet* 361, 598–604.

Child, N.J., Yang, J.A., Pulletz, M.C. et al. (2003). Polymorphisms in Toll-like receptor 4 and the systemic inflammatory response syndrome. *Biochem Soc Trans* 31, 652–653.

Choi, E.H., Zimmerman, P.A., Foster, L.B. et al. (2001). Genetic polymorphisms in molecules of innate immunity and susceptibility to infection with Wucheria bancrofti in South India. *Genes Immun* 2, 248–253.

Chung, T.P., Laramie, J.M., Meyer, D.J., Downey, T., Tam, L.H., Ding, H., Buchman, T.G., Karl, I., Stormo, G.D, Hotchkiss, R.S. et al. (2006). Molecular diagnostics in sepsis: from bedside to bench. *J Am Coll Surg* 203, 585–598.

Creagh, E.M. and O'Neill, L.A. (2006). TLRs, NLR and RLRs: A trinity of pathogen sensors that co-operate in innate immunity. *Trends Immunol* 27, 352–357.

Creagh, E.M. et al. (2006). *Trends Immunol*.

Croker, B., Crozat, K., Berger, M., Xia, Y., Sovath, S., Schaffer, L., Eleftherianos, I., Imler, J.L. and Beutler, B. (2007). ATP-sensitive-potassium channels mediate survival during infection in mammals and insects. *Nat Genet* 39, 1453–1460.

D'Avila, L.C., Albarus, M.H., Franco, C.R., Aguiar, B.B., Oliveira, J.R., Dias, F.S. and Alho, C.S. (2006). Effect of CD14 -260C > T polymorphism on the mortality of critically ill patients. *Immunol Cell Biol* 84, 342–348.

Delogu, G., Casula, M.A., Mancini, P., Tellan, G. and Signore, L. (1995). Serum neopterin and soluble interleukin-2 receptor for prediction of a shock state in gram-negative sepsis. *J Crit Care* 10, 64–71.

Densen, P., Weiler, J.M., Griffiss, J.M. and Hoffmann, L.G. (1987). Familial properdin deficiency and fatal meningococcemia: correction of the bactericidal defect by vaccination. *New Eng J Med* 316, 922–926.

Drennan, M.B., Nicolle, D., Quesniaux, V.J., Jacobs, M., Allie, N., Mpagi, J., Fremond, C., Wagner, H., Kirschning, C. and Ryffel, B. (2004). Toll-Like Receptor 2-Deficient Mice Succumb to Mycobacterium tuberculosis Infection. *Am J Pathol* 164, 49–57.

Eisen, M.B., Spellman, P.T., Brown, P.O. and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95(14), 863–868.

Feezor, R.J., Oberholzer, C., Baker, H.V. et al. (2003). Molecular characterization of the acute inflammatory response to infections with gram-negative versus gram-positive bacteria. *Infect Immun* 71, 5803–5813.

Freimer, N. and Sabatti, C. (2004). The use of pedigree, sib-pair and association studies of common diseases for genetic mapping and epidemiology. *Nat Genet* 36, 1045–1051.

Gibot, S., Cariou, A., Drouet, L., Rossignol, M. and Ripoll, L. (2002). Association between a genomic polymorphism within the CD14 locus and septic shock susceptibility and mortality rate. *Crit Care Med* 30, 969–973.

Goring, H.H., Terwilliger, J.D. and Blangero, J. (2001). Large upward bias in estimation of locus-specific effects from genomewide scans. *Am J Hum Genet* 69, 1357–1369.

Haralambous, E., Hibberd, M.L., Hermans, P.W., Ninis, N., Nadel, S. and Levin, M. (2003). Role of functional PAI-1 4G/5G promoter polymorphism in susceptibility, severity, and outcome of meningococcal disease in Caucasian children. *Crit Care Med* 31, 2788–2793.

Härtel, C., Schultz, C., Herting, E. and Göpel, W. (2007). Genetic association studies in VLBW infants exemplifying susceptibility to sepsis – recent findings and implications for future research. *Acta Paediatr* 96, 158–165.

Hawn, T.R., Verbon, A., Lettinga, K.D. et al. (2003). A common dominant TLR5 stop codon polymorphism abolishes flagellin signaling and is associated with susceptibility to Legionnaires' disease. *J Exp Med* 198, 1563–1572.

Hayashi, F., Smith, K.D., Ozinsky, A. et al. (2001). The innate immune response to bacterial flagellin is mediated by toll-like receptor 5. *Nature* 410, 1099–1103.

Hermans, P.W. and Hazelzet, J.A. (2005). Plasminogen activator inhibitor type 1 gene polymorphism and sepsis. *Clin Infect Dis* 41(Suppl 7), S453–S458.

Hill, A.V. (2006). Aspects of genetic susceptibility to human infectious diseases. *Annual Rev Genet* 40, 469–486.

Hirschhorn, J.N., Lohmueller, K., Byrne, E. and Hirschhorn, K. (2002). A comprehensive review of genetic association studies. *Genet Med* 4, 45–61.

Hoekstra, T., Geleijnse, J.M., Schouten, E.G. and Kluft, C. (2004). Plasminogen activator inhibitor-type 1: Its plasma determinants and relation with cardiovascular risk. *Thromb Haemost* 91, 861–872.

Huang, Q., Liu, D., Majewski, P. et al. (2001). The plasticity of dendritic cell responses to pathogens and their components. *Science* 294, 870–875.

Hughes, A.L. and Nei, M. (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335, 167–170.

Inohara, N. and Nunez, G. (2003). NODS: intracellular proteins involved in inflammation and apoptosis. *Nat Rev Immuno* 3, 371–382.

Ioannidis, J.P., Ntzani, E.E., Trikalinos, T.A. and Contopoulos-Ioannidis, D.G. (2001). Replication validity of genetic association studies. *Nat Genet* 29, 306–309.

Jepson, A.P., Banya, W.A., Sisay-Joof, F., Hassan-King, M., Bennett, S. and Whittle, H.C. (1995). Genetic regulation of fever in Plasmodium falciparum malaria in Gambian twin children. *J Infect Dis* 172, 316–319.

Kato, H., Takeuchi, O., Sato, S. et al. (2006). Differential roles of MDA5 and RIG-1 helicases in the recognition of RNA viruses. *Nature* 441, 101–105.

Kingsmore, S.F., Lindquist, I.E., Mudge, J., Gessler, D.D. and Beavis, W.D. (2008). Genome-Wide Association Studies: progress and potential for drug discovery and development. *Nat Rev Drug Discov* 7, 221–230.

Kingsmore, S.F. (2006). Multiplexed protein measurement: technologies and applications of protein and antibody arrays. *Nat Rev Drug Discov* 5, 310–320.

Kingsmore, S.F., Lejnine, S.J., Driscoll, M. and Tchernev, V.T. (2005). Biomarkers for sepsis. US Patent Application 20050196817.

Kondaveeti, S., Hibberd, M.L., Booy, Nadel, S. and Levin, M. (1999). Effect of the Factor V Leiden mutation on the severity of meningococcal disease. *Pediatr Infect Dis J* 18, 893–896.

Kornelisse, R.F., Hazelzet, J.A., Savelkoul, H.F. et al. (1996). The relationship between PAI-1 and proimflammatory and counterinflammatory mediators in children with meningococcal septic shock. *J Infect Dis* 173, 1148–1156.

Kronborg, G. and Garred, P. (2002). Mannose-binding lectin genotype as a risk factor for invasive pneumococcal infection. *Lancet* 360, 1176.

Kuhlman, M., Joiner, K. and Ezekowitz, R.A. (1989). The human mannose-binding protein functions as an opsonin. *J Exp Med* 169, 1722–1745.

Lander, E. and Kruglyak, L. (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 11, 241–247.

Lasker, M.V., Gajjar, M.M. and Nair, S.K. (2005). Cutting edge: Molecular structure of the IL-1R-associated kinase-4 death domain and its implications for TLR signaling. *J Immunol* 175, 4175–4179.

Leon, P., Redmond, H.P., Shou, J. and Daly, J.M. (1992). Interleukin 1 and its relationship to endotoxin tolerance. *Arch Surg* 127, 146–151.

Lin, M.T. and Albertson, T.E. (2004). Genomic polymorphisms in sepsis. *Crit Care Med* 32, 569–579.

Liu, R. et al. (1996). Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* 86, 367–377.

Lohmueller, K.E., Pearce, C.L., Pike, M., Lander, E.S. and Hirschhorn, J.N. (2003). Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 33, 177–182.

Lorenz, E., Mira, J.P., Cornish, K.L., Arbour, N.C. and Schwartz, D.A. (2000). A novel polymorphism in the toll-like receptor 2 gene and its potential association with staphylococcal infection. *Infect Immun* 68, 6398–6401.

Lorenz, E., Mira, J.P., Frees, K.L. and Schwartz, D.A. (2002). Relevance of mutations in the TLR4 receptor in patients with gram-negative septic shock. *Arch Intern Med* 162, 1028–1032.

Lorenz, E. (2006). TLR 2 and TLR 4 expression during bacterial infections. *Curr Pharm Des* 12, 4185–4193.

Ma, P., Chen, D., Pan, J. and Du, B. (2002). Genomic polymorphism within interleukin-1 family cytokines influences the outcome of septic patients. *Crit Care Med* 30, 1046–1050.

Majetschak, M., Obertacke, U., Schade, F.U. et al. (2002). Tumor Necrosis Factor gene polymorphism, leukocyte function, and sepsis susceptibility in blunt trauma patients. *Clin Diagn Lab Immunol* 9, 1205–1211.

Martin, G.S., Mannino, D.M., Eaton, S. and Moss, M. (2003). The epidemiology of sepsis in the United States from 1979 through 2000. *N Engl J Med* 348, 1546–1554.

Martin, C., Saux, P., Mege, J.L., Perrin, G., Papazian, L. and Gouin, F. (1994). Prognostic values of serum cytokines in septic shock. *Inten Care Med* 20, 272–277.

Martin, G.S., Mannino, D.M., Eaton, S., Moss, M. (2003). The epidemiology of sepsis in the United States from 1979 through 2000. *N Engl J Med* 348(16), 1546–1554.

Mira, J.P., Cariou, A., Crell, F. et al. (1999). Association of TNF2, TNFalpha promoter polymorphisms with septic shock susceptibility and mortality: a multicenter study. *JAMA* 282, 561–568.

Mira, J.P., Cariou, A., Grall, F., Delclaux, C., Losser, M.R., Heshmati, F., Cheval, C., Monchi, M., Teboul, J.L., Riché, F. et al. (1999). Association of TNF2, a TNF-alpha promoter polymorphism, with septic shock susceptibility and mortality: A multicenter study. *JAMA* 282, 561–568.

Moore, C.E., Segal, S., Berendt, A.R., Hill, A.V. and Day, N.P. (2004). Lack of association between Toll-like receptor 2 polymorphisms and susceptibility to severe disease caused by Staphylococcus aureus. *Clin Diagn Lab Immunol* 11, 1194–1197.

Nau, G.J., Richmond, J.F., Schlesinger, A., Jennings, E.G., Lander, E.S. and Young, R.A. (2002). Human macrophage activation programs induced by bacterial pathogens. *Proc Natl Acad Sci USA* 99, 1503–1508.

Nicolaes, G.A. and Dahlback, B. (2003). Acitvated protein C resistance [FV(Leiden)] and thrombosis: factor Va mutations causing hypercoagulable states. *Hematol Oncol Clin North Am* 17, 37–61.

Opal, S.M. and Cohen, J. (1999). Clinical gram-positive sepsis: does it fundamentally differ from gram-negative bacterial sepsis? *Crit Care Med* 27, 1608–1616.

Oureshi, S.T., Lariviere, L., Leveque, G. et al. (1999). Endotoxin-tolerant mice have mutations in toll-like receptor 4 (TLR4). *J Exp Med* 189, 615–625.

Pacelli, F., Doglietto, G.B., Alfieri, S. et al. (1996). Prognosis in intra-abdominal infections. Multivariate analysis on 604 patients. *Arch Surg* 131, 641–645.

Parrillo, J.E., Parker, M.M., Natanson, C. et al. (1990). Septic shock in humans. Advances in the understanding of pathogenesis, cardiovascular dysfunction, and therapy. *Ann Intern Med* 113, 227–242.

Perlee, L., Christiansen, J., Dondero, R., Grimwade, B., Lejnine, S., Mullenix, M., Shao, W., Sorette, M., Tchernev, V., Patel, D. et al. (2004). Development and standardization of multiplexed antibody microarrays for use in quantitative proteomics. *Proteome Sci* 15, 2–9.

Pociot, F., Molvig, J., Wogensen, L., Worsaae, H. and Nerup, J. (1992). A TaqI polymorphism in the human interleukin-1B gene correlates with IL-1β secretion in vitro. *Eur J Clin Invest* 22, 396–402.

Ramilo, O., Allman, W., Chung, W. et al. (2007). Gene expression patterns in blood leukocytes discriminate patients with acute infections. *Blood* 109, 2066–2077.

Rangel-Frausto, M.S., Pittet, D., Costigan, M., Hwang, T., Davis, C.S. and Wenzel, R.P. (1995). The natural history of the systemic inflammatory response syndrome (SIRS). A prospective study. *JAMA* 1, 117–123.

Read, R.C., Pullin, J., Gregory, S. et al. (2001). A functional polymorphism of toll-like receptor 4 is not associated with likelihood or severity of meningococcal disease. *J Infect Dis* 484, 640–642.

Redden, D.T. and Allison, D.B. (2003). Nonreplication in genetic association studies of obesity and diabetes research. *J Nutr* 133, 3323–3326.

Risch, N.J. (2000). Searching for genetic determinants in the new millennium. *Nature* 405, 847–856.

Rivers, E., Nguyen, B., Havstad, S., Ressler, J., Muzzin, A., Knoblich, B., Peterson, E. and Tomlanovich, M. (2001). Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N Engl J Med* 345(19), 1368–1377.

Roy, S., Knox, K. and Segal, S. (2002). MBL genotype and risk of invasive pneumococcal disease: A case−control study. *Lancet* 359, 1569–1573.

Sackton, T.B., Lazzaro, B.P., Schlenke, T.A., Evans, J.D., Hultmark, D. and Clark, A.G. (2007). Dynamic evolution of the innate immune system in Drosophila. *Nat Genet* 39, 1461–1468.

Samson, M. et al. (1996). Resistance to HIV-1 infection in Caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* 382, 722–725.

Sands, K.E., Bates, D.W., Lanken, P.N. et al. (1997). Epidemiology of sepsis syndrome in 8 academic medical centers. *JAMA* 278, 234–240.

Schlenke, T.A. and Begun, D.J. (2003). Natural selection drives Drosophila immune system evolution. *Genetics* 164, 1471–1480.

Schluter, B., Raufhake, C., Erren, M. et al. (2002). Effect of the IL-6 promoter polymorphism (-174 G/C) on the incidence and outcome of sepsis. *Crit Care Med* 30, 32–37.

Selberg, O., Hecker, H., Martin, M., Klos, A., Bautsch, W. and Kohl, J. (2000). Discrimination of sepsis and systemic inflammatory response syndrome by determination of circulating plasma concentrations of procalcitonin, protein complement 3a, and interleukin-6. *Crit Care Med* 28, 2793–2798.

Shorr, A.F., Bernard, G.R., Dhainaut, J.F., Russell, J.R., Macias, W.L., Nelson, D.R. and Sundin, D.P. (2006). Protein C concentrations in severe sepsis: an early directional change in plasma levels predicts outcome. *Crit Care* 10, R92.

Sillanpaa, M.J. and Auranen, K. (2004). Replication in genetic studies of complex traits. *Ann Hum Genet* 68, 646–657.

Skamene, E., Schurr, E. and Gros, P. (1998). Infection genomics: Nramp1 as a major determinant of natural resistance to intracellular infections. *Annu Rev Med* 49, 275–287.

Smirnova, I., Mann, N., Dols, A. et al. (2003). Assay of locus specific genetic load implicates rare toll-like receptor 4 mutations in meningococcal susceptibility. *Proc Natl Acad Sci USA* 100, 6075–6080.

Smyth, D.J., Cooper, J.D., Bailey, R. et al. (2006). A genome-wide association study of nonsynonomous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. *Nat Genet* 38, 617–619.

Steinmetz, L.M. and Davis, R.W. (2004). Maximizing the potential of functional genomics. *Nat Rev Genet* 5, 190–201.

Storey, J.D. and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 100, 9440–9445.

Stüber, F., Petersen, M., Bokelmann, F. and Schade, U. (1996). A genomic polymorphism within the tumor necrosis factor locus influences plasma tumor necrosis factor-alpha concentrations and outcome of patients with severe sepsis. *Crit Care Med* 24, 381–384.

Sutherland, A.M., Walley, K.R. and Russell, J.A. (2005). Polymorphisms in CD14, mannose-binding lectin, and Toll-like receptor-2 are associated with increased prevalence of infection in critically ill adults. *Crit Care Med* 33, 638–644.

Texereau, J., Pene, F., Chiche, J.D., Rousseau, C. and Mira, J.P. (2004). Importance of hemostatic gene polymorphisms for susceptibility to and outcome of severe sepsis. *Crit Care Med* 32(5 Suppl), S313–S319.

Thach, D.C., Agan, B.K., Olsen, C. et al. (2005). Surveillance of transcriptomes in basic military trainees with normal, febrile respiratory illness, and convalescent phenotypes. *Genes and Immunity* 6, 588–595.

van der Poll, T. and van Deventer, S.J. (1999). Cytokines and anticytokines in the pathogenesis of sepsis. *Infect Dis Clin North Am* 13, 413–426.

Von Bernuth, H., Puel, A., Ku, C.L. et al. (2005). Septicemia without sepsis: inherited disorders of nuclear factor kappa B-mediated inflammation. *Clin Infect Dis* 41, S436–S439.

Walley, K.R. and Russell, J.A. (2007). Protein C-1641 AA is associated with decreased survival and more organ dysfunction in severe sepsis. *Crit Care Med* 35, 12–17.

Walsh, M.C., Bourcier, T., Takahashi, K. et al. (2005). Mannose-binding lectin is a regulator of inflammation that accompanies myocardial ischemia and reperfusion injury. *J Immunol* 175, 541–546.

Wunder, C., Eichelbronner, O. and Roewer, N. (2004). Are IL-6, IL-10, and PCT plasma concentrations reliable for outcome prediction in severe sepsis? A comparison with APACHE III and SAPS II. *Inflamm Res* 53, 158–163.

Zhan, Z.Y., Wang, H.W., Chen, D.F., Cheng, B.L., Wang, H.H. and Fang, X.M. (2005). Relationship between sepsis and 4G/5G polymorphism within the promoter region of plasminogen activator inhibitor-1 gene. *Zhonghua Yi Xue Za Zhi* 85, 2404–2407.

CHAPTER

Genomics and the Management of Hepatitis

N.A. Shackel, K. Patel and J. McHutchison

INTRODUCTION

Viral hepatitis is a significant global health problem with hepatitis B (HBV) and hepatitis C (HCV) infecting in excess of 300 million people. HBV and HCV are complicated by chronic persistent infection characterized in a proportion of patients by progressive hepatic injury leading to complications of end-stage liver disease including hepatocellular carcinoma (HCC). HCC is the fifth most prevalent human malignancy, and the majority of cases can be directly attributed to liver injury secondary to chronic HBV and/or HCV infection. Although both HAV and HEV are significant health problems they are typically characterized by a self-limiting course and are not complicated by significant clinical sequelae in the majority of cases. Therefore, research into infectious hepatitis has focused mainly on HBV and HCV pathogenesis, including the development of liver fibrosis, the immune response in acute infection, mechanisms of viral persistence and the development of HCC. The use of functional genomics approaches has significantly advanced our understanding of viral hepatitis pathogenesis and as well as our understanding of therapeutic strategies.

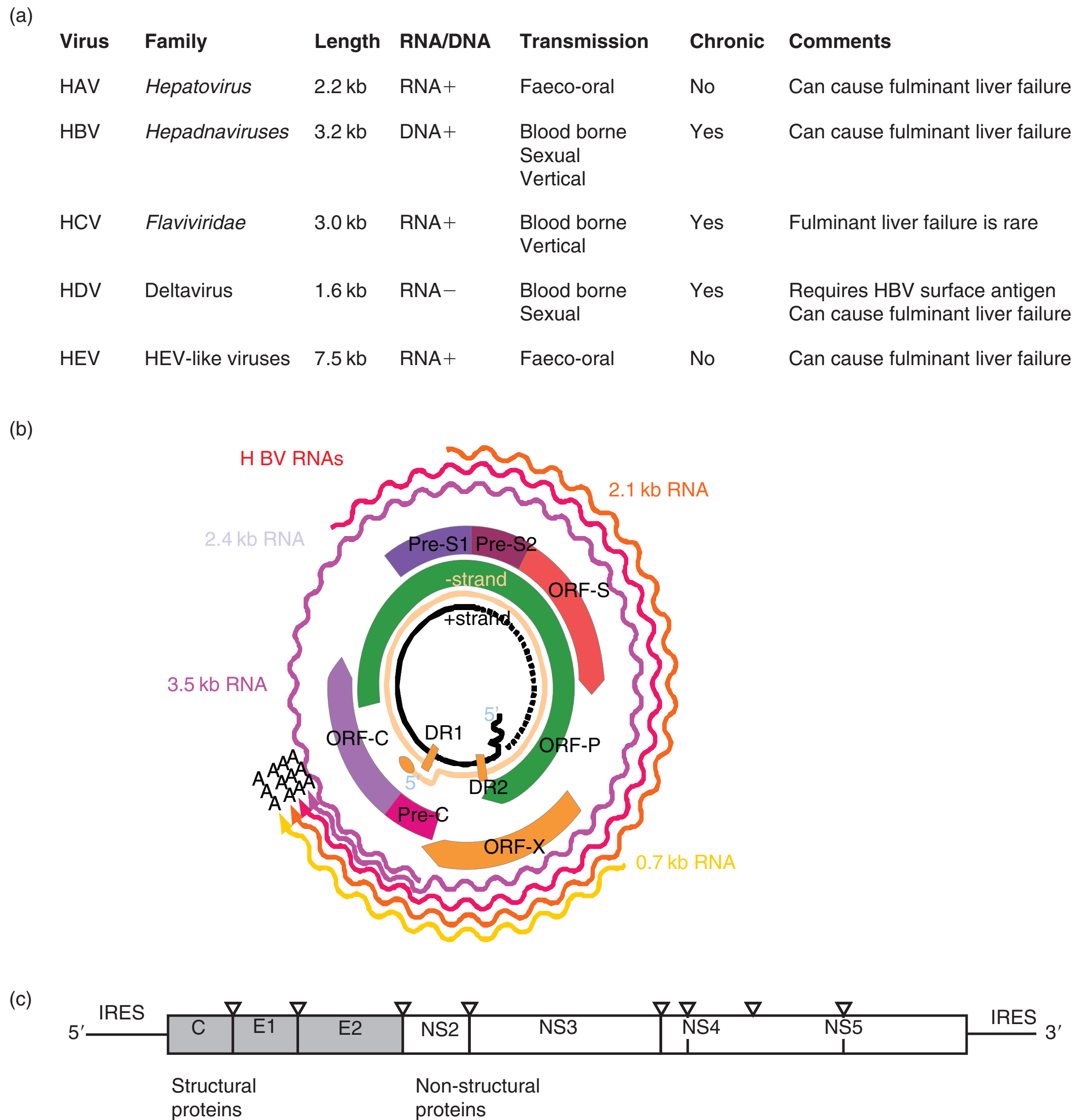
VIROLOGY OF HEPATITIS VIRUSES

The hepatitis viruses are characterized by specificity for the liver and in particular the hepatocyte (Figure 60.1). However, the mechanism by which these viruses are specific for the liver is largely unknown but is thought to involve hepatocyte receptor and co-receptor interactions and possible involvement of liver-specific pathways such as lipoprotein trafficking and synthesis.

Hepatitis A (HAV) is an RNA virus of the *Hepatovirus* genus belonging to *Picornaviridae* family (Flehmig, 1990; Lee, 2003; Martin and Lemon, 2006). HAV is a positive strand RNA virus that has a 7.5 kb genome that is translated into a 2225 to 2227 polyprotein that gives rise to a number of structural and non-structural proteins. The viral particles are 27–32 nm in diameter and there are distinct genotypes and sub-genotypes of HAV. In contrast to another hepatotrophic RNA virus, HCV, HAV has a high degree of genomic and resultant antigenic conservation. The spontaneous mutation rate of HAV is low and antibody-mediated immunity from previous exposure or vaccination is effective in preventing HAV infection. HAV is characterized by slow replication, and it is rarely cytopathic and is stable in the environment for at least a month.

Essentials of Genomic and Personalized Medicine
by Ginsburg & Willard

Copyright © 2010, Elsevier Inc.
All rights reserved.


(a)

Virus	Family	Length	RNA/DNA	Transmission	Chronic	Comments
HAV	*Hepatovirus*	2.2 kb	RNA+	Faeco-oral	No	Can cause fulminant liver failure
HBV	*Hepadnaviruses*	3.2 kb	DNA+	Blood borne Sexual Vertical	Yes	Can cause fulminant liver failure
HCV	*Flaviviridae*	3.0 kb	RNA+	Blood borne Vertical	Yes	Fulminant liver failure is rare
HDV	Deltavirus	1.6 kb	RNA−	Blood borne Sexual	Yes	Requires HBV surface antigen Can cause fulminant liver failure
HEV	HEV-like viruses	7.5 kb	RNA+	Faeco-oral	No	Can cause fulminant liver failure

Figure 60.1 Hepatitis virus virology. (a) The virology of viral hepatitis (HAV to HEV) is summarized. (b) The HBV circular DNA and the multiple overlapping reading frames. (c) In contrast to HBV the HCV has a linear RNA genome encoding viral proteins. The HCV 5′ and 3′ internal ribosome re-entry site (IRES) are important viral regulatory elements.

HBV is a circular DNA virus that belongs to the *hepadnaviruses* that include woodchuck hepatitis virus and ground squirrel hepatitis virus (Chisari, 1992; Lee, 1997; Lok, 2000). The partially double stranded HBV genome is 3.2 kb in length and is organized with multiple overlapping open reading frames (ORF) (Chisari, 1992; Lee, 1997; Lok, 2000) (Figure 60.1). Greater than half of the genome is translated in more than one ORF and this limits the viral mutations that will be tolerated. HBV DNA incorporates into the host genome (Lau and Wright, 1993; Moyer and Mast, 1994). The viral proteins include the surface (envelope) protein, core, polymerase and X-proteins. Variations of the nucleotide sequence and corresponding amino acids that constitute the HBV surface protein give rise to 8 common HBV genotypes (A–H).

HCV is a positive strand RNA virus, a member of the *Flaviviridae* family that includes flaviviruses and pestiviruses (Shimotohno, 2000). There is considerable diversity in the HCV genome with at least 6 distinct genotypes, more than 50 subtypes and a propensity of the virus to mutate giving rise to "*quasi species*" (Forns and Bukh, 1999; Forns et al., 1999). The virus genome is 9.5 kb that encodes a large single polyprotein of 3010 to 3033 amino acids (Forns and Bukh, 1999) (Figure 60.1). The 5′ and 3′

untranslated regions are important for the replication of the virus and translation of the polyprotein. The HCV proteins comprise the structural proteins; core protein and the envelope glycoproteins E1, E2 and p7 followed by the non-structural proteins; proteases, helicase (NS2 and NS3), protease cofactor (NS4A), NS4B, replication associated phosphoprotein (NS5A) and the RNA-dependent polymerase (NS5B) (Shimotohno, 2000; Simmonds, 1996). Possible receptors for HCV have been identified and these include CD-81, SRB-1 and Claudin-1.

Hepatitis D (HDV also known as delta hepatitis) is unlike any other transmissible agent in animals (Taylor, 2006a, b). This minus strand circular RNA virus requires the hepatitis B surface antigen (HBsAg) for encapsulation and entry into hepatocytes. HDV only co-infects up to 5% individuals with HBV infection. Importantly, the prognosis of HDV co-infection is worse than with HBV alone.

Hepatitis E (HEV) is related to *Calciviradie* but now has been classified into a separate genus of *hepatitis E-like viruses* (Krawczynski et al. 2000; Worm et al., 2002). The HEV genome consists of a single positive RNA strand of 7.5 kb that is translated into three polyproteins as there are three overlapping ORF. Consequently HEV has stable genome sequence and exists as a single serotype and at least four unique genotypes. HEV is stable in the environment and is common waterborne pathogen especially in developing countries.

ACQUISITION AND PREDISPOSITION TO VIRAL HEPATITIS

HAV is characterized by an enteric route of infection and classically is transmitted via personal contact, illicit drug use or ingestion of contaminated food or water (Martin and Lemon, 2006). The incubation period following exposure can be up to 90 days. HAV infection typically occurs in sporadic outbreaks as well as isolated cases and can be endemic in poorly developed countries with poor sanitation. Although molecular phenotyping of the virus helps trace outbreaks, there is very little known about individual genetic factors that may predispose to HAV infection. Host genetic factors may not be that important as indicated by the epidemiologic studies showing that HAV infection is often transmitted to persons with similar risk factors. HEV is similarly transmitted to HAV and is characterized by a great propensity for waterborne transmission. Like HAV, HEV is endemic in many poorly developed countries. HEV has the propensity to be associated with miscarriage and an increased mortality in pregnant women especially during the third trimester. The basis for this is unclear, but a number of host factors are clearly important in determining susceptibility to infection.

HBV or HCV enter the host through blood contact by direct inoculation (i.e., needle or transfusion) or via a disrupted percutaneous barrier (i.e., sexual or perinatal transmission) (Shimotohno, 2000; Simmonds, 1996). Entry of HBV or HCV into liver hepatocytes is not understood, and there are extra-hepatic reservoirs of infection in peripheral blood leukocytes (Cabrerizo et al., 2000; Yoffe et al., 1990). Receptor-based cell entry has been implicated in the both HBV and HCV pathogenesis (Cocquerel et al., 2006; Pileri et al., 1998; Yerly et al., 2006). The immunopathogenesis of both HBV and HCV infection is characterized by both an innate and adaptive immune response to the virus resulting in predominantly non-specific inflammation. Typically, neither HBV nor HCV are directly cytopathic as viral load does not correlate with organ damage and the level of antigen expression in hepatocytes does not correlate with hepatocyte injury (Chisari and Ferrari, 1995). Importantly, both HBV and HCV *in vivo* infectivity is limited to higher primates (chimpanzees and humans).

Predisposing factors to viral hepatitis include a number of non-genetic and genetic factors (Thomas, 2000; Wasley and Alter, 2000; Yee, 2004) (see Table 60.1). HBV and HCV are more prevalent in communities with increased rates of intravenous drug use, unsafe therapeutic injections and with the use of unscreened blood products (Maddrey, 2000; Merican et al., 2000; Wasley and Alter, 2000). Fortunately, the recognition of routes of transmission, the use of safe injecting practices and the adoption of volunteer donor blood transfusion services are widely instigated global public health measures used to control the spread of viral hepatitis. Vertical transmission of HBV is dependent on the stage of maternal HBV infection and the viral antigen expression. Approximately, 10–20% of surface antigen positive mothers transmit the virus to their offspring. However, both surface and HBeAg expression is associated with a 90% rate of transmission. Maternal acute HBV infection results in a 10% neonatal rate of infection in the first trimester, which increases to 80–90% if acute infection occurs in the third trimester. HCV infection is characterized by lower rates of vertical transmission with only 2.7–8.4% of offspring being infected. The presence of HIV co-infection increases the rate of vertical transmission of both HBV and HCV. Transmission via sexual activity with mucosal disruption is more prevalent in HBV than HCV infection.

Genetic factors predispose to persistence of HBV infection given the rate of concordance for surface antigen expression is greater in monozygotic compared to dizygotic twins (Table 60.1). The twin concordance data in HCV infection are not as convincing. There are a number of HLA alleles which are associated with both the persistence and the clearance of both HBV and HCV (Wang, 2003; Yee, 2004). Interestingly, in contrast to HIV progression homozygosity for HLA class II locus increases the risk of HBV persistence. The HLA class II locus DRB1*1302 is associated with HBV clearance and the DQB1*0301 locus is associated with a self-limiting course of HCV infection. Non-HLA immunogenic is also implicated in viral hepatitis with TNF promoter polymorphisms resulting in higher TNF secretion being associated with HBV clearance. The killer cell immunoglobulin-like receptors (KIR) genes interact with HLA class I molecules and specific KIR heliotypes are associated with HCV clearance (Martin and Carrington, 2005; Williams et al., 2005). Importantly, the majority of genetic predispositions identified in viral hepatitis is linked to viral persistence or clearance and can be directly implicated in the adaptive immune response. The current

TABLE 60.1 Viral hepatitis genetic susceptibility associations

Allele/Polymorphism	Hepatitis C	Hepatitis B	Comment
HLA-DRB1*1302 and HLA-DRB1*0301		Spontaneous elimination of infection	
HLA-DQA1*0501, DQB1*301, HLA-DRB1*1102 and HLA-DRB1*0301		Persistence of infection	
HLA-DRB1*0101, HLA-DRB1*0401, HLA-DRB1*15, HLA-DRB1*1101, HLA-DRB1*0301, HLA-A*2301, HLA-A*1101, HLA-A*03, HLA-B*57 and HLA-Cw*0102	Spontaneous elimination of infection		
HLA-DRB1*0701, HLA-A*01-B*08-Cw*07-DRB1*0301-DBQ1*0201, HLA-Cw*04, HLA-Cw*04-B*53	Persistence of infection		
TNF promoter	Viral replication and clearance	Viral replication and clearance	Polymorphisms at -308 and -238 best characterized
Interleukin-10		Spontaneous elimination of infection	
Vitamin D receptor		Control of viral replication	Expressed on monocytes and lymphocytes
GDNF family receptor alpha 1	At risk of HCC in HCV		
Chemokine (C-X-C motif) ligand 14	At risk of HCC in HCV		Previously known as SCYB14

HCC: Hepatocellular carcinoma.

documented genetic disease associations with viral hepatitis are of limited use in clinical practice. Presently genome-wide association studies are being undertaken to comprehensively characterized hepatitis genetic susceptibility. Therefore, future clinical practice is likely to see panels of genetic susceptibility markers being screened to determine prognosis and the likelihood of a treatment response.

SCREENING AND DIAGNOSIS OF VIRAL HEPATITIS

HAV and HEV are characterized by jaundice, fever and a diarrhea illness. Given the propensity for epidemic outbreaks and known endemic regions of the globe (mainly developing countries) establishing the diagnosis is not difficult. HAV and HEV are readily diagnosed using serology with acute infection distinguished by the presence of antiviral IgM (Acharya and Panda, 2006; Panda et al., 2006). The diagnosis of HDV is also made on serology and anti-HDV IgM is detectable by 30 days following infection (Fiedler and Roggendorf, 2006; Weston and Martin, 2001).

HBV and HCV infections results in a comparatively non-specific cluster of symptoms ranging from asymptomatic viral infection (most common in HCV), jaundice and fever (30–50% of HBV infections) to fulminant hepatic failure and death (<1% of HBV and exceedingly rare in HCV). Importantly, both HBV and HCV viral replication can be associated with a normal liver panel. Although routine liver tests such as transaminases are used to monitor disease effects on the liver, they do not provide prognostic information and cannot be used for the diagnosis of viral hepatitis. The cornerstone of screening and diagnosis is serology for viral-specific antibodies and antigens. The commonly available serology of HBV is complex with the pattern of antibody and antigen expression determining the nature and timing of infection (Table 60.2).

HCV antibody production is a simpler test of exposure which becomes positive within 4 weeks of exposure (Bhandari and Wright, 1995). The commonly used HCV antibody screening is performed with using an enzyme immunoassay (EIA) and is prone to false positive readings. Specificity of the EIA has improved with successive generations of test but confirmation of a positive result with supplemental methodology is considered mandatory. A positive HCV EIA is routinely confirmed using a

TABLE 60.2 Hepatitis B virus serology and DNA quantitation

Serologic markers	Definition/clinical use
HBsAg – Hepatitis B surface antigen	• General marker of infection • First serologic marker to appear after infection • Persistence for >6 months = chronic infection
HBeAg – Hepatitis B e antigen	• Indicates active replication of virus
Anti-HBs – Antibody to hepatitis B s antigen	• Documents recovery and/or immunity to HBV • Detectable after immunity conferred by HBV vaccination
Anti-HBe – Antibody to hepatitis B e antigen	• Marker of reduced level of replication
Anti-HBc (IgM) – Antibody to hepatitis B core antigen	• Marker of acute HBV
Anti-HBc (IgG) – Antibody to hepatitis B core antigen	• Marker of current or past infection
HBV DNA – Hepatitis B virus genomic DNA	• Marker of HBV replication • Used for monitoring response to therapy

recombinant immunoblot assay in which serum reactivity to a number of HCV recombinant viral proteins is assessed.

Detection of HBV DNA and HCV RNA is not recommended routinely to establish the diagnosis (Acharya and Panda, 2006; Mondelli et al., 2005; Patel et al., 2006; Servoss and Friedman, 2004). However, as viral replication can occur with normal biochemistry and antibody titers do not demonstrate viral replication quantitative assessment of viral nucleotide presence is used in many centers as the final confirmatory diagnostic test for active HBV or HCV infection. Quantitation of HBV DNA or HCV RNA is used in planning and following treatment of the infection. HBV DNA can be quantitated using an approach of capturing HBV DNA on a full length HBV RNA transcript and using antibodies to detect the RNA:DNA hybrid (known as capture assays or "Digene"™). An alternative approach uses quantitative PCR methodologies to amplify the HBV DNA (COBAS Amplior™) or a variation of the capture approach using branched DNA ("Versant"™).

HBV and HCV viral levels are typically determined once the infection has been confirmed as they carry prognostic significance as well as being used to monitor treatment responses. HBV genotyping is not routinely performed in the clinical setting unless the patient is being considered for interferon-alpha therapy.

An important aspect of screening and diagnosis of viral hepatitis involves an assessment of the extent of liver injury attributable to the virus. In particular the extent of hepatic inflammation and fibrosis has implications for the progression and the likely response to treatment. Routine biochemistry gives only limited information about the extent of inflammation or fibrosis, and most centers will use imaging of the liver with ultrasound or other modalities combined with a liver biopsy in assessing virus induced liver injury. However, biopsy is an invasive procedure prone to significant sampling and observer error (Siegel et al., 2005). This has led to the development of a number of new non-invasive screening modalities. The most promising approaches use multiple variables to assess for liver injury, and these methods are likely to become commonplace in the next decade.

PATHOGENESIS OF VIRAL HEPATITIS

It is clear from the natural history and prognosis that the aim of chronic HBV and/or HCV treatment is to stop the development of cirrhosis and the sequelae of hepatic decompensation and HCC. Genomic medicine is likely to have a significant impact in the future determination of viral hepatitis outcomes and identification of at risk individuals for development of cirrhosis and sequelae such as HCC. Genomic studies have made significant contributions to our understanding of HBV and HCV pathogenesis and treatment responses (see Table 60.3). The pathogenesis of viral hepatitis is unique for HBV and HCV (Chisari and Ferrari, 1995; Lee and Locarnini, 2004; Tanikawa, 2004). However, there are similarities in the mechanisms of HBV and HCV clearance and persistence. The initial innate immune response characterized by interferon (IFN) and IFN stimulated gene (ISG) expression. Chronic infection with either virus is characterized by both antigen-specific and non-specific CD4+ and CD8+ T-cell responses that are pivotal in virus clearance or in chronic infection responsible for ongoing inflammation resulting in liver injury. Further, both HBV and HCV employ the strategy of mutational escape to evade the adaptive immune response. The mode of entry of HBV or HCV into the hepatocytes is unknown. Receptor-mediated viral entry is thought likely, given the restricted cell population infected (Cocquerel et al., 2006; Pileri et al., 1998; Yerly et al., 2006). In HBV carboxypeptidase D (gp180) has been shown to interact with the preS portion of the large viral surface protein (Yerly et al., 2006). In HCV CD-81, scavenger receptor class B type 1 and claudin-1 have all been implicated in viral entry (Cocquerel et al., 2006; Pileri et al., 1998). Currently, no conclusive evidence of a

Table 60.3 Key findings arising from genomic studies of viral hepatitis

	Hepatitis C	Hepatitis B	Comments
Acute infection	Interferon stimulated gene expression (Mx1, ISG15) correlated with viral clearance	"Stealth virus" Innate immune response is abrogated Immune evasion	Acute HBV or HCV infection is poorly characterized by genomic studies
Chronic infection	Th1 immuno-phenotype perpetuating chronic injury	T-cell effector function activated and latter B-cell related gene expression observed	
Treatment responses	Interferon stimulated gene induction correlate with treatment responses		HBV treatment response have been poorly characterized by genomic studies
Hepatocellular carcinoma	HCV core protein oncogenic	HBV x (HBx) protein oncogenic	
Bio-markers	A number of non-specific inflammatory markers identified Potential tumor markers identified	A number of non-specific inflammatory markers identified	
Future studies	Predicting HCV IFN treatment responses	Predicting HBV IFN and nucleotide/nucleoside treatment responses	Determining the pathogenesis of and predicting chronicity, disease progression and development of sequelae such as HCC

single receptor or the presence of a receptor complex has been demonstrated for either HBV or HCV (Cocquerel et al., 2006; Yerly et al., 2006).

There are marked differences in the pathogenesis of both HBV and HCV. Innate immune responses are blunted and do not appear to play an important role in HBV clearance. In contrast a strong innate immune response is thought to be important in HCV clearance. Finally, antibody-derived immunity is present in HBV infection whilst it appears to be of no importance in HCV infection.

HBV Pathogenesis

HBV infection results in the formation of a double-stranded HBV genome in the nucleus that is converted into a covalently closed circular double-stranded DNA (cccDNA). Further, HBV DNA integrates into the host genome and this has significant implications in the long-term management of HBV, as the complete elimination of the virus is not possible and viral reactivation characterized by active replication is possible in the future. The cccDNA acts as a template for the formation of an RNA replicative intermediate, which is prone to a high rate of mutation of 1 in 10^5 bases. However, HBV replication after infection is characterized by low level replication reaching levels of 10^2–10^4 genome equivalents per ml for up to 6 weeks after infection. Once established HBV is associated with extremely high viral titers of 10^8–10^{13} genome equivalents per ml. This has led to the assertion that HBV initially evades the immune response before becoming established. Importantly, once established, active HBV replication can result in infection of 100% of the intrahepatic hepatocytes.

The innate immune activation in HBV infection is abrogated. However, HBV clearance can occur prior to the induction of an adaptive immune response. This is thought to be mediated by IFN-α and -β (type I IFN) in a non-classical manner that is proteosome dependent. Further antigen independent natural killer (NK) cell activation is thought to be responsible for IFN-γ induction and induction of an adaptive immune response.

Adaptive immunity in HBV is characterized by CD8+ T-cell response to surface antigen epitopes with secretion of IFN-γ and TNF which have direct antiviral effects principally by controlling HBV replication at the stage of formation of the RNA replicative intermediate. In contrast to HCV, HBV antibody-mediated humoral immunity effectively neutralizes the virus. However, antibody production is often absent in chronic HBV infection by mechanisms that are poorly understood. Viral protein such as the X-protein inhibit proteosome dependent control of virus replication also both surface antigen and precore protein act as tolerogens abrogating the T-cell response to the virus.

HCV Pathogenesis

HCV replication involves formation of a negative sense replicative RNA strand and the subsequent formation of dsRNA. This has several important implications for HCV pathogenesis. First, the formation of RNA intermediates means that there is no stable genomic replicative form of the virus and as a result HCV must produce new viral RNA and proteins to maintain persistence.

Secondly, the formation of dsRNA associated with HCV replication is a target for endogenous RNA interference and elicits the endogenous IFN response (Yu et al., 2000). Further, the NS5B RNA polymerase of HCV lacks proof reading activity and as a result virus replication is highly error prone (1 in 10^3 bases) resulting in remarkable genetic diversity. HCV is currently divided into six major genotypes with many subtypes that differ by up to 35% in their nucleotide sequence. HCV infection is characterized by a rapid increase in circulating levels to 10^5–10^7 equivalents per ml. The rapid induction of an immune response means that not all hepatocytes are infected although the true proportion of hepatocytes infected is unknown.

The innate immune response is characterized by induction of type I IFN, interferon-stimulted genes (ISGs) and a NK response. The IFN gene expression is induced by the induction of endogenous RNA interference pathways, the formation of dsRNA that binds to the RNA helicases RIG-1 and MDA5 and binding of the phagocytosed infected cell fragments to toll-like receptor (TLR)-3 (Honda et al., 2006; Kato et al., 2006; Yoneyama et al., 2005). These upstream events then signal through IRF-3 phosphorylation resulting in IFN gene transcription. The IFN gene expression then signals via cognate receptors and activates the JAK/stat pathway resulting in the induction of ISGs including protein kinase R, RNA–specific adenosine deamsinase-1 (ADAR-1), P56 and 2′-5′ oligoadenylate synthetase. Most of these ISGs act on the formation of the negative replicative strand of HCV. The cellular innate immune response is characterized by induction of an NK cell response. NK cells destroy infected cells in an antigen independent manner via cytotoxic cell lysis. These activated NK cells secrete large amounts of IFN-γ, which activates and maintains a cellular adaptive immune response.

Adaptive immune responses in HCV infection is characterized by the virus-specific CD4+ and CD8+ T-cell response to multiple HCV epitopes, including many highly promiscuous epitopes formed due to the high spontaneous rate nucleotide mutation rate associated with HCV replication (Bowen and Walker, 2005a, b). This T-cell response is accompanied and maintained by induction of IFN-γ and TNF, both of which can directly inhibit viral replication without killing an infected cell. Although HCV antibody production is universal in immunocompetent individuals, it does not prevent infection or correlate with outcome. Further the virus-specific T-cell response is maintained for decades after HCV clearance in contrast to the antibody responses which can become undetectable.

Genomic Analyses of HCV Infection

Gene array analysis of HCV recurrence in liver transplant allografts has provided novel insights into the molecular mechanisms of viral recurrence (Mansfield and Sarwal, 2004; McCaughan and Zekry, 2004). HCV recurrence in the liver graft is associated with expression of IFN-γ associated genes such as CXCL10 (IP-10), CXCL9 (HuMIG) and RANTES (McCaughan and Zekry, 2004). Further, antiviral IFN-α associated gene expression is seen in chronic HCV recurrence and during acute rejection associated with HCV recurrence (McCaughan and Zekry, 2004). Additionally, upregulation of NF-kappa β pathway during acute rejection in association with HCV recurrence appears to alter cellular apoptosis via changes in the expression of TRIAL-associated genes (McCaughan and Zekry, 2004). Importantly chronic HCV recurrence in grafts is associated with Th1 associated gene expression similar to that seen in chronically HCV infected individuals that have not been transplanted (McCaughan and Zekry, 2004). In contrast cholestatic HCV recurrence, which follows an aggressive course, is associated with a Th2 cytokine profile (McCaughan and Zekry, 2004). This suggests that the Th1 immune response suppresses viral replication whilst being profibrogenic (McCaughan and Zekry, 2000, 2004; Shackel et al., 2002). In cholestatic HCV recurrence, the unchecked viral replication is directly fibrogenic (McCaughan and Zekry, 2004; McCaughan and Zekry, 2000).

A particular challenge in the study of the effect of viruses on liver cells is the difficulty in infecting liver cells with virus. The studies described below have involved models in which cultured cells are infected with viral proteins or viral genome. Progress in this field has been rapid and most recently, a cellular model of HCV infection has been reported that is likely to stimulate further study (Heller et al., 2005; Lindenbach et al., 2005).

Proteomic methodologies have been applied to a number of aspects of HCV-related liver injury. However, to date most proteomic studies have focused on the identification of a number of biomarkers of disease rather than trying to unravel aspects of HCV pathobiology. Proteomic studies have defined potential protein therapeutic targets that interact with HCV in detail. Large-scale proteome analysis of a full-length HCV replicon revealed prominent expression of proteins involved in lipid metabolism (Jacobs et al., 2005). Several *in vitro* proteomic studies have identified proteins that interact with specific HCV proteins. Heat shock protein 27 (Hsp27) was shown to specifically interact with NS5A via the N-terminal regions (Choi et al., 2004). Fourteen cellular proteins binding to the core protein were identified by proteomics (Kang et al., 2005). These proteins include DEAD-box polypeptide 5 (DDX5) and intermediate microfilament proteins, including cytokeratins (cytokeratin 8, cytokeratin 19 and cytokeratin 18) and vimentin. Interestingly, DDX5 gene polymorphisms are associated with accelerated fibrosis development in HCV infected individuals (see Chapter 50) (Huang et al., 2006).

The development of HCC and IFN treatment response are two further aspects of HCV infection studied using proteomics. In the study of HCV-related HCC development, over-expression of alpha enolase was identified and correlated with poorly differentiated HCC (Kuramitsu and Nakamura, 2005; Takashima et al., 2005). The response of hepatocyte cell lines to IFN-γ treatment has uncovered over 54 IFN response genes, including many novel targets an approach that may pave the way for novel therapies. Examination of protein extracts that bind to the HCV IRES

has identified a number of novel protein targets such as Ewing Sarcoma breakpoint 1 region protein EWS and TRAF-3. The final aspect of HCV liver injury receiving attention is the study of potential biomarkers such as heat shock protein 70 Hsp-70 associated with HCV infection progression to HCC (Takashima et al., 2003).

THERAPEUTICS AND PHARMACOGENOMICS

The principal treatment goal in viral hepatitis is clearance of the virus with a secondary goal of averting or delaying the onset of cirrhosis, hepatic decomposition and HCC. Immune modulators in the form of IFN treatment have been the mainstay of treatment for years (see Table 60.4). Antiviral therapy has now become an effective treatment option in HBV, but in HCV is useful only when combined with IFN treatment. Finally, HAV and HBV infection are reliably protected against by immunization. There is no prospect in the foreseeable future of a vaccine for HCV. The treatment options for viral hepatitis are summarized in Table 60.4. Predicting individual's treatment responses based on gene expression is likely to be an area in which genomic medicine will enable highly directed individual therapy in treatment of viral hepatitis.

Treatment of HBV Infection

IFN is the only treatment shown to clear HBV infection in chronically infected individuals without the development of drug resistance. In both HBeAG positive and negative individuals with chronic HBV infection, IFN treatment for 4–6 months has been shown to normalize liver function abnormalities, result in clearance of HBsAg and HBeAg and result in a sustained loss of HBV DNA. In a meta-analysis of 15 studies, suppression of HBV DNA was seen in 37% of patient's loss of HBeAg was seen in 33% of subjects and HBeAg seroconversion was seen in 18% of patients. Subjects responded to treatment if they had lower pretreatment HBV DNA levels and higher pretreatment liver transaminase levels. The advent of long-lasting IFN preparations using pegylated IFN has been shown to have an additional benefit over conventional IFN in treating HBV infection resulting in improved HBsAg and HBeAg seroconversion. However, the current meta-analysis of the treatment outcomes with IFN in HBV does not support its use in preventing HCC. Compared to nucleoside analogs, discussed below, one principle advantage of IFN therapy for HBV is the durability of the treatment response with $<10\%$ of individuals having a relapse in HBeAg expression up to 8 years latter.

Nucleoside analogs are now being increasingly used to treat HBV infection and these agents target the HBV DNA polymerase. Lamivudine is the most widely used nucleoside analog and effectively suppresses HBV replication as evident by a greater than two log decline in HBV viral DNA titers. Lamivudine results in 16–18% HBeAg seroconversion in 1 year and 50% after 5 years. The durability of the response is 77% at 3 years. However, the sustained virological response in other studies has been reported as 39% at 4 years, and sustained response to lamivudine following cessation of therapy is significantly less than IFN. Continued treatment with lamivudine results in sustained suppression of HBV viral replication but is limited by the appearance of mutant forms of the HBV polymerase typically in a conserved YMDD motif at methoine 204 of the enzyme. After 1 year resistance develops in 14–32% of cases and this increase to 50% at 2 years and 74% at 5 years.

Other approved nucleoside analogs include adefovir, tenofovir, telbirudine and entecavir. All have significant activity against HBV replication although both tenofovir and entecavir would appear to have greater activity against HBV. Importantly resistance with all of these newer agents is uncommon, with entecavir resistance being less than 5% at 2 years. Unfortunately resistance to these newer agents appears inevitable. In a situation analogous to HAART treatment of HIV-1 infection, combination therapy is now being studied in HBV infection. The conclusive outcome of these studies is not yet available but the initial results are promising especially in cases of lamivudine resistance.

TABLE 60.4 Treatment of viral hepatitis

Virus	Chronicity	Treatment	Vaccine	Comments
HAV	No	Immune globulin	Yes	
HBV	Yes	Immune globulin Nucleoside Analogues (target DNA polymerase) Immune mediators (i.e., IFN-α)	Yes	Viral resistance to nucleoside analogues common
HCV	Yes	Immune mediators (i.e., IFN-α and Ribavirin)	No	Small molecular inhibitors in clinical trials
HDV	Yes	Treatment of HBV	No	
HEV	No	None	No	No specific treatment Avoidance

Treatment of HCV Infection

IFN treatment is the only effective antiviral therapy available for the treatment of HCV. The current recommendations are for a 24–48 week course of treatment using pegylated IFN combined with the antiviral ribavirin. Ribavirin is a guanasine analog able to inhibit the replication of viruses but in the absence of IFN has no significant effect on HCV RNA levels. The overall chance of a sustained virological response (SVR) varies according to HCV genotype. In genotype 1 infection, SVR can be achieved in 42–46% of patients, with better response rates of 76–88% for those with genotype 2 or 3 infection. Virologic response to treatment can be predicted from the decline in HCV RNA at 12 weeks; The absence of a 2 log drop or undetectable HCV RNA at week 12 has a high negative predictive value for the absence of SVR in genotype 1 patients with continued therapy. In genotype 2 and 3 the decline in viral load at 4 weeks may be predictive of achieving SVR with only 12–16 weeks of therapy. The treatment of patients with cirrhosis is controversial but there appears to be a benefit in avoiding progression of disease, decompensation, and the development of HCC. However, in the presence of hepatic decompensation, IFN-based treatment is contraindicated and patients should be referred for transplant evaluation.

Genomic Studies Related to the Treatment of Viral Hepatitis

IFN-α is currently part of the standard of care treatment for HCV infection. Several studies have used microarray analysis to identify the mechanisms by which IFN-α acts on hepatocytes and the HCV. IFN-α activated the multiple signal transducer and activator of transcription factors (STAT) 1, 2, 3, 5 in cultured hepatocytes (Radaeva et al., 2002). Other upregulated genes include a variety of antiviral and tumor suppressors/pro-apoptotic genes. Downregulated genes include c-myc and c-Met and the hepatocyte growth factor (HGF) receptor (Radaeva et al., 2002). In a second and comparable study, IFN-α antiviral efficacy was associated with 6–16 (G1P3) expression. Involvement of STAT3 in IFN-α signaling was confirmed (Zhu et al., 2003). Resistance to IFN-α antiviral activity may be mediated the HCV viral protein, NS5A. To identify the mechanisms through which NS5A blocks IFN activity, gene expression profile was studied in IFN-treated Huh7 cells expressing NS5A. The strongest effect of NS5A on IFN response was observed for the OAS-p69 gene (Girard et al., 2002). Another key response of hepatocytes to the HCV virus is cellular proliferation. Gene array studies identified upregulation of growth-related genes, in particular wnt-1 and its downstream target gene WISP (Fukutomi et al., 2005). In another study, CDK activity, hyperphosphorylation of Rb, and E2F activation was shown to be associated with hepatocyte proliferation induced by a full-length HCV clone (Tsukiyama-Kohara et al., 2004).

Global quantitative proteomic analysis in a human hepatoma cell line (Huh7) in the presence and absence of IFN was performed to examine liver-specific responses to IFN and the mechanisms of IFN inhibition of virus replication (Yan et al., 2004). Fifty-four proteins were induced by IFN and 24 were repressed, representing several novel and liver-specific key regulatory components of the IFN response.

Molecular markers used on an individual patient basis might improve prediction of treatment responses prior to commencement of therapy. Previously Chen and colleagues examined liver biopsies prior to IFN-α and ribavirin therapy from 16 responders, 15 non-responders and 20 normal individuals by gene array analysis and determined that 18 genes were predictive of an SVR (Chen et al., 2005). These investigators identified a gene expression signature of 8 genes that could predict the likelihood to achieve an SVR in 30 of 31 individuals (GIP2/IFI15/ISG15, ATF5, IFIT1, MX1, USP18/UBP43, DUSP1, CEB1, and RPS28) (Chen et al., 2005). The striking outcome from this study was that these genes, known to be involved in IFN responsiveness, were overexpressed in non-responders and formed part of the predictive gene signature profile. Additionally, two genes (ISG15/IFI15 and USP18/UBP43) were identified as part of a previously unrecognized novel IFN regulatory pathway (Chen et al., 2005; Randall et al., 2006). In a further study of peripheral blood mononuclear cells (PBMC), a relative lack of ISG expression was associated with a poor response to antiviral therapy with pegylated IFN (Taylor et al., 2007) and Feld and colleagues present work demonstrating unique patterns of liver gene expression that correlate with IFN and ribavirin treatment responses in HCV genotype 1 infection (Feld et al., 2007).

Pharmacogenomics of Viral Hepatitis

Pharmacogenomic studies of viral hepatitis are currently evolving as the treatment options become better and the understanding of the disease pathobiology improves. In HBV infection a group of 82 chronic active carriers received standard IFN-α treatment for 6 months (King et al., 2002; Randall et al., 2006). These patients were concurrently studied for single nucleotide polymorphisms (SNPs) in genes involved in the JAK/stat signaling of IFN and in genes leading to the expression of ISGs (King et al., 2002; Randall et al., 2006). Two SNPs were identified that appeared to predict response; one in the promoter region of the ISG MxA and the other in the IFN regulated eIF-2α gene. This study is significant in demonstrating that host polymorphisms may correlate with treatment response in HBV infection (King et al., 2002; Randall et al., 2006). In HCV infection the non-structural protein NS5a is known to bind viral RNA and alter HCV replication (Feld et al., 2007; Goyal et al., 2006; Kohashi et al., 2006; Taylor et al., 2007). Amino acid substitutions within the interferon sensitivity-determining region (ISDR) of NS5A correlates with IFN-α treatment responses (Feld et al., 2007; Goyal et al., 2006; King et al., 2002; Kohashi et al., 2006; Taylor et al., 2007; Watanabe et al., 2005). A meta-analysis of ISDR NS5a mutations demonstrates a relative risk of 4.66 to 5.73 of IFN-α treatment response compared to non-mutant ISDR

HCV (Goyal et al., 2006; Schinkel et al., 2004). However, this effect is more pronounced in Japanese compared to European patients (Goyal et al., 2006; Kohashi et al., 2006; Pascu et al., 2004; Schinkel et al., 2004). These studies highlight the gaps in our knowledge of disease pathogenesis and host genomics that influence disease and the lack of high-specific treatment options means that pharamacogenomic measures are still in their infancy.

2009 UPDATE

There has been renewed interest in adopting emerging high-throughput technologies for large-scale gene expression profiling to predict antiviral treatment response in chronic hepatitis C (CHC) patients. A hepatic gene expression study in biopsies from 69 CHC patients identified a two-gene signature (*IFI27* and *CXCL9*) that could predict treatment response in 79% of patients (Asselah et al., 2008). This confirms prior observations in studies using both PBMC and liver tissues (ISG) appear to be upregulated in pre-treatment liver tissues from patients that do not respond to current standard-of-care antiviral therapy with pegylated IFN-alpha and ribavirin. A defective hepatic response to IFN may be mediated through activation of suppressor of cytokine signaling 3 (*SOCS3*) (Huang et al., 2007). An association study that evaluated three SNPs of *SOCS3* in 162 CHC genotype-1 nonresponder and 184 CHC sustained responder patients noted that *SOCS3*-4874 AA genotype was strongly associated with nonresponse to IFN-based therapy, and these patients had significantly higher *SOCS3* mRNA and protein levels (Persico et al., 2008). The IFN-gamma promoter SNP -764G/C also appears functionally important in determining viral clearance and treatment response in HCV infection (Huang et al., 2007).

Emerging specific targeted antiviral therapies for HCV will continue to be combined with IFN and there is likely to be continued interest in determining mechanisms of IFN resistance and predictors of virologic response to therapy through SNP associations of cytokine and other host signaling pathways (Selzner et al., 2008). Plasma proteome profiling studies incorporating emerging quantitative and identification bioinformatics tools to assess CHC virologic responses to IFN-based therapy are ongoing. In contrast, the availability of several oral nucleoside and nucleotide analogs that are effective against HBV may explain the absence of functional genomic or plasma protein profiling studies to assess therapeutic responses in chronic hepatitis B infection.

2009 UPDATE REFERENCES

Asselah, T., Bieche ,I., Narguet, S., Sabbagh, A., Laurendeau, I., Ripault, M.P., Boyer, N., Martinot-Peignoux, M., Valla, D., Vidaud, M. et al. (2008). Liver gene expression signature to predict response to pegylated interferon plus ribavirin combination therapy in patients with chronic hepatitis C. *Gut* 57, 516–524.

Huang, Y., Feld, J.J., Sapp, R.K., Nanda, S., Lin, J.H., Blatt, L.M., Fried, M.W., Murthy, K. and Liang, T.J. (2007). Defective hepatic response to interferon and activation of suppressor of cytokine signaling 3 in chronic hepatitis C. *Gastroenterology* 132, 733–744.

Huang, Y., Yang, H., Borg, B.B., Su, X., Rhodes, S.L., Yang, K., Tong, X., Tang, G., Howell, C.D., Rosen, H.R. and Thio, C.L. (2007). A functional SNP of interferon-gamma gene is important for interferon-alpha-induced and spontaneous recovery from hepatitis C virus infection. *Proc Natl Acad Sci U S A* 104, 985–990.

Persico, M., Capasso, M., Russo, R., Persico, E., Croce, L., Tiribelli, C. and Iolascon, A. (2008). Elevated expression and polymorphisms of SOCS3 influence patient response to antiviral therapy in chronic hepatitis C. *Gut* 57, 507–515.

Selzner, N., Chen, L., Borozan, I., Edwards, A., Heathcote, E.J. and McGilvray, I. (2008). Hepatic gene expression and prediction of therapy response in chronic hepatitis C patients. *J Hepatol* 48, 708–713.

REFERENCES

Acharya, S.K. and Panda, S.K. (2006). Hepatitis E virus: Epidemiology, diagnosis, pathology and prevention. *Trop Gastroenterol* 27(2), 63–68.

Bhandari, B.N. and Wright, T.L. (1995). Hepatitis C: An overview. *Annu Rev Med* 46, 309–317.

Bowen, D.G. and Walker, C.M. (2005). Adaptive immune responses in acute and chronic hepatitis C virus infection. *Nature* 436(7053), 946–952.

Bowen, D.G. and Walker, C.M. (2005). Mutational escape from CD8+ T cell immunity: HCV evolution, from chimpanzees to man. *J Exp Med* 201(11), 1709–1714.

Cabrerizo, M., Bartolome, J., Caramelo, C., Barril, G. and Carreno, V. (2000). Molecular analysis of hepatitis B virus DNA in serum and peripheral blood mononuclear cells from hepatitis B surface antigen-negative cases. *Hepatology* 32(1), 116–123.

Chen, L., Borozan, I., Feld, J., Sun, J., Tannis, L.L., Coltescu, C. et al. (2005). Hepatic gene expression discriminates responders and nonresponders in treatment of chronic hepatitis C viral infection. *Gastroenterology* 128(5), 1437–1444.

Chisari, F.V. (1992). Hepatitis B virus biology and pathogenesis. *Mol Genet Med* 2, 67–104.

Chisari, F.V. and Ferrari, C. (1995). Hepatitis B virus immunopathogenesis. *Annu Rev Immunol* 13, 29–60.

Choi, Y.W., Tan, Y.J., Lim, S.G., Hong, W. and Goh, P.Y. (2004). Proteomic approach identifies HSP27 as an interacting partner of

the hepatitis C virus NS5A protein. *Biochem Biophys Res Commun* 318(2), 514–519.

Cocquerel, L., Voisset, C. and Dubuisson, J. (2006). Hepatitis C virus entry: Potential receptors and their biological functions. *J Gen Virol* 87(Pt 5), 1075–1084.

Feld, J. J., Nanda, S., Huang, Y., Chen, W., Cam, M., Pusek, S.N. et al. (2007). Hepatic gene expression during treatment with peginterferon and ribavirin: Identifying molecular pathways for treatment responses. *Hepatology* 46, 1548–1563.

Fiedler, M. and Roggendorf, M. (2006). Immunology of HDV infection. *Curr Top Microbiol Immunol* 307, 187–209.

Flehmig, B. (1990). Hepatitis A. *Baillieres Clin Gastroenterol* 4(3), 707–720.

Forns, X. and Bukh, J. (1999). The molecular biology of hepatitis C virus. Genotypes and quasispecies. *Clin Liver Dis* 3(4), 693–716, vii.

Forns, X., Purcell, R.H. and Bukh, J. (1999). Quasispecies in viral persistence and pathogenesis of hepatitis C virus. *Trends Microbiol* 7(10), 402–410.

Fukutomi, T., Zhou, Y., Kawai, S., Eguchi, H., Wands, J.R. and Li, J. (2005). Hepatitis C virus core protein stimulates hepatocyte growth: Correlation with upregulation of wnt-1 expression. *Hepatology* 41(5), 1096–1105.

Girard, S., Shalhoub, P., Lescure, P., Sabile, A., Misek, D.E., Hanash, S. et al. (2002). An altered cellular response to interferon and up-regulation of interleukin-8 induced by the hepatitis C viral protein NS5A uncovered by microarray analysis. *Virology* 295(2), 272–283.

Goyal, A., Hofmann, W.P., Hermann, E., Traver, S., Hissar, S.S., Arora, N. et al. (2006). The hepatitis C virus NS5A protein and response to interferon alpha: mutational analyses in patients with chronic HCV genotype 3a infection from India. *Med Microbiol Immunol (Berl)* 6.

Heller, T., Saito, S., Auerbach, J., Williams, T., Moreen, T.R., Jazwinski, A. et al. (2005). An *in vitro* model of hepatitis C virion production. *Proc Natl Acad Sci USA* 102(7), 2579–2583.

Honda, K., Takaoka, A. and Taniguchi, T. (2006). Type I interferon [correction of inteferon] gene induction by the interferon regulatory factor family of transcription factors. *Immunity* 25(3), 349–360.

Huang, H., Shiffman, M.L., Cheung, R.C., Layden, T.J., Friedman, S., Abar, O.T. et al. (2006). Identification of two gene variants associated with risk of advanced fibrosis in patients with chronic hepatitis C. *Gastroenterology* 130(6), 1679–1687.

Jacobs, J.M., Diamond, D.L., Chan, E.Y., Gritsenko, M.A., Qian, W., Stastna, M. et al. (2005). Proteome analysis of liver cells expressing a full-length hepatitis C virus (HCV) replicon and biopsy specimens of posttransplantation liver from HCV-infected patients. *J Virol* 79(12), 7558–7569.

Kang, S.M., Shin, M.J., Kim, J.H. and Oh, J.W. (2005). Proteomic profiling of cellular proteins interacting with the hepatitis C virus core protein. *Proteomics* 5(8), 2227–2237.

Kato, H., Takeuchi, O., Sato, S., Yoneyama, M., Yamamoto, M., Matsui, K. et al. (2006). Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature* 441(7089), 101–105.

King, J.K., Yeh, S.H., Lin, M.W., Liu, C.J., Lai, M.Y., Kao, J.H. et al. (2002). Genetic polymorphisms in interferon pathway and response to interferon treatment in hepatitis B patients: A pilot study. *Hepatology* 36(6), 1416–1424.

Kohashi, T., Maekawa, S., Sakamoto, N., Kurosaki, M., Watanabe, H., Tanabe, Y. et al. (2006). Site-specific mutation of the interferon sensitivity-determining region (ISDR) modulates hepatitis C virus replication. *J Viral Hepat* 13(9), 582–590.

Krawczynski, K., Aggarwal, R. and Kamili, S. (2000). Hepatitis E. *Infect Dis Clin North Am* 14(3), 669–687.

Kuramitsu, Y. and Nakamura, K. (2005). Current progress in proteomic study of hepatitis C virus-related human hepatocellular carcinoma. *Expert Rev Proteomics* 2(4), 589–601.

Lau, J.Y. and Wright, T.L. (1993). Molecular virology and pathogenesis of hepatitis B. *Lancet* 342(8883), 1335–1340.

Lee, J.Y. and Locarnini, S. (2004). Hepatitis B virus: Pathogenesis, viral intermediates, and viral replication. *Clin Liver Dis* 8(2), 301–320.

Lee, S.D. (2003). Recent advance on viral hepatitis A. *J Chin Med Assoc* 66(6), 318–322.

Lee, W.M. (1997). Hepatitis B virus infection. *N Engl J Med* 337(24), 1733–1745.

Lindenbach, B.D., Evans, M.J., Syder, A.J., Wolk, B., Tellinghuisen, T.L., Liu, C.C. et al. (2005). Complete replication of hepatitis C virus in cell culture. *Science* 309(5734), 623–626.

Lok, A.S. (2000). Hepatitis B infection: Pathogenesis and management. *J Hepatol* 32(Suppl 1), 89–97.

Maddrey, W.C. (2000). Hepatitis B: An important public health issue. *J Med Virol* 61(3), 362–366.

Mansfield, E.S. and Sarwal, M.M. (2004). Arraying the orchestration of allograft pathology. *Am J Transplant* 4(6), 853–862.

Martin, A. and Lemon, S.M. (2006). Hepatitis A virus: From discovery to vaccines. *Hepatology* 43(2 Suppl 1), S164–S172.

Martin, M.P. and Carrington, M. (2005). Immunogenetics of viral infections. *Curr Opin Immunol* 17(5), 510–516.

McCaughan, G.W. and Zekry, A. (2000). Effects of immunosuppression and organ transplantation on the natural history and immunopathogenesis of hepatitis C virus infection. *Transpl Infect Dis* 2(4), 166–185.

McCaughan, G.W. and Zekry, A. (2004). Mechanisms of HCV reinfection and allograft damage after liver transplantation. *J Hepatol* 40(3), 368–374.

Merican, I., Guan, R., Amarapuka, D., Alexander, M.J., Chutaputti, A., Chien, R.N. et al. (2000). Chronic hepatitis B virus infection in Asian countries. *J Gastroenterol Hepatol* 15(12), 1356–1361.

Mondelli, M.U., Cerino, A. and Cividini, A. (2005). Acute hepatitis C: Diagnosis and management. *J Hepatol* 42(Suppl 1), S108–S114.

Moyer, L.A. and Mast, E.E. (1994). Hepatitis B: Virology, epidemiology, disease, and prevention, and an overview of viral hepatitis. *Am J Prev Med* 10(Suppl), 45–55.

Panda, S.K., Thakral, D. and Rehman, S. (2006). Hepatitis E virus. *Rev Med Virol* 19.

Pascu, M., Martus, P., Hohne, M., Wiedenmann, B., Hopf, U., Schreier, E. et al. (2004). Sustained virological response in hepatitis C virus type 1b infected patients is predicted by the number of mutations within the NS5A-ISDR: A meta-analysis focused on geographical differences. *Gut* 53(9), 1345–1351.

Patel, K., Muir, A.J. and McHutchison, J.G. (2006). Diagnosis and treatment of chronic hepatitis C infection. *BMJ* 332(7548), 1013–1017.

Pileri, P., Uematsu, Y., Campagnoli, S., Galli, G., Falugi, F., Petracca, R. et al. (1998). Binding of hepatitis C virus to CD81. *Science* 282(5390), 938–941.

Radaeva, S., Jaruga, B., Hong, F., Kim, W.H., Fan, S., Cai, H. et al. (2002). Interferon-alpha activates multiple STAT signals and down-regulates c-Met in primary human hepatocytes. *Gastroenterology* 122(4), 1020–1034.

Randall, G., Chen, L., Panis, M., Fischer, A.K., Lindenbach, B.D., Sun, J. et al. (2006). Silencing of USP18 potentiates the antiviral activity of interferon against hepatitis C virus infection. *Gastroenterology* 131(5), 1584–1591.

Schinkel, J., Spoon, W.J. and Kroes, A.C. (2004). Meta-analysis of mutations in the NS5A gene and hepatitis C virus resistance to interferon

therapy: Uniting discordant conclusions. *Antivir Ther* 9(2), 275–286.

Servoss, J.C. and Friedman, L.S. (2004). Serologic and molecular diagnosis of hepatitis B virus. *Clin Liver Dis* 8(2), 267–281.

Shackel, N.A., McGuinness, P.H., Abbott, C.A., Gorrell, M.D. and McCaughan, G.W. (2002). Insights into the pathobiology of hepatitis C virus-associated cirrhosis: Analysis of intrahepatic differential gene expression. *Am J Pathol* 160(2), 641–654.

Shimotohno, K. (2000). Hepatitis C virus and its pathogenesis. *Semin Cancer Biol* 10(3), 233–240.

Siegel, C.A., Silas, A.M., Suriawinata, A.A. and van Leeuwen, D.J. (2005). Liver biopsy 2005: When and how? *Cleve Clin J Med* 72(3), 199–201, 6, 8 passim.

Simmonds, P. (1996). Virology of hepatitis C virus. *Clin Ther* 18(Suppl B), 9–36.

Takashima, M., Kuramitsu, Y., Yokoyama, Y., Iizuka, N., Toda, T., Sakaida, I. et al. (2003). Proteomic profiling of heat shock protein 70 family members as biomarkers for hepatitis C virus-related hepatocellular carcinoma. *Proteomics* 3(12), 2487–2493.

Takashima, M., Kuramitsu, Y., Yokoyama, Y., Iizuka, N., Fujimoto, M., Nishisaka, T. et al. (2005). Overexpression of alpha enolase in hepatitis C virus-related hepatocellular carcinoma: Association with tumor progression as determined by proteomic analysis. *Proteomics* 5(6), 1686–1692.

Tanikawa, K. (2004). Pathogenesis and treatment of hepatitis C virus-related liver diseases. *Hepatobiliary Pancreat Dis Int* 3(1), 17–20.

Taylor, J.M. (2006). Structure and replication of hepatitis delta virus RNA. *Curr Top Microbiol Immunol* 307, 1–23.

Taylor, J.M. (2006). Hepatitis delta virus. *Virology* 344(1), 71–76.

Taylor, M.W., Tsukahara, T., Brodsky, L., Schaley, J., Sanda, C., Stephens, M.J. et al. (2007). Changes in gene expression during pegylated interferon and ribavirin therapy of chronic hepatitis C virus distinguish responders from nonresponders to antiviral therapy. *J Virol* 81(7), 3391–3401.

Thomas, D.L. (2000). Hepatitis C epidemiology. *Curr Top Microbiol Immunol* 242, 25–41.

Tsukiyama-Kohara, K., Tone, S., Maruyama, I., Inoue, K., Katsume, A., Nuriya, H. et al. (2004). Activation of the CKI-CDK-Rb-E2F pathway in full genome hepatitis C virus-expressing cells. *J Biol Chem* 279(15), 14531–14541.

Wang, F.S. (2003). Current status and prospects of studies on human genetic alleles associated with hepatitis B virus infection. *World J Gastroenterol* 9(4), 641–644.

Wasley, A. and Alter, M.J. (2000). Epidemiology of hepatitis C: Geographic differences and temporal trends. *Semin Liver Dis* 20(1), 1–16.

Watanabe, K., Yoshioka, K., Yano, M., Ishigami, M., Ukai, K., Ito, H. et al. (2005). Mutations in the nonstructural region 5B of hepatitis C virus genotype 1b: Their relation to viral load, response to interferon, and the nonstructural region 5A. *J Med Virol* 75(4), 504–512.

Weston, S.R. and Martin, P. (2001). Serological and molecular testing in viral hepatitis: An update. *Can J Gastroenterol* 15(3), 177–184.

Williams, A.P., Bateman, A.R. and Khakoo, S.I. (2005). Hanging in the balance KIR and their role in disease. *Mol Interv* 5(4), 226–240.

Worm, H.C., van der Poel, W.H. and Brandstatter, G. (2002). Hepatitis E: An overview. *Microbes Infect* 4(6), 657–666.

Yan, W., Lee, H., Yi, E.C., Reiss, D., Shannon, P., Kwieciszewski, B.K. et al. (2004). System-based proteomic analysis of the interferon response in human liver cells. *Genome Biol* 5(8), R54.

Yee, L.J. (2004). Host genetic determinants in hepatitis C virus infection. *Genes Immun* 5(4), 237–245.

Yerly, D., Di Giammarino, L., Bihl, F. and Cerny, A. (2006). Targets of emerging therapies for viral hepatitis B and C. *Expert Opin Ther Targets* 10(6), 833–850.

Yoffe, B., Burns, D.K., Bhatt, H.S. and Combes, B. (1990). Extrahepatic hepatitis B virus DNA sequences in patients with acute hepatitis B infection. *Hepatology* 12(2), 187–192.

Yoneyama, M., Kikuchi, M., Matsumoto, K., Imaizumi, T., Miyagishi, M., Taira, K. et al. (2005). Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity. *J Immunol* 175(5), 2851–2858.

Yu, S.H., Nagayama, K., Enomoto, N., Izumi, N., Marumo, F. and Sato, C. (2000). Intrahepatic mRNA expression of interferon-inducible antiviral genes in liver diseases: dsRNA-dependent protein kinase overexpression and RNase L inhibitor suppression in chronic hepatitis C. *Hepatology* 32(5), 1089–1095.

Zhu, H., Zhao, H., Collins, C.D., Eckenrode, S.E., Run, Q., McIndoe, R.A. et al. (2003). Gene expression associated with interferon alfa antiviral activity in an HCV replicon cell line. *Hepatology* 37(5), 1180–1188.

Index

Page numbers followed by *f* indicate a figure; page numbers followed by *t* indicate tabular material

A

Abacavir hypersensitivity reactions
  genetic basis, 224
ABCA1, 276–277
  blood lipid associations, 279
  common gene variants, 276
  deficiency (Tangier disease), 276
  high-density-lipoprotein metabolism, 271
  therapeutic targeting, 282
ABCB1, 68, 387
  methylation status, tumor treatment response prediction, 67
ABCG5 mutations
  sitosterolemia, 273–274
ABCG8 mutations, 283
  sitosterolemia, 274
Abducens nerve disorders, 713
Abetalipoproteinemia, 274, 282
ABI SOLiD (Supported Oligonucleotide Ligation and Detection), 34, 35
Academy of Managed Care Pharmacy (AMCP), 235, 238, 240
ACCE framework, 249
ACCESS (A Case Control Etiologic Study of Sarcoidosis), 618
Access to research results by participants, 211–212
Acetaldehyde dehydrogenase
  alcoholic liver disease, 652
N-Acetylcysteine, chronic obstructive pulmonary disease, 611
N-Acetyltransferase 2 (NAT2) polymorphism, 464–465
ACTC mutations, hypertrophic cardiomyopathy, 338
Actin
  hypertrophic cardiomyopathy, 338
  multiple sclerosis biomarker, 570
Activated protein C, sepsis therapy, 765, 767, 765
ACTN2 mutations (alpha-actinin-2), hypertrophic cardiomyopathy, 338, 341
Acute coronary syndromes, 303–310
  biochemical risk stratification markers, 308–309, 309*t*
  diagnosis, 305–308, 305*f*
    metabolomics, 306–307, 307*f*
    protein biomarkers, 305–306, 306*t*
    proteomics, 307–308, 307*f*, 308*f*
  leukotriene inhibitor therapy, 297
  risk factors *see* Cardiovascular risk
  screening, 304
Acute lymphoblastic leukemia
  L-asparaginase treatment, 425–426
  corticosteroid-resistant, 83
    sirolimus response study, 116
  cytogenetics, 421, 426
  gene expression profiling, 76, 78, 113, 113*f*, 424
    drug response studies, 116, 426, 427
    prognostically important subtypes, 166, 424
  6-mercaptopurine response, *TPMT* pharmacogenetics, 185
  treatment, 423
Acute myeloid leukemia
  cytogenetics, 421, 426
  Flt-3 targeted therapy, 542
  gemtuzumab ozogamicin therapy, 538
  gene expression profiling, 78, 113, 113*f*, 166, 424, 425*f*
    drug response prediction, 427
  genomic data reporting, 143
  prognostic factors, 423
  subtypes, 166, 421
Acute promyelocytic leukemia
  all-trans retinoic acid treatment, 422, 424, 536
  pharmacogenomics, 424
Acute stress ulcers, 627
Acylcarnitines
  targeted metabolic profiling, 102
    insulin resistance, 103, 104
Acylcoenzyme A, targeted metabolic profiling, 102
ADA, obesity, 666
Adalimumab, rheumatoid arthritis, 556
ADAM22, multiple sclerosis, 567
ADAM33
  asthma, 593
  chronic obstructive pulmonary disease, 607
ADAMTS10, multiple sclerosis, 567
ADBR2, obesity, 666
ADBR3, obesity, 666, 669, 671
Adducin, antihypertensive agent pharmacogenomics, 265
Adefovir, hepatitis B treatment, 781
Adenomatous polyps, 458, 461
  adenoma to carcinoma sequence, 458–460, 459*f*
  APC mutation-negative, 461

Adenomatous polyps, (*Continued*)
  criteria for genetic testing, 466
  familial adenomatous polyposis, 461
  family history, 464
  see also Colorectal adenoma
S-Adenosyl methionine synthetase, alcoholic liver disease, 652
ADH *see* Alcohol dehydrogenase
Adhesins, *Helicobacter pylori*, 633
Adipocyte metabolism
  fatty acids, 270
  genetic factors, 668
Adipose tissue, lipoprotein metabolism, 270
Admixed populations, 146
  Mexican Genome Project, 211
Admixture mapping, 19
  multiple sclerosis, 567
ADRB2 gene, 597, 612
ADRB2 polymorphism
  asthma, 591, 593
  β_2-agonist pharmacogenomics, 597
  see also β_2 adrenergic receptor
Adrenocorticotropic hormone 4–10 (ACTH4-10), obesity management, 671
Adrenocorticotropin (ACTH), depression, 724
Adult T-cell leukemia/lymphoma, 401
Advanced glycation end products (AGEs), 377
Adverse drug reactions, 175
  antihypertensive agents, 265
  drug development process, 230
  pharmacogenetics applications, 224
    preventve testing, 230
Affymetrix HU6800 oligonucleotide array, 410
  lymphoma, 410
Affymetrix Human Genome-Wide SNP Array 6.0, 43
Affymetrix Mapping 500K array, 40, 42, 51
Affymetrix microarrays, 41, 73, 485, 677
  Burkitt lymphoma, 415–416
  colorectal cancer, 468
  comparative aspects, 42
  copy number variation (CNVs) detection, 42, 46
  diffuse large B-cell lymphoma subtype differentiation, 409, 410–411
  DNA processing steps, 41, 41*f*
  fibroblast gene expression profiling, 78
  follicular lymphoma, 413
  GeneChip, 41, 41*f*, 42, 74
  gliomas, 522–527
  hepatic stellate cells, 645, 649–650
  Hodgkin lymphoma, 412–413
  primary mediastinal large B-cell lymphoma, 411–412
  probe length, 42
  tiling arrays, 82
  whole genome amplified DNA, 34
Affymetrix U95 microarray, 410, 413
Affymetrix U133 microarray, 468
Affymetrix U133A+B, 410, 411, 412, 414, 416
African Americans
  heart failure
    BiDil treatment, 175, 319
    hypertension relationship, 314
  lung cancer, 433, 434
  myocardial infarction risk, 294
    LTA4H HapK haplotype association, 304
    PCSK9 nonsense/missense variants, 294, 295*f*
  prostate cancer risk, 478
African-American Heart Failure (A-HeFT), 319
AFT-5, Hodgkin lymphoma/Reed-Sternberg cells, 412–413
Age associations
  fibroblast gene expression profiling, 78
  hypertension, 260, 262*t*
  sporadic colorectal cancer, 460
  see also Aging process
Age-related macular degeneration *see* Macular degeneration
Agilent microarrays, 73
  oligonucleotide arrays, 74–75
Aging process
  gene expression profiles, 78
  liver, 655
  see also Age associations
AGR2, ovarian cancer, 495
AGRP (agouti-related protein)
  body weight regulation, 662
  obesity, 669
AHEAD, 63*t*
AIB1 (SRC3)
  ovarian cancer metastasis, 497
Air pollution
  chronic obstructive pulmonary disease, 603
AKAP9 (LQT11), 365
AKT
  colorectal cancer, 459
  ovarian cancer, 495
  see also P13 kinase/Akt pathway
Alagille syndrome (*JAG1/NOTCH2*), 394*f*
  congenital heart disease, 393
Albumin cobalt binding (ACB) test, 306
Albumin, sarcoid bronchoalveolar fluid, 619
Albuterol, asthma, 596
Alcohol consumption
  cancer risk
    colorectal, 464, 465
    head and neck, 511
Alcohol dehydrogenase (ADH)
  alcoholic liver disease, 652
Alcoholic liver disease
  proteomics, 653–655
  transcriptomics, 651–653
Alcoholism
  secondary dementia, 687
Alcoholomics, 654
Aldehyde dehydrogenase, alcoholic liver disease, 652, 654
Alemtuzumab, 538
  chronic lymphocytic leukemia, 538
Alkaline lysis plasmid isolation, 29
Alkaptonuria, 13
Alkylating agents, DNA methylation profiles in tumor response prediction, 67
All-trans retinoic acid, acute promyelocytic leukemia, 422, 424, 536, 540
Allele-specific primer extension (ASPE), 40
Allelic imbalance, 20
Allergen hypersensitivity, asthma, 591
AlloMap® Molecular Expression test, 334
AlloMap test (cardiac allograft rejection)
  clinical use, 333–334
  corticosteroid dose effects, 332, 333*f*
  development, 330
  endomyocardial biopsy "gold standard" relationship, 331
  pathways monitored, 331, 331*t*
  post-transplant time effects, 330
  rejection predicition, 330–331
ALOX5, leukotriene inhibitor pharmacogenomics, 596
ALOX5AP
  myocardial infarction risk, 294, 297
  therapeutic inhibition, 297
AlpA/B, *Helicobacter pylori* adhesins, 633
Alpha enolase, hepatocellular carcinoma, 780
Alpha myosin heavy chain (*MYH6*) mutations, hypertrophic cardiomyopathy, 338
α-1 antitrypsin
  hepatitis B infection, 651
  sarcoid bronchoalveolar fluid, 619
α-1 antitrypsin deficiency
  chronic obstructive pulmonary disease, 604
α-2 adrenoceptors, antihypertensive agent pharmacogenomics, 264
Alpha-actinin-2 (*ACTN2*) mutations, hypertrophic cardiomyopathy, 338, 341
α-methylmalonyl co-A racemase, prostate cancer urine biomarker, 481
α-synuclein
  gene mutations
    diffuse Lewy body dementia/Parkinson disease with dementia, 692
    Parkinson's disease (*SNCA; PARK1; PARK4*), 702
  Lewy bodies, 702
Alpha-tropomyosin (*TPM1*), hypertrophic cardiomyopathy, 338, 341, 342
$\alpha\beta$ crystallin gene mutations
  autosomal dominant congenital lamellar cataract, 715
  dilated cardiomyopathy, 715
  protein aggregation cardiomyopathy (PAC; desmin-related myopathy), 321
AlzGene, 691
Alzheimer's disease, 687, 688, 689–691
  APOE4 association, 191, 270, 689–690, 691, 694

early-onset, 689
amyloid precursor protein processing abnormalities, 689, 690*f*
inheritance patterns, 689
late onset, 689–691, 695
candidate genes, 691
genetic factors, 689
pathology, 689, 689*f*
AMACR, gene expression in prostate cancer, 483
American Indians
lung cancer, 433
American Society of Clinical Oncology (ASCO), 186
Amino acids, targeted metabolic profiling, 102
AML1-ETO fusions gene, acute myeloid leukemia, 421
Amoxicillin, *Helicobacter pylori* eradication, 628, 635
AMPD1, obesity, 666
AmpliChip Cytochrome P450 test, 185, 195
AmpliTaq Gold, 32
Amy1 gene, 57
Amy2 gene, 57
Amygdala reactivity, 726
Amyloid precursor protein, 689
α-/β-/γ-secretase cleavage, 689, 690*f*
candidate gene studies, 691
Alzheimer's disease
Aβ42 fragment toxicity, 689
APP mutations, 689
processing abnormalities, 689, 690*f*
Amyloidosis, interstitial lung disease, 616
Amyotrophic lateral sclerosis (Lou Gehrig's disease), 691
frontotemporal dementia association, 691
Analysis of variance, 118
Analyte-specific reagents (ASRs), 226
regulatory issues, 226, 227
Anastrozole, 455, 533
Ancestry informative markers (AIMs), 19
Ancestry testing, 19
Andersen syndrome *see* Andersen-Tawil syndrome
Andersen-Tawil syndrome, 352, 353, 358
clinical features, 358
genetics, 353, 358
LQT7 (Kir 2.1; KCNJ2), 358
Androgen receptor
prostate cancer, 485
circulating tumor cells, 482
hormone-refractory disease, 485
transcriptional targets, 485–487
Androgen response elements (AREs), 485
Androgens
ablation therapy, 486
prostate cancer dependence, 485
Anemia, chronic with heart failure, 314
Aneuploidy, 14
Angiogenesis
therapeutic targets
head and neck cancer, 526
ovarian cancer, 498
Angiopoietins
ovarian cancer angiogenesis, 498
Angiotensin I (AT1) receptor variants, hypertension-related stroke, 263
Angiotensin II receptor blockers
cirrhosis, 343
hypertrophic cardiomyopathy, 343
pharmacogenomics, 264
Angiotensin-converting enzyme (ACE) inhibitors
adverse drug reactions, 265
cirrhosis, 343
hypertrophic cardiomyopathy, 343
pharmacogenomics, 264
heart failure, 319
Angiotensin-converting enzyme (ACE) polymorphism
ACE inhibitor-related cough, 265
depression, 727
heart failure treatment response, 319
host response to *Neisseria meningitidis*, 752
obesity, 668
sarcoidosis, 617
ANGPTL4 gene variants, 276, 277, 278, 283
Aniridia, PAX6 mutation, 712
Ankryn-β (*LQT4*; Ankβ), 353, 356
Annexin A2, alcoholic liver disease, 652
ANRIL, 298
Anterior chamber, 712
Antiangiogenic agents
small molecules, 542–543
Antibiotics
chronic obstructive pulmonary disease, 610
drug-induced long QT syndromes, 352
Antibodies
microarrays, surface plasmon resonance imaging (SPRI), 135–136
Anti-cancer agents
antibody-based, 535–540, 536*t*
drug metabolizing enzyme measurement, 180
small molecule-based, 541*t*, 542–543
susceptibility
bioinformatics applications, 114–115
DNA methylation markers, 67
gene microarray measurements relationship, 114–115, 115*f*
pharmacogenomics, 426
Anticardiolipin antibodies, thrombotic event prediction, 380, 381*f*, 382*f*
Antidepressants
cytochrome P450 metabolism, 726, 726*t*
long-term effects, 728
pharmacogenomics, 726–727
Antigen arrays, antibody repertoire investigation in multiple sclerosis, 570
Antihistamines, drug-induced long QT syndromes, 352
Anti-hormonal therapies, DNA methylation profiles in tumor response prediction, 67
Antihypertensive agents
adverse drug reactions, 265
pharmacogenomics, 264–265
Antimalarials, rheumatoid arthritis, 558
Antioxidants, 435, 512, 607, 608
Antiplatelet agents
pharmacogenomics, 296–297
Anti-*Saccharomyces cerevisiae* antibodies (ASCA)
inflammatory bowel disease, 578, 579, 580, 581
Antithrombin III, 374
sepsis pathogenesis, 765
Antithrombotic therapy, pharmacogenomics, 384–385
see also Warfarin
Antral G cells
hyperplasia, peptic ulceration, 627
Anxiety
depression comorbidity, 723
Aortic dissection, Marfan syndrome, 393
APACHE II score, 767
APC (adenomatous polyposis coli)
mutations
colorectal cancer, 65, 458, 460, 466
familial adenomatous polyposis, 461
promoter hypermethylation in lung cancer, 434
APOA1 (apolipoprotein A-1), 654
APOA4 (apolipoprotein A-1V), hepatitis B infection proteomics, 671
APOA5 (apolipoprotein A-V)
blood lipid genome-wide association studies, 279
variants, 276, 277, 278
ApoA-I, 269, 271
deficiency/structural mutation, high-density lipoprotein cholesterol reduction, 274, 276
cardiovascular risk, 279
hepatitis B, 651, 654, 655
ApoA-I_{Milano}, 274, 276
ApoA-II, 269
ApoA-V deficiency, 277
ApoB, 269–270, 271, 377
mutation, familial hypobetalipoproteinemia, 274
therapeutic targeting, 281
antisense oligonucleotides, 281–282
variants, 277, 278, 279
low-density lipoprotein C interindividual variation, 274
obesity, 671
ApoB-48, 270
ApoB-100, 270
mutation (familial defective apolipoprotein B-100), 272
ApoC-I, 270

ApoC-II, 270
mutation, familial chylomicronemia syndrome, 277, 279
ApoC-III, 270
ApoE, 270
ApoE4, 198
APOE (apolipoprotein E), 30
blood lipid genome-wide association studies, 279
E4 allele *see* *APOE4*
mutations, hyperlipoproteinemia type III (familial dysbetalipoproteinemia), 277
thrombotic event prediction with anticardiolipin antibodies, 380
variants
LDL cholesterol interindividual variation, 274
ApoE2, 277, 283
ApoE3, 277
APOE4, 191, 283
Alzheimer's disease association, 689–690, 691
diffuse Lewy body dementia, 693
Apolipoprotein(a), 270
Apolipoproteins, 269
Apoptosis
cardiomyocytes in heart failure, 316–317, 318*f*
APP mutations, Alzheimer's disease, 689
Appetite regulation, genetic factors, 667
AR methylation status, tumor treatment response prediction, 67
AR-targeting therapies, 487
Arachidonate 5-lipoxygenase-activating protein *see* ALOX5AP
ARF, protate cancer methylation markers, 67
Arg381Gln, 583
Arginase 1, 597
ARH mutations, autosomal recessive hypercholesterolemia, 272
Arimidex, 455
Array comparative genomic hybridization, 50
congenital heart disease, 396
copy number variation (CNVs) detection, 46, 53
clinical cytogenetic diagnostics, 54, 55*f*
gliomas, 523, 524
prostate cancer biochemical relapse after surgery, 483
Array Express, 119, 120
Arrhythmias, 350–366
atrial arrhythmia syndromes, 363*t*
familail ventricular tachycardia/ catecholaminergic polymorphic ventricular tachycardia (CPVT), 350, 361–362
primary conduction abnormalities, 362–364
short QT interval syndrome, 361
see also Long QT syndromes; Short QT interval syndrome
Arrhythmogenic right ventricular cardiomyopathy, 314, 337
ARSC1, corneal disease, 713
Arterial thrombosis
recurrence risk, 397
risk factors, 383*t*
risk prediction (Bloodomics project), 385–386
screening investigations, 380–381
Arthropathy, inflammatory bowel disease association, 581
Artificial neural networks, 76, 118
proteomics
hepatitis B infection, 654
liver function, 655
Aryl hydrocarbon receptor (AHR), 572
Asbestos exposure
L-Asparaginase, 425
pharmacogenomics, 425–426
response prediction, 426
Aspirin
colorectal cancer chemoprevention, 468
pharmacogenomics, 296–297
resistance, 296–297
Assembly (genome assembler software), 30
Association studies
Alzheimer's disease (late onset), 690
asthma, 593–594
copy number variation (CNVs), 54
depression, 724–725, 725*t*
hypertension, 262
infectious diseases susceptibility, 735, 748
multiple sclerosis, 567
myocardial infarction, 291–294
obesity, 664, 666
replicability, 763
sepsis, 762–763
single nucleotide polymorphisms (SNPs), 27
thrombosis, 379–380
translation into medical practice, 164
see also Genome-wide association studies; Linkage studies (genome scans)
Asthma, 27, 590
allergen hypersensitivity, 591
family history, 591
gene–environment interactions, 591
genomics, 594, 595*t*
candidate-gene association studies, 593–594
genome-wide association studies, 594
linkage analysis, 591–593, 592*t*
susceptibility genes, 593
IL-4RA/IL-13 polymorphism, 523
pathobiology, 590–591, 591*f*
chronic obstructive pulmonary disease comparison, 607, 608*t*
pharmacogenomics, 594–597
pharmacotherapy, 594–597
Astrocytoma, 522, 523*t*
anaplastic, alkylating agent chemoresistance, 527
cytogenetics, 524
gene expression profiles, 526
genetic alterations, 523, 524*f*
progression, 523
Ataxia-telangiectasia
breast cancer association, 448
pancreatic cancer association, 502
ATBF1, Hodgkin lymphoma/Reed-Sternberg cells, 413
ATG16L1, Crohn's disease, 578
ATG16L1 gene, 583
Atherosclerosis
DNA methylation in pathogenesis, 65
fibrin variation, 377
leukotriene pathway associations, 294
plaque rupture risk, 290
biomarkers, 297
genetic influence, 289, 290, 294, 304
matrix metalloproteinases, 290
tissue inhibitors of metalloproteinases (TIMPs), 290
vascular-occlusive dementia, 687
Atherosclerosis Risk In Communities (ARIC) study, 298
Athletes, hypertrophic cardiomyopathy, 343
Atlas assembler, 30
Atlas cDNA Expression array, follicular lymphoma, 413
ATM mutations, breast cancer, 448
Atomoxitine, 230
ATP1A2, obesity, 666
ATP-binding cassette protein A1 *see* ABCA1
ATP-binding cassette protein B1 *see* ABCB1
ATP-binding cassette protein G5 *see* ABCG5 mutations
ATP-binding cassette protein G8 *see* ABCG8 mutations
Atrial arrhythmia syndromes, 363*t*
Atrial fibrillation, 363
clinical features, 363–364
electrophysiology, 364
genetics, 364
mechanisms, 364
Atrial natriurteic peptide gene, hypertension-related stroke association, 263
Atrial septal defects, 391
Atrial tachyarrhythmias, 350
Atrioventricular block, 350
Lev-Lenegre progressive cardiac conduction disease, 362, 363*t*
Attenuated familial adenomatous polyposis, 461
Autism
copy number variation (CNVs), 194
Autism spectrum disorder (ASD), 56
Autoimmune disease
DNA methylation alterations, 65
Autoimmune liver disease
immunogenetics, 648, 648*t*
proteomics, 653–655
transcriptomics, 651–653

Autosomal dominant congenital lamellar cataract, 715
Autosomal dominant hypercholesterolemia, 273, 278, 282
Autosomal recessive hypercholesterolemia, 272–273, 278
Avastin, 528
see also Bevacizumab
Avellino's (combined granular-lattice) corneal dystrophy, 714, 715
Azathioprine, 327
inflammatory bowel disease pharmacogenomics, 582
response biomarkers, 185

B

B cells
chronic obstructive pulmonary disease, 606
multiple sclerosis immune response, 570
sarcoidosis, 617
transplant rejection
heart, 327
kidney, 116
B-type natriuretic peptide
acute coronary syndromes risk stratification, 308, 309
heart failure diagnosis, 317–318
BAALC, acute myeloid leukemia prognosis, 423
BabA1, 633
BabA2, *Helicobacter pylori* adhesins, 633, 634
peptic ulcer disease correlations, 634, 636
Bacteria, 29, 101, 194, 556, 606, 631
Bacterial infections, 610, 744–754
genomics
bacterial, 744–745
host, 745–748
host susceptibility
Gram-negative organisms, 751–753
Gram-positive organisms, 748–751
IRAK4 mutations, 736
study methods, 748
see also Infectious disease
BAG1/Bag-1
head and neck cancer, 517
Oncotype Dx assay, 533
Bannyan-Ruvalcabe-Riley syndrome, PTEN pathway mutations, 464
BAT26, colorectal cancer, 467
Bayesian methods, 118
Bayesian networks, 118
Bcl-2
cardiomyocyte apoptosis in heart failure, 316
diffuse large B-cell lymphoma, 411
follicular lymphoma, 413, 414
head and neck cancer, 516
pathway, 442
therapeutic targeting, 543
Bcl-6, diffuse large B-cell lymphoma, 410, 411
Bcl-Xl, head and neck cancer, 516
BCR-ABL fusions gene
acute lymphoblastic leukemia, 424
Philadelphia chromosome (chronic myelogenous leukemia), 421
BCR-ABL tyrosine kinase
chronic myelogenous leukemia, 422, 424
imatinib mesylate inhibition, 422, 424, 534
BDNF *see* Brain-derived neurotrophic factor
Beclomethasone, asthma management, 596, 596*f*
Bengin prostatic hypertrophy, urine biomarkers, 481
Benzo(a)pyrene carcinogenesis
head and neck cancer, 512
Benzo(a)pyrene diol epoxide, 512
β_1-adrenergic receptor, pharmacogenomics
antihypertensive agents, 265
beta blocker response in heart failure, 319
β_2 adrenergic receptor
heart, 357
pharmacogenomics
antihypertensive agents, 265
β_2-agonists, 596
post-myocardial infarction therapy response, 296
β_2-agonists
asthma, 594
non-responsers, 596
chronic obstructive pulmonary disease, 610
pharmacogenomics, 594–597
β myosin heavy cain mutations (*MYH7*), hypertrophic cardiomyopathy, 313, 336, 338, 340, 341
β-Blocker Evaluation of Survival Trial (BEST), 319
Beta-blockers
catecholaminergic polymorphic ventricular tachycardia (CPVT), 350
hypertrophic cardiomyopathy, 343
long QT syndromes, 359–360
pharmacogenomics, 265
heart failure, 319
post-myocardial infarction response, 296
β-defensins, 52
11β-hydroxylase/aldosterone synthase fusion gene, glucocorticoid-remediable aldosteronism, 263
11β-hydroxysteroid dehydrogenase mutations, syndrome of apparent mineralocorticoid excess, 263
β-interferon, multiple sclerosis treatment, 569
Betet quid chewing, 511
Bevacizumab, 539, 545
prostate cancer trials, 485
BF, obesity, 666
Bias
case–control studies, 177
Bicuspid aortic valve (*NOTCH1*), 393
BiDil (nitrate/hydralazine combination), heart failure management in African-Americans, 175, 319
Bile acid sequestrants, 281
familial hypercholesterolemia, 272
Bile ducts, 652
Biliary atresia, 652
Biliary disease transcriptomics, 652–653
Biobanks, 145, 146, 165–166
electronic medical records, 146
policy issues, 210, 211–213
public consultation approaches, 212
public health applications, 249
Bioconductor, 123
Bioinformatics, 111–126, 167
analytic methods, 118–119
challenges, 167
medical applications, 112–117
diagnosis, 112–113
histopathology, 116
nosology, 116–117
sepsis studies, 766
therapeutics, 114–116
proteomics, 92–94
modified peptide identification, 93
public health genomics, 245
software tools (free), 121–124
analytic, 122–123
interpretation, 123–124
vocabularies/ontologies, 120–121, 120*t*
Biological processes, Gene Ontology (GO), 121
The Biomarker Consortium, 171, 309
Biomarker studies, 612
Biomarkers, 167
applications, 203
clinical event prediction, 203
clinical trial populations
genome-based trial design, 167–168
colorectal cancer, 467
coronary artery disease
acute coronary syndromes, 303, 305, 307–308, 308*f*
myocardial infarction, 296
prognostic stratification, 308–309, 309*t*
screening, 304
early disease, 133
FDA Critical Path Initiative, 309
head and neck cancer, 513, 515
locoregional recurrence prediction, 516–517
heart failure, 320–322, 322*f*
inflammatory bowel disease, 578, 579
leukemia outcome prediction, 426–427
liver disease, 654, 655
lung cancer diagnosis, 434, 435
multiple sclerosis, 570–571
prostate cancer, 481*f*
circulating tumor cells, 482
serum, 482
urine, 481
sepsis, 765, 766*f*
staging/severity assessment, 767–769

Biomarkers, (*Continued*)
systems biology approach, 131
emerging quantitative techniques, 134–136
in vitro measurement technology, 134–136
in vivo techniques, 136
nanotechnology, 134
organ-specific fingerprints, 133–134
proteomics, 135
validation
economic aspects, 238
Biorepositories *see* Biobanks
Biospecimen respiratory, nationwide, 171
Birth defects, copy number variation (CNVs), 194
Bisulfite DNA conversion, methylation analysis, 61, 61*f*, 62, 63, 64
Bladder cancer
serum protein antibody microarrays, 134
Bleomycin-induced pulmonary fibrosis, 617
Blood-based gene expression profiling, 3
Blood coagulation, 374, 374*f*, 375*f*, 765
genetics, 374–375
proteins, 374–375
phylogenetic analysis, 375, 375*f*, 376*f*
vitamin K-dependent, 375, 376
Blood lipids
genome-wide association studies, 279, 280*t*
screening, 272
variation, 271–272
Blood pressure, hypertension
diagnosis, 262–263
Blood samples, DNA methylation assessment, 64–65, 67
Bloodomics project, 385–386
Bmi1 (miR-128), 528
BMP4 gene, 470
Body mass index, data extraction from electronic medical records, 147–148, 148*t*
Body weight regulation, 662, 662*f*
genetic factors, 662, 662*f*
Boolean networks, 118
Bop, ventricular development regulation, 395
Borrelia burgdorferi, 407
cutaneous lymphomas, 401
BRAC1/BRAC2 testing, 1
BRAF
colorectal cancer, 65, 460, 467
Brain tumor stem cells, therapeutic targeting, 527
Brain tumors, 522–527
chromosomal alterations, 524–525
diagnosis, 524–528
epidemiology, 523–524
gene expression profiles, 526, 526*t*
genomic mapping, 524
microRNAs, 527
pharmacogenomics, 527
predisposition, 523–524
prognosis, 524–527
gene expression signatures, 525
molecular signatures, 137, 527
proteomics, 527
transcriptional alterations, 525–527
Brain-derived neurotrophic factor (BDNF)
antidepressant effects, 726
depression, 723
hepatic stellate cells, 649
Brain-derived neurotrophic factor receptor (trkB), 723
Branched-chain α-ketoacid dehydrogenase complex defect, newborn screening, 102
BRCA1
breast cancer, 196, 448
risk of second primary cancer, 191–192
DNA methylation status, 496
tumor treatment response prediction, 67
DNA sequence data, electronic medical records, 143–144
genetic testing, 164, 214
costs, 196
direct-to-consumer marketing, 214
family history criteria for referral, 215
ovarian cancer, 195, 493–494
aberrant methylation, 496
DNA repair defects, 494
BRCA1/2 mutations, 242
BRCA2
brain tumor predisposition, 523
breast cancer, 196, 214, 448
risk of second primary cancer, 191–192
DNA sequence data, electronic medical records, 143–144
genetic testing, 196, 213
costs, 196
direct-to-consumer marketing, 214, 214
family history criteria for referral
ovarian cancer, 196, 493, 494
DNA repair defects, 494
pancreatic cancer, 501
BRD2
see β_2-adrenergic receptor
Breast cancer, 191, 446–454
BRCA1/BRCA2, 196, 214, 448
management options, 448
risk of second primary cancer, 191–192
cell line microarrays, drug response investigations, 115–116
DNA methylation analysis, 67
estrogen/progesterone-receptor status, 448, 449
luminal A/B subgroups, 450
targeted hormonal therapy, 533, 534*f*
treatment responsiveness, 448
familial risk, 447–448
gene expression assay
see also MammaPrint; Oncotype Dx assay
gene expression profiles, 76, 78, 73, 75, 450, 451*f*, 452*t*, 542–543
cost-effectiveness, 236–237, 237*t*
distant recurrence, 452–453
Duke-Taipei study, 452
lymph node status, 452, 453*f*
Netherlands Cancer Institute study, 450
NSABP study, 452–453
pathway prediction, 453–454
prognostic applications, 80, 82, 114, 450, 452
risk stratification, 454
subtype differentiation, 450, 453*f*
tissue sampling, 75
treatment planning, 82, 210
genetic factors, 164, 448
genetic testing, 213–214, 448
access/reimbursement issues, 214–215
diagnostic-therapeutic combinations, 533
prognosis, 226
genomics applications, 454
Her2/neu positive, 449, 448, 449
diagnostic tests, 532, 533*t*, 535
pharmacogenomics, 183, 185
trastuzumab therapy, 183, 185, 194, 448, 532, 535
treatment approach, 165
hormonal status associations, 448
molecular basis, 448–449
prognosis, 80, 82, 114, 449
molecular signature analysis, 137
receptor negative tumor subgroups, 450
recurrence risk, 235, 236, 237
risk assessment, 447–448
screening, 447
systems medicine approach, 133
TAILORx, 210, 236
TNM staging, 449
treatment, 446
clinical trials, 447*t*
Brugada syndrome, 350, 360–362
clinical features, 360–361
genetics
clinical, 360–361
molecular, 357*t*, 360–361
mutations, 357*t*
therapy, 361
BTNL2 (butyrophilin-like 2), sarcoidosis, 619
Bucindolol, 319
Budesonide, 596
Bupropion, 610
Burkitt lymphoma, 415–416
Epstein–Barr virus association, 401
gene expression profile, 416

C

C4, systemic lupus erythematosis-related copy number variation, 53
C5
asthma, 594
gene polymorphism, 289
C-Path Institute, 230
C-reactive protein

acute coronary syndrome risk stratification, 308, 309*t*
chronic obstructive pulmonary disease, 606
complement activation, 745
Crohn's disease, 579–580
rheumatoid arthritis, 557
thrombosis gene linkage studies, 379
C-type lectin receptors (CLRs), 745, 747
CADASIL (cerebral autosomal dominant arteriopathy with subcortical infarctas and leucoencephalopathy), 687, 688*f*, 693–694
Cag A (*Helicobacter pylori* cytotoxin associated gene product), 633, 633*f*
peptic ulcer disease correlations, 633
CALBG trial, 485
Calcium channel antagonists
catecholaminergic polymorphic ventricular tachycardia (CPVT), 350
hypertrophic cardiomyopathy, 343
pharmacogenomics, 265
Calcium-dependent arrhythmia syndromes, 359*t*
Calcium-handling hypertrophic cardiomyopathy, 338–339
Calgranulin A *see* MRP8
Calgranulin B
prostate cancer urine biomarker, 481
Calsequestrin mutations, catecholaminergic polymorphic ventricular tachycardia (CPVT), 350
Canadian Agency for Drugs and Technologies in Health (CADTH), 235
Cancer, 94, 224
cost-effectiveness of genetic tests, 214
diagnostic-therapeutic combinations, 532–543
DNA methylation profiles, 65
early detection/diagnosis, 65, 67
prognosis, 67
treatment response predicition, 67–69
epigenetic gene silencing, 16, 496
gene expression profiles, 80
oncogene expression signatures, 117, 117*f*
prognostic applications, 78–79, 82, 114
subgroups identificaiton, 78, 167
treatment otpimization, 194
microarray experiments, 73
data analysis, 76–77
tumor tissue samples, 75
microRNAs (miRNAs)
expression profiling, 82
molecular signature analysis, 164
outcome predicition, 136
obesity association, 663
pharmacogenomics, 542–543
single cell analysis, DNA-encoded antibody libraries (DEAL), 136
spread *see* Metastasis
systems medicine approach to diagnosis, 133
targeted therapies, 532–543
antibody-based, 535–540, 536*t*
definitions, 532
ideal target, 532–533, 533*t*
small molecule-based, 542–543*t*
Cancer cells
microarray experiments, 78
single cell analysis, DNA-encoded antibody libraries (DEAL), 136
The Cancer Genome Atlas (TCGA), 35, 171, 528
glioblastoma, 525
Cancer Outlier Profile Analysis (COPA), 76, 77
Candida albicans infection, dendritic cell gene expression response, 767
Candidate genes, 691
asthma, 593–594
chronic obstructive pulmonary disease, 604, 605*t*, 608
cirrhosis, 646, 647*t*
colorectal cancer, 464–465
depression, 724–725
diabetes type 1, 676
diabetes type 2, 677
drug response biomarkers, 179–180
familial pulmonary fibrosis, 620–621
hypertension, 262
salt-sensitivity, 261
infectious disease susceptibility, 736, 748
inflammatory bowel disease, 576
multiple sclerosis, 567
obesity, 665*t*, 666*t*, 666
Parkinson's disease, 705
rheumatoid arthritis, 554
sarcoidosis, 617–618, 618*t*
sepsis, 762
Capecitabine, 469
CAPN10, diabetes type 2, 677
CAPSL, diabetes type 1, 680
CAPTURE, 309
Capture assay, hepatitis B DNA quantitation, 778
Carbamazepine, 186
metabolism, 727
Carcinogenesis
colorectal cancer, 457–458
adenoma to carcinoma sequence, 458–460, 459*f*
DNA methylation patterns, 61, 65
CARD8, *Salmonella* infection susceptibility, 764
CARD11, 416
CARD15, 583
CARD15/NOD2
Crohn's disease, 1045
phenotype relationship, 579
inflammatory bowel disease association, 576, 578
Cardiac calcium ion channel mutations, 359*t*
Cardiac ion channels
interacting protein gene mutations, long QT syndromes, 357, 357*t*
mutations
long QT syndrome ECG manifestation correlations, 359
Romano-Ward syndrome, 354–356
see also Cardiac potassium ion channel mutations; Cardiac sodium channel gene (SCN5A; *LQT3*) mutations
Cardiac potassium ion channel mutations, 357
Andersen-Tawil syndrome, 358
atrial fibrillation, 363–364
Jervell and Lange-Nielsen syndrome, 358
Romano-Ward syndrome, 354–355, 356
short QT interval syndrome, 361, 361*t*
Cardiac ryanodine receptor mutations *see* *RyR2* mutations
Cardiac sodium channel gene (SCN5A; *LQT3*) mutations, 353, 355–356, 357, 359, 360
atrial fibrillation, 363–364
Brugada syndrome, 360*f*, 361
gene-specific treatment approach, 360
Lev-Lenegre progressive cardiac conduction disease, 362
Romano-Ward syndrome, 353, 355–356
sinus node dysfunction/sick sinus syndrome, 362–363
sudden infant death syndrome, 352
Cardiac transplantation, 326–334
heart failure, 326, 337
immunosuppressive regimens, 327, 328*t*
monitoring, 328, 329*f*
rejection, 205
acute antibody-mediated, 327
acute cellular, 327, 328, 329*f*
chronic (cardiac allograft vasculopathy), 327
gene expression network studies, 331
gene expression signature test *see* AlloMap test
grading system, 328, 331*t*
molecular diagnosis in peripheral blood mononuclear cells (CARGO study), 322, 330, 331, 332, 333
prevention, 327
Cardiac troponin I (*TNNT13*) mutations, hypertrophic cardiomyopathy, 338
Cardiac troponin T (*TNNT2*) mutations, hypertrophic cardiomyopathy, 338
Cardiomyocytes
apoptosis in heart failure, 316–317, 318*f*
β_1-adrenergic receptor polymorphism, 319
Cardiomyopathy, microarray experiments, 78
Cardiovascular disease
chronic obstructive pulmonary disease comorbidity, 606
DNA methylation profiles, 65
genome-wide association studies, 279–280, 280*t*

Cardiovascular disease (*Continued*)
hypertension association, 264
metabolic profiling, 103
obesity association, 663
risk factors *see* Cardiovascular risk
see also Atherosclerosis
Cardiovascular dynamics, 259–260, 261*f*
Cardiovascular risk, 303–304, 378
cholesterol plasma level, 271
PCSK9 nonsense/missense variant effects, 294, 295*f*
chromosome 9p21 region of interest, 294, 304
clinical decision support using electronic medical records, 145, 145*f*
environmental factors, 295
family history, 289
genetic factors, 289–290, 304
genetic lipid disorders, 269, 278–279
high-density lipoprotein cholesterol, 274, 277
low-density lipoprotein cholesterol, 272, 274, 278
triglycerides, 277
genetic lipid variants, 279
hypertension, 264
screening, 295–296
CARE study, 296
CARGO (Cardiac Allograft Rejection Gene Expression Observational) study, 322, 328–330, 331, 332, 333
CARGO II, 334
Carmustine
oligodendroglioma, 524
tumor response prediction, 67
CART, obesity, 667
Case-control studies
bias, 177
computational analysis, 118
genome-wide association studies, 38–39, 146
control genotyping databases, 38
population stratification, 40
sample size, 39, 39*t*
infectious diseases susceptibility, 748
pharmacogenetics, 177–178, 179*f*
thrombosis, 379–380
Caspases, cardiomyocyte apoptosis, 316–317
Caspase-9, lung cancer, 434
Caspase inhibitors, cirrhosis, 648
Cataract, inherited forms, 715
CATCH22, 391
Catechol-O-methyltransferase *see* COMT
Catecholaminergic polymorphic ventricular tachycardia (CPVT), 350, 361–362
genetics, 361–362
management, 362
Cathepsin D
breast cancer, 533
glioblastoma, 527
Cathepsin L2, Oncotype Dx assay, 533
Ca(V)1.2 (L-type calcium channel) mutations, Timothy syndrome, 358
Caveolae, 356
cardiac ion channels, 357
Caveolin-3 (*LQT9*; CAV3), Romano-Ward syndrome, 356–357
Caveolins, 356
CBF (core binding factor), acute myeloid leukemia, 421
CBFB-MYH11 fusions gene, acute myeloid leukemia, 421
CCL1 *see* MCP-1
CCL3L1 copy number variation, HIV/AIDS, 53–54
CCL5 *see* RANTES
CCL16, hepatitis B infection, 651
CCND1 *see* Cyclin D1
CCND2, diffuse large B-cell lymphoma prognosis, 411
CCR2 mutation
human immunodeficiency virus (HIV) resistance, 736
CCNY, 583
CCR5
HIV infection/AIDS, 53–54, 762
CD4, 54, 552, 569, 590
CD4 T cells, 571
asthma, 590
hepatitis C, 778, 779
response to infection, 745
sarcoidosis, 617
CD8 T cells
chronic obstructive pulmonary disease, 606
hepatitis B, 778, 779
hepatitis C, 778, 779
sarcoidosis, 617
CD14, 747
host pathogen recognition/signaling, 764
Enterobacteriacea, 751
Staphylococcus aureus, 751
Streptococcus pneumoniae, 749, 750
promoter polymorphism
asthma, 593, 594
CD25, diabetes type 1, 680
CD36, host response to *Staphylococcus aureus*, 751
CD40 ligand, acute coronary syndromes risk stratification, 309
CD81, hepatitis C receptor, 776, 778
CD133, 528
CD209, alcoholic liver disease, 652
CD226, diabetes type 1, 680
Cdc7, hepatic stellate cells, 651
CDH1 gene, 470
CDH1 DNA methylation
breast cancer, 67
lung cancer, 434
CDH3 DNA methylation, breast cancer, 67
CDH13 DNA methylation, lung cancer, 434
CDK6, 528
CDKAL1, 583
diabetes type 2, 678
CDKN2A
cardiovascular risk, 294, 304
DNA methylation
lung cancer, 434, 435
protate cancer, 67
CDKN2B
cardiovascular risk, 294, 304
cDNA microarrays, 73, 74*f*, 75
disease diagnosis/classification, 78
multiple sclerosis immune response, 568
ovarian cancer, 495
cDNA-mediated anneling selection extension and ligation (DASL) assay, 81
CEACAM6, ovarian cancer, 495
CEBPA, acute myeloid leukemia, 423, 424
Cediranib (AZD2171), 528
Cell-based therapy
Parkinson's disease, 706–707
Cellular components, Gene Ontology (GO), 121
Center for Devices and Radiological Health (CDRH), 225
Centers for Medicare and Medicaid Services (CMS)
Clinical Laboratory Improvement Amendments (CLIA) program administration, 227
genomic medicine regulation, 223
Centre de recherche en droit public (CRDP), 249
Ceramides, targeted metabolic profiling, 102
CETP
blood lipid genome-wide association studies, 279
mutations, 276
variants, 276, 277, 279, 284
Cetuximab, 231, 539
CFTR (cystic fibrosis transmembrane conductance regulator) mutations, 27
diabetes type 1, 680
Char syndrome (*TFAP2B*), congenital heart disease, 393
CHD, 284, 297, 398–399
Chemical Genomics project, 210
Chemical individuality, 13
Chemokines
asthma, 590
host response to infection, 733
sarcoidosis, 617
Chemoprevention, lung cancer, 435
Chemotherapy-resistance, 441
ChemSpider database, 100
CHFR methylation status, tumor treatment response prediction, 67
Chitinase 3 like-1 (CH12L1) *see* YKL-40
CHMP2B mutations, frontotemporal dementia, 691
Cholangiocarcinoma, 654

Cholesterol
intestinal absorption, 270
lipoprotein transport, 269
metabolism, 270, 271
plasma level
cardiovascular risk, 271
screening, 272
Cholesterol absorption inhibitors, 281
Cholesteryl ester transfer protein (CETP), 271
deficiency, 276, 279
therapeutic inhibition, 282
see also CETP
CHRNA7, 57
Chromatin, 19
active/inactive
DNA methylation relationship, 60
epigenetic patterns, 19
Chromatin-modifying 2B protein (*CHMP2B*) mutations, frontotemporal dementia, 691
Chromosomal aberrations
brain tumors, 524–525, 525*t*
congenital heart disease, 394
leukemia, 421–422
lymphoma, 407, 408*t*
diffuse large B-cell, 409
obesity, 664
prostate cancer, 479–480, 479*f*, 483
see also Cytogenetics
Chromosomal syndromes, strabsimus, 713
Chromosome 4q25, 365
Chromosome 21, DNA sequencing by hybridization, 34
Chromosome Abnormality Database, 48
Chromosomes
gene distribution, 14–16
human genome, 14*t*, 15*f*
variation, 14*t*
study methods, 14
Chronic bronchitis, 603
see also Chronic obstructive pulmonary disease
Chronic disease
pharmacogenetic research targets, 176
see also Complex disease
Chronic hepatitis C (CHC) patients, 783
Chronic lymphocytic leukemia, 401
alemtuzumab treatment, 538
fludarabine treatment, 424–425
gene expression profiling, 424
prognostic factors, 423
ZAP-70 marker, 424, 426–427
Chronic myelogenous leukemia
imatinib mesylate, 194, 534
diagnostic test of patient eligibility, 532
targeted therapy, 532, 537
Philadelphia chromosome, 421
treatment, 422
pharmacogenomics, 424
Chronic obstructive bronchiolitis, 603
see also Chronic obstructive pulmonary disease
Chronic obstructive pulmonary disease, 603–613
clinical features, 610*f*
systemic, 606
corticosteroid resistance, 609, 611
diagnosis, 609–610, 610*f*
environmental factors, 603–604, 604*t*
exacerbations, 604, 606
treatment, 608
familial clustering, 604
genetic factors, 604, 604*t*, 605*t*
genomics, 603
inflamatory mechanisms, 606–609, 607*f*
asthma comparison, 607, 608*t*
macrophages, 606–608
mediators, 607
NK-κB activation, 607
management, 610
anti-smoking measures, 610
antibiotics, 610
bronchodilators, 610
corticosteroids, 611
lung volume reduction, 611
mucolytics, 611
oxygen, 611
pumonary rehabilitation, 611
natural history, 605*f*
pathophysiology, 604–606, 606*f*
emphysema, 605
pulmonary hypertension, 605
small airways, 605
pharmacogenomics, 611–612
prognosis, 610
proteomics, 608–609
screening, 609–610
CHST6, corneal disease, 713
Chylomicrons, 269
apolipoproteins, 270
metabolism, 270, 271
Cigarette smoke
carcinogenic components, 512
metabolising enzyme gene polymorphisms, 433
Cigarette smoking
asthma associations, 591
chronic obstructive pulmonary disease, 603, 604, 605*f*, 607
colorectal cancer risk, 464, 465
familial interstitial pneumonia, 616
gene expression profiles, 608
head and neck cancer risk, 511, 512
incidence in USA, 433
inflammatory bowel disease association, 576
lung cancer association, 432
at-risk smoker identification, 434–435
genetic influences, 432, 433–435
preventive approaches, 441
nicotine addiction-related genes, 604
pancreatic cancer risk, 502
rheumatoid arthritis association, 555
Cigarette smoking cessation
chronic obstructive pulmonary disease management, 603
gene expression profiles, 434
Cilengitide, 528
Cirrhosis, 645–655
causes, 647*t*
diagnosis, 646
genetics, 646–648
candidate genes, 646, 647*t*
hepatitis B, 778, 779
hepatitis C, 778, 779
liver structural changes, 649, 650*f*
progression to hepatocellular carcinoma, 645, 653
see also Liver disease
Cisplatin
head and neck cancer, 516
resistance, 516
tumor response prediction, 67
Citrate/isocitrate carrier (CIC), 102
CK19 expression, 507
Clarithromycin
Helicobacter pylori eradication, 628, 635
resistance, 636
pharmacogenomics, 636
Classification of disease, 120–121, 120*t*
Claudin-1, hepatitis C receptor, 776, 778
Claudin-4, pancreatic mucinous cystic neoplasms, 505
Claudins, alcoholic liver disease, 652
Clinical decision support
electronic medical records, 144–145, 145*f*
Clinical event prediction, 203
Clinical Laboratory Improvement Amendments (CLIA), 195, 227
laboratory certification, 223, 226, 227
Clinical laboratory services, 213
Clinical Molecular Genetics Society, genetic testing audits, 214
Clinical practice guidelines (CPGs), 144
Clinical trials
cost-effectiveness analysis, 234
diagnostic biomarkers/*in vitro* diagnostics, 226
drug-diagnostic test combinations, 229
genome-based, 167–168
patient randomization, 164
genotype-guided, 182
results disclosure to participants, 212–213
surrogate endpoints, 300*t*
Clinical utility, building evidence for, 7
Clinical utility of genomic tests, 196
Clopidogrel, 297
Cluster analysis, 118, 737
disease sub-type discovery, 119
Cocaine, 383*t*
Cohort studies, 81
pharmacogenetics, 179

COL8A2, corneal disease, 713
Collagens
  hepatic stellate cell production, 645, 649–650
Colorectal adenoma
  adenoma to carcinoma sequence, 458–460, 459*f*
    chromosomal instability pathway, 458–459, 460
    mismatch repair pathway, 458, 459–460, 467, 468
Colorectal cancer, 457–470
  bevacizumab treatment, 539
  biomarkers, 467
  candidate genes, 464–465
  causation, 457
  chemoprevention, 468–470
  chemotherapy, 467, 468
    genetic targets, 469, 469*t*
    response prediction, 469–470
  copy number variation (CNVs), 460, 465
  DNA methylation analysis, 65, 67, 458, 460
  early diagnosis, 457
  edrecolomab treatment, 541
  epidemiology, 457
  family history/familial forms, 215, 457, 461, 464, 466
    inherited monogenic syndromes, 457, 461–464, 462*t*
  gene expression profiles, 460, 468
  genetic predisposition, 457, 458*f*, 460–465
  genetic testing, 164, 465–467
  genome-wide association studies, 465
  genomic model, 458–460
  inflammatory bowel disease association
    ulcerative colitis, 575, 579
  K-ras mutations, 503, 505, 507
  metastasis, 458
  microsatellite instability, 459–460
  pharmacogenomics, 468–470
  plasma methylation markers, 67
  prognosis, 467–468
    molecular signatures, 137
  risk assessment, 465–467, 466*f*
  risk factors, 460–461, 464
  screening, 457, 467
  sporadic/isolated, 460, 464–465, 467
Colorectal polyps *see* Colorectal adenoma
CombiMatrix microarrays, 81
Common allele common disease hypothesis, 4
  colorectal cancer, 464
Common host response to infection, 737–738, 737*f*, 745*f*
Comparative genomic hybridization, 194, 410, 460
  colorectal cancer, 460
  diffuse large B-cell lymphoma, 411
  prostate cancer, 479
  see also Array comparative genomic hybridization
Complement, 745
Complete atrioventricular canal, 391
Complex disease, 201, 202*f*
  ancestral susceptibility alleles, 679
  association studies, 27–28
  clinical utility of genetic testing, 195–196
  copy number variation (CNVs), 54
  genome-wide association studies *see* Genome-wide association studies
  genomic approaches, 164–165
    data integration, 167, 167*f*
  metabolomics, 97
  pharmacogenetic research targets, 176
  predictive factors, 203–205
    healthcare approaches, 164–165
Computational analysis, 111–125
  methods, 118–119
    molecular signature classifiers, 137
Computerized tomography (CT)
  cirrhosis, 647
  Crohn's disease, 579–580
  head and neck cancer, 513
  heart failure, 318
  lung cancer screening, 434, 435
  pancreatic cancer screening, 502–503
  pancreatic mucinous cystic neoplasms, 505
  prostate cancer, 482
  see also Positron emission tomography-computerized tomography (PET-CT)
Computerized tomography (CT)-guided fine-needle aspiration, pancreatic cancer diagnosis, 503
COMT (catechol-O-methyltransferase), 706, 707
Confidentiality of data
  biobanking, 212
  public health genomics, 249
Confoundings factors
  case–control studies, 177
Congenital adrenal hyperplasia
  11β-hydroxylase deficiency, 263
  CYP11B1 mutations, 263
  hypertension, 263
  newborn screening, 193
Congenital fibrosis of extraocular muscles type 1 (CFEOM1), 713
Congenital heart disease, 390–398
  cytogenetic/molecular genetic testing, 396–397, 397*t*
  family history, 397
  gene discovery
    conventional genetics, 390–394
    large scale cDNA sequencing, 395
    microarray transcriptional profiling, 395, 396*f*
    mutagenesis and phenotypic screens, 395–396
    subtractive hybridization/differential display, 395
  genetic counseling, 397–398
  genetic factors, 390–391, 391*t*
    cell signaling genes, 393
    chromosomal aneuploidies, 394
    extracellular matrix protein genes, 393–394
    transcription factor genes, 391–393
  imaging investigations, 397
  medical evaluation approach, 397–398, 398*f*
  physical examination, 397
  right ventricular heart failure, 314
Congenital night blindness, rhodopsin mutations, 716
Connective tissue growth factor (CTGF), 607
Connectivity Map, 83
Connexins, early onset cataract, 715
Consent
  access to genetic information, 216
  biobanking, 212
  genetic testing, 196
  investigatory *in vitro* diagnostic studies, 226
  public health genomics, 249
Conserved sequences, 19
Consumer education, 196
Consumer health literacy, 152
Consumer online health information retrieval, 152–155
Consumer revolution, 6
Control groups, 179, 750
COPD, 612
Copy number variation (CNVs), 17, 22, 43, 46–55, 56, 194, 428, 708
  biallelic, 46, 47*f*
  causal genomic rearrangements, 46–47
    non-allelic homologous recombination, 46–47, 48, 48*f*
    non-homologous end joining, 47–48, 49*f*
  clinical cytogenetic diagnostics, 54, 55*f*
  colorectal cancer, 460, 465
  databases, 48, 50
  definition, 46
  detection, 50–52
    array-based comparative genomic hybridization, 50, 51*f*
    genotyping arrays, 42, 50–52
    whole-genome sequence comparisons, 52
  disease susceptibility associations, 52–54
  functional impact, 50
  generation, during DNA replication, 56
  genomic distribution, 48, 50, 51*f*
  HuRef genome analysis, 15*f*
  within individual and between identical twin pairs, 56
  local regulatory variants, 20
  multiallelic, 46, 47*f*
  neurodevelopmental disorders, 48, 52, 54
  new associations with, 56
  pharmacogenetics, 54
  prostate cancer biochemical relapse after surgery, 483

role, in schizophrenia and autism, 56
single nucleotide polymorphisms (SNPs) tagging, 54
Cornea, 712
genetic disorders, 713–715
Corneal dystrophies, 713, 714–715
Corneal keratitis, PAX6 mutation, 712
Cornelia de Lange syndrome, 713
Coronary artery disease, 289
genetic predisposition, 304
genome-wide association studies, 290–291, 290*t*
heart failure, 313, 314, 315*f*
metabolic profiling, 101
twin studies, 289
see also Acute coronary syndromes; Myocardial infarction
Corticosteroids
asthma management, 594–597
glucocorticoid resistance, 596
non-responsers, 596
cardiac transplantation rejection prevention, 327
AlloMap test performance, 332, 333*f*
chronic obstructive pulmonary disease, 607
unresponsiveness, 604, 608*f*
pharmacogenomics, 582, 597
ulcerative colitis, 579
Corticotrophin-releasing hormone, depression, 724
Cortisol metabolism, depression, 724, 728
Cost-benefit analysis, 234
Cost-consequence analysis, 234
Cost-effectiveness analysis, 233–235
genomic medicine, 233, 234*t*
genomic testing, 234*t*, 236
access/reimbursement issues, 214–215
pharmacogenetics, 224
resource allocation decision-making, 234–235
Cost issues, in genomic testing, 7
Cost-minimization, 234
Cost-utility analysis, 234, 235
Costs
genetic/genomic tests, 196
third-party payments, 196
genomic medicine, 167
pharmacogenetic test clinical utility, 183
Cowden syndrome
breast cancer, 448
PTEN mutations, 464
Cox Proportional Hazards Model, 119
Coxsackie B3 virus, idiopathic dilated cardiomyopathy, 314
CpG islands, 60
DNA methylation, 60
colorectal cancer, 65, 458, 460
ovarian cancer, 496
see also DNA methylation
CpG methylation-sensitive restriction enzymes, 61
Creatine kinase cardiac-specific isoenzyme, acute coronary syndrome diagnosis, 305
CREB (cAMP response element binding protein)
antidepressant effects, 727
CRH1/2, obesity, 665
Crohn's disease, 56, 57, 198, 575
anti-*Saccharomyces cerevisiae* antibodies (ASCA), 578, 579, 580, 581
CARD15/NOD2, 576, 578
phenotype relationship, 579
classification, 575, 579
clinical features, 579
colorectal cancer risk, 464
DEFB4 copy number variation, 52
diagnosis, 579–580
gene expression profiling, 78
genetic factors, 576, 577*t*, 581
disease course prediction, 581
genome-wide association studies, 582
incidence, 581
infliximab treatment, 581
pharmacogenomics, 582
peptic ulceration, 627
prognosis, 580–582
severity assessment, 579
surgical management, 580
twin studies, 576
see also Inflammatory bowel disease
Crystallins, early onset cataract, 715
CSRP3 mutations (LIM protein), hypertrophic cardiomyopathy, 338
CTLA4
diabetes type 1, 676, 677
rheumatoid arthritis, 554, 555
CTSE, ovarian cancer, 495
Cutaneous diffuse large B-cell lymphoma, 411
CX3CR1 polymorphism, asthma, 594
CX37 polymorphism, coronary artery disease, 290
CXCL1 (GROα), chronic obstructive pulmonary disease, 607
CXCL8, chronic obstructive pulmonary disease, 607
CXCL9 (MIG), hepatitis C infection, 651, 780
CXCL10 (IP10)
chronic obstructive pulmonary disease, 607
hepatitis C infection, 651, 780
CXCR4
pancreatic mucinous cystic neoplasms, 505
Cyclic citrullinated peptide antibodies, rheumatoid arthritis, 552, 556
Cyclin D1 (CCND1)
gene expression profiling in cancer progression, 80
head and neck cancer, 511, 515, 517
mantle cell lymphoma, 415
Cyclin-dependent kinase inhibitor genes, cardiovascular risk, 294, 304
Cyclin-H, hepatitis B infection, 651
Cyclooxygenase pathway, ovarian cancer metastasis, 497
Cyclooxygenase-2 (Cox-2)
chronic obstructive pulmonary disease, 606
head and neck cancer, 517
Cyclophosphamide
pharmacogenomics, 542–543
Cyclosporine, 327
CYP gene variant387
CYP1A1
promoter methylation, prostate cancer, 478
CYP1A2 polymorphism
antidepressants pharmacogenomics, 726
colorectal cancer risk, 465
CYP2C9, 186
CYP2C9 polymorphism
non-steroidal anti-inflammatory agent (NSAID)-induced peptic ulcer disease, 636
pharmacogenomics
antidepressants, 637
phenytoin, 181
warfarin response variability, 180, 183, 385, 386*t*
CYP2C19 genetic variants, 187, 387
CYP2C19 genotype, 637
CYP2C19 polymorphism
AmpliChip test, 185, 195
pharmacogenomics
proton pump inhibitors, 628
CYP2D6 polymorphism
AmpliChip test, 185, 195
copy number variation (CNVs), 54
pharmacogenomics
antidepressants, 637
atomoxitine, 230
healthy volunteer studies, 179
paroxetine, 181
CYP3A4 polymorphism
pharmacogenomics
antidepressants, 637
clarithromycin, 628
CYP11B1 mutations, congenital adrenal hyperplasia, 263
CYPs (cytochrome P450 isoenzymes), 180
AmpliChip, 185, 195
antidepressants metabolism, 726, 726*t*
cigarette smoke metabolism, 433
pharmacogenomics
Helicobacter pylori eradication, 628
Cystatin A, head and neck cancer, 517
Cystic fibrosis, 27, 246
newborn screening, 193
pancreatic cancer association, 502
see also CFTR (cystic fibrosis transmembrane conductance regulator) mutations

Cystic fibrosis transmembrane conductance regulator *see* *CFTR* mutations
Cytochrome oxidase subunit I, prostate cancer association, 478
Cytochrome P450s
  see also CYPs
Cytogenetic assays
  copy number variation (CNVs), 53*f*, 54
Cytogenetics
  acute lymphoblastic leukemia, 421, 426
  acute myeloid leukemia, 421, 426
  leukemia, 421–422, 422*t*
  lymphoma, 408*t*
  see also Chromosomal aberrations
Cytokines
  asthma, 590
  infectious disease response, 733, 747
    common host response, 737, 738
    Gram-positive organisms, 748
    polymorphisms, 764–765
  rheumatoid arthritis proteomics, 557
  sarcoidosis, 617
Cytomegalovirus
  cardiac transplantation rejection AlloMap test performance, 332
  susceptibility genes, 762
Cytotoxic T cells, 607*f*, 736

D
Dacarbazine
  tumor response prediction, 67
Daclizumab, 538
Danon's syndrome, hypertrophic cardiomyopathy, 339
DAPK1, promoter hypermethylation in lung cancer, 434
DASL 6000-gene expression profiling method, 656
Dasatinib, 545
Data access, 169
  policy issues, 211–213
  public health genomics, 249
  see also Confidentiality of data
Data acquisition, 167
Data analysis
  microarray information, 76–77
    mining, 76–77
    supervised learning, 76, 77*f*
    unsupervised learning, 76, 77*f*
Data repositories, 119–120
  search problems, 120
Data standards, 167
Data-sharing policies, 213
Data-withholding practices, 213
Database of Genomic Variants, 48
Databases
  access policies, 213
Daunorubicin pharmacogenomics, 426
DC-SIGN, host response to mycobacteria, 754
DCC (deleted in colorectal carcinoma), 458, 468
DDX5 (DEAD-box polypeptide-5)
  cirrhosis, 343
  hepatitis C proteomics, 780
De novo CNVs, 56
Death-associated protein kinase (DAP-kinase), head and neck cancer monitoring, 516
DECIPHER, 48, 55
Decision analysis, economic evaluation, 234
Decision trees, 118
Decitabine, 422
deCode Genetics Inc., 145, 212
deCODEme, 198
Deep brain stimulation, Parkinson's disease, 706
Deep vein thrombosis, 382
DEFB genes, 56
DEFB4, Crohn's disease-related copy number variation, 52
Deleted in colorectal carcinoma (DCC), 458, 468
Deletions, 14, 17
  see also Insertion/deletion (in/del) polymorphism
Delta hepatitis *see* Hepatitis D
Delta-like protein 3 (DLL3), astrocytomas, 526
Dementia, 687
  antemortem diagnosis, 693, 694
  causes, 687, 688*f*
  clinical approach, 694
  genetic testing, 694
  incidence, 688, 688*f*
  primary, 687, 689–694
    pathophysiology, 688
  secondary, 687
Dendritic cells
  innate immune response, 745, 763
  pathogen-specific responses, 733, 738
  response to infection, gene expression profiles, 737, 767
Depression, 722–728
  anxiety comorbidity, 723
  causes, 723, 723*f*
  clinical course, 722
  diagnosis, 722–723
  environmental factors, 724
    5-HTT gene interactions, 725
  gene expression studies, 727
  genetic factors, 724
    association studies, 724–725, 725*t*
    emotional regulation/stress response, 725–726
    linkage analysis, 724
  hypothalamus-pituitary-adrenal axis involvement, 724
  neurobiology, 723–724, 723*f*
    signal transduction, 723
    stress response system, 724
  pharmacogenomics, 726–727
    cytochrome P450 polymorphisms, 726–727, 726*t*
  prevalence, 723
  proteomics, 727
  response to antidepressants, 727, 727*t*
DIA1, 56
Diabetes, 156, 191, 676–681
  epidemiology, 676
  family history, 678
  pancreatic cancer risk, 502
Diabetes Heart Study, 290
Diabetes type 1, 676
  applications of genetic research, 681
  candidate gene studies, 676
  gene expression profile comparison with other autoimmune disorders, 569
  gene-environmental factor interactions, 681
  genetic risk, 676, 677, 681
  genome-wide association studies, 677, 680–681
  linkage studies, 676–677
  permanent neonatal diabetes, 681
  prospective health care, 205
Diabetes type 2, 27, 676
  applications of genetic research, 682
  candidate gene studies, 677
  cardiovascular risk, chromosome 9p21 region of interest, 294
  epidemiology, 676
  future genetic research, 679–680
  gene expression data analysis, 76
  genetic risk, 677, 678
  genome-wide association studies, 677–680, 678*t*
    blood lipids, 279, 280*t*
  obesity association, 662, 676, 678
Diagnostic applications, 112–113, 164–165
Diagnostic markers *see* Biomarkers
Diagnostic tests
  clinical utility, 196
  clinical validation, 196, 246
  costs, 196
    third-party reimbursement, 196
    value-based reimbursement, 239–240
  economic incentives for development, 237–239
  laboratory standards, 194–195
  performance
    predictive values, 182
    quality control, 182
    sensitivity, 182
    specificity, 182
  pharmacogenomics, 182–183
    drug combinations development, 228–230, 238
    regulatory issues, 229
  policy issues, 211
  pre-analytical variation effects, 182
  public health genomics, 248, 249
  regulatory issues, 223–224, 225, 238

results communication, 196
targeted therapies, 532
translational challenges, 211
see also In vitro diagnostics
Dietary factors
cardiovascular risk, 295
colorectal cancer, 464, 465
preventive strategies, 468
head and neck cancer, 512
obesity, 662
pancreatic cancer, 502
see also Nutrition
Differential display, 73
Differential hybridization, 73
Differential methylation hybridization (DMH), 62
Diffuse large B-cell lymphoma, 410–411
activated B-cell subtype, 410
chromosomal aberrations, 410
diagnosis, 410
gene expression profiles
prognostic applications, 78, 80, 410–411
skin locations, 411
subtypes discrimination, 114, 114*f*, 167, 411
therapeutic target identification, 411
germinal center B-cell subtype, 410
histogenesis, 410
International Prognostic Index (IPI), 410
Diffuse Lewy body dementia, 687, 692–693, 700
Lewy bodies (α-synuclein aggregates), 687, 692, 693*f*
Diffuse parenchymal lung disease, 619–620
classification, 616
genetic factors, 617–618
inbred mouse studies, 617
surfactant protein C mutations, 619–620
genomics, 620
microarray studies, 620–623
pathogenesis, 620
sarcoidosis *see* Sarcoidosis
surfactant proteins, 619–620
see also Familial interstitial pneumonia
DiGeorge syndrome, 395, 396
congenital heart disease, 391–392
Dihydropyrimidine dehydrogenase, pharmacogenomics, 469
5-fluorouracil, 469, 516
Dilated cardiomyopathy, 313, 337, 341
alpha B-crystallin, 715
Diltiazem, hypertrophic cardiomyopathy, 343
Dipeptidylpeptidase 10 *see* DPP10
Direct-to-consumer (DTC) genetic tests, 187
Direct-to-consumer genetic test marketing, 214
BRCA1 testing, 213
policy issues, 214
Discrimination, genetic
policy issues, 209
public health genomics, 245
Disease, 4
Disease mechanisms, 613
Disease sub-type discovery, bioinformatics, 118–119
Disease tracking, 203, 205
predictive models, 203
Disease-modifying anti-rheumatic drugs (DMARDS)
rheumatoid arthritis, 558
Disopyramide, hypertrophic cardiomyopathy, 343
Diuretics, pharmacogenomics, 265
DJ-1 (*PARK7*), Parkinson's disease, 704
DLBCL, RT-PCR gene expression profiling, 82
DLG5, inflammatory bowel disease, 578
corticosteroid pharmacogenomics, 582
DNA methylation, 20, 68–69, 62*f*, 63–171, 729
assessment technology, 61–65, 61*t*
clinical applications, 65
methylation "content"/"pattern"/"level", 62, 63*f*
methylation-sensitive restriction enzymes, 61, 62
methylation-specific antibodies/proteins, 61–62
methylation-specific bisulfite modification, 61, 62, 63, 64
screening sample analysis, 64–65
tissue sample anaylsis, 62–64
chromatin structure correlations, 60
colorectal cancer, 458, 460
screening/surveillance, 467
disease
associations, 65, 66*t*
early detection/diagnosis, 65, 67
prognosis, 67
treatment response prediction, 67–68
genome-wide epigenetic analysis, 62, 63*t*
head and neck cancer monitoring, 516
lung cancer early diagnosis, 434–435
marker discovery methods, 62
ovarian cancer, 493
pancreatic cancer screening, 502–503
pattern, 62
prostate cancer, 478
DNA methyltransferase 1 (DNMT1), 60
DNA methyltransferase 3b (DNMT3b), 60
gene mutations, 60
DNA methyltransferases (DNMTs), 60
de novo activities, 60
maintenance activities, 60
DNA repair enzyme genes
defects
brain tumors, 523
head and neck cancer, 512
ovarian cancer, 494, 495
methylation status, 68
polymorphisms
pancreatic cancer diagnosis, 503–504
DNA sequencing, 28–34
by synthesis, 33–34
congenital heart disease gene discovery, 395
copy number variation (CNVs) detection, 50–52
electronic medical records, 143–144
future goals, 35
hybridization on microarrays, 34
methodologies, 28, 30
PCR amplicon re-sequencing, 31–33
polymorphisms detection, 28
whole genome shotgun method *see* Whole genome shotgun sequencing
see also Comparative sequence analysis; Genome sequencing; Viral sequence determination
DNA-encoded antibody libraries (DEAL), blood biomarker analysis, 136
DNMT3B mutations, ICF disorder, 60
Docetaxel, prostate cancer, 484–485
Domains, 16
Dopa decarboxylase, 707
Dopamine metabolism
Parkinson's disease therapeutics, 706, 706*f*
pharmacogenomics, 707
Down syndrome *see* Trisomy 21
Doxorubicin, pharmacogenomics, 450
DPC4/SMAD4, pancreatic cancer, 501
DPP10 (dipeptidylpeptidase 10)
asthma, 593
chronic obstructive pulmonary disease, 607
DRD2 *see* Dopamine-2 receptor
DRD4 *see* Dopamine-4 receptor
Drug development, 210, 227–228
economics, 235, 238
incentives, 237–239
monoclonal antibody anti-cancer agents, 535
pharmacogenomics, 224, 238
applications, 183
drug-diagnostic test combinations, 228–229, 238
regulatory issues, 224, 227–230, 238
Voluntary Genomic Data Submission, 228
Drug discovery
gene expression profiling, 87
genomics applications, 169–170
target identification, 168
pharmacogenomics, 168
process, 168
Drug pricing, 238
Drug response biomarkers, 168
candidate gene approaches, 179–180
whole-genome association studies, 179–180
Drug response pathways, pharmacogenetic studies, 181
Drug therapy
bioinformatics applications, 114–115
anti-cancer agent susceptibility, 114–115

Drug therapy (*Continued*)
cancer cell line drug responses, 115–116, 116*f*
side-effects prediction, 115
Drug-eluting intracoronary stents, 297
Drug-induced long QT syndromes, 352
Drug-metabolizing enzymes, 235
cigarette smoke metabolism, 433
genetic tests
clinical acceptance, 228
drug labels, 224, 227
regulatory issues, 224, 225
pharmacogenetic studies, 180
DTC genomics, 156
Duane's syndrome, *SALL4* mutations, 713
Duodenal ulcer, 627, 627*f*
susceptibility, 631, 636
see also Peptic ulcer disease
Duplications, 14
segmental, 17
Dyskeratosis congenital, pulmonary fibrosis, 620

E

E2A-PBX1, acute lymphoblastic leukemia, 424
E40K variant, 283
E-cadherin, 496
E-selectin, 293
chronic obstructive pulmonary disease, 607
Early disease detection, 164–165
colorectal cancer, 457
lung cancer, 434–435
population genomic screening, 193
Echocardiography
cardiac allograft rejection monitoring, 328
heart failure, 318
hypertrophic cardiomyopathy, 337, 341
screening, 342*f*, 343
septal morphology, 337, 337*t*, 338*f*, 341
Economic evaluation, 233–235
genomic medicine, 233, 235–239
genomic technologies, 235–237
breast cancer, 236–237
cost-effectiveness framework, 234*t*, 236
disease risk testing, 235
pharmacogenomics, 235
methodological approaches, 234
Edrecolomab, 539
Education
genomic literacy, 197–198
health professionals, 196
competencies, 209
policy issues, 214, 215
EGFR, 528
see also Epidermal growth factor receptor
EGFR pharmacogenomic testing, 242
Egr-1, chronic obstructive pulmonary disease, 608
Eicosanoids, targeted metabolic profiling, 102
ELAC2, prostate cancer, 477
Elastin
hepatic stellate cell production, 650
ELASTIN (Williams syndrome), congenital heart disease, 394
Electrocardiogram
acute coronary syndrome diagnosis, 305
Brugada syndrome, 360, 361, 361*f*
catecholaminergic polymorphic ventricular tachycardia (CPVT), 361–362
heart failure, 317
hypertrophic cardiomyopathy screening, 343
Lev-Lenegre progressive cardiac conduction disease, 363*f*
long QT syndromes, 350, 351*f*, 353*f*
cardiac ion channel mutation correlations, 359
Electron transfer dissociation (ETD), 95
Electronic medical records, 142–149
biorespositories, 145, 146
clinical decision support, 144–145, 145*f*
data sources, 143
family history, 144
genomic data integration, 143–144, 144*t*
genomic medicine, 143–144, 144*t*
genomic research applications, 145–149
data extraction, 147–148
genome-wide association studies, 145–146
patient information, 143
Electrophoresis *see* Gel electrophoresis, two-dimensional; Polyacrylamide gel electrophoresis, two-dimensional (2D-PAGE)
Electrospray ionization-tandem mass spectrometry, metabolic profiling, 98, 100
ELISA, 135
ELMOD2, familial interstitial pneumonia, 623
eMERGE network, 149
EMR data, 148
Emotional behavior, 725
Emphysema, 605
pathophysiology, 605
surgical lung volume reduction, 611
see also Chronic obstructive pulmonary disease
Employment discrimination, 165–166, 216
ENaC (epithelial sodium channel) mutations, Liddle's syndrome, 263
ENCODE (Encyclopedia of DNA Elements) Project, 63*t*
Endometrial cancer, 463
Endomyocardial biopsy
cardiac allograft rejection
AlloMap test score relationship, 331–332
diagnosis, 326
grading system, 331*t*
interobserver variability, 331
molecular score discordance, 331–332
monitoring, 328
gene chip analysis, 320–321
heart failure, 318
Endoscopic retrograde cholangiopancreatography, pancreatic cancer
diagnosis, 503
screening, 502–503
Endoscopic ultrasound
pancreatic cancer screening, 502–503
pancreatic cystic neoplasm monitoring, 505
Endoscopic ultrasound-guided fine-needle aspiration
pancreatic cancer diagnosis, 503–505
combined telomerase assay, 505
microdissection-based genotyping, 504
pancreatic mucinous cystic neoplasms, 502
Endothelial cells
activation, 378
phenotypic variation, 378
prothrombotic stimuli response, 378, 379*f*
response to infection, gene expression profiles, 737
vascular bed-specific thrombosis, 378
Endothelial lipase, 271
Endothelin, antihypertensive agent pharmacogenomics, 265
Endothelin-1
hepatic stellate cells, 650
Endotoxin exposure, 763
see also Lipopolysaccharide
Energy expenditure regulation, genetic factors, 667
Enhancers, 19
ultraconserved elements, 19
Enolase, laryngeal carcinoma, 517
eNOS polymorphism *see* Nitric oxide synthase
ENPR1, obesity, 666
Entactin, hepatic stellate cell production, 650
Entecavir, hepatitis B treatment, 781
Enterobacteriacea, genetics of host response, 751–752
Entrez Gene, 124
Environment–gene interactions
cardiovascular risk, 295
chronic obstructive pulmonary disease, 603–604
colorectal cancer, 457
depression, 724
5-HTT gene polymorphism, 726
heart failure, 314
hemostasis, 377–378
hypertension, 259, 260, 260*f*
inflammatory bowel disease, 576
obesity, 661, 662, 663
Parkinson's disease, 700, 702, 704
population-based studies
susceptibility screening, 193
public health genomics, 246, 247*f*, 250
rheumatoid arthritis, 555

systems biology approach, 132
viral hepatitis, 776
Environmental factors *see* Environment–gene interactions
Enzyme immunoassay
hepatitis C diagnosis, 777–778
see also ELISA
Eosinophilic granuloma, 616
Ep-CAM (17-11A), monoclonal antibody targeted therapy, 539
EPHX1 gene, 612
EpiCare system, 143
Epidermal growth factor receptor (EGFR)
glioblastoma, 525, 526
gene expression profiles, 526
gliomas, 522, 523
immunotherapy, 527
head and neck cancer, 511, 515
cytotoxic agent resistance, 516
relapse prognosis, 516
therapeutic targeting, 517, 518
lung cancer, 438, 439*f*
therapeutic targeting, 438–439, 440
monoclonal antibody targeted therapy, 538
ovarian cancer, 495
prostate cancer circulating tumor cells, 482
Epidermal growth factor receptor inhibitors, 438–439, 440, 518
brain tumor pharmacogenomics, 527
small molecules, 532
erlotinib, 542
gefitinib, 542
Epigenetics, 2, 68
definition, 2
ovarian cancer, 495
prostate cancer, 478–479
Epigenome
definition, 2
Epigenomics
definition, 2
Epistatic associations, pharmacogenetic studies, 181
Epithelial sodium channel (ENaC) mutations, Liddle's syndrome, 263
Epstein–Barr virus
Burkitt lymphoma, 401, 839
lymphomas association, 401
multiple sclerosis association, 570
Equilibrium radionuclide angiograpy, heart failure diagnosis, 318
Erbitux, 231, 545
ERBB1 *see* Epidermal growth factor receptor
ERBB2 *see* Her2/neu
ERBB3, ovarian cancer, 495
ERG, rearrangements in prostate cancer, 77
ERG-TMPRSS2 fusion, prostate cancer, 483
Erlotinib, 542
head and neck cancer, 518
Erythroplakia, 512
Escherichia coli
genome mapping, 744
host gene expression in response to infection, 748, 767
outer membrane porin C antibodies, inflammatory bowel disease, 578, 579
Esotropia, 713
ESR1/2 methylation status, tumor treatment response prediction, 67
Essential hypertension *see* Hypertension
Essential myosin light chain (*MYL3*) mutations, hypertrophic cardiomyopathy, 338
Estramustine, 533
Estrogen receptor
breast cancer targeted hormonal therapy, 533, 534*f*
Oncotype Dx assay, 533
Ethical issues, 197, 198
genomics research policy, 210
pharmacogenetic tests, 183
public health genomics, 246, 249
research results disclosure to participants, 213
Ethnic/racial variation
breast cancer, 447
cardiovascular risk, 294
chronic obstructive pulmonary disease, 604
colorectal cancer, 457
genomics research analytical variables, 210
policy issues, 211–213
hypertension, 262
nephrosclerosis association, 264
lung cancer, 433
multiple sclerosis, 567
pharmacogenomics and heart failure management, 175, 319
sarcoidosis, 617
Ethylnitrosourea mutagenesis, 734
ETS translocations, prostate cancer, 483
ETV1, rearrangements, prostate cancer, 77
ETV1-TMPRSS2 fusion, prostate cancer, 483
Europe
genomics research policy, 210
European Agency for the Evaluation of Medicinal Products (EMEA), pharmacogenomics information sharing, 228
European Bioinformatics Institute (AMBL-EBI) Array Express repository, 119, 120
EV11, acute myeloid leukemia, 424
Evaluation of genomic applications in practice and prevention (EGAPP), 171–172, 249, 252
Everolimus *see* RAD001
Evidence-based guidelines, genomics translation into medical practice, 164
EWS, hepatitis C, 654
Extracellular Matrix Protein 1 (ECM1), 583
EXP, immunomodulatory vaccines, 570
Experimental autoimmune encephalomyelitis, 568, 569
Experimental design pharmacogenetics, 177–179, 178*f*
Extracellular matrix
cirrhosis, therapeutic targets, 648
liver fibrosis, 649
remodeling, 645
protein genes, congenital heart disease, 393–394
Extraocular muscles, genetic disorders, 713
Eye
genetic disorders, 714*t*
movements, 713
structure, 712, 712*f*
Ezetimibe, pharmacogenomics, 281
EZH2
prostate cancer metastasis, 484

F

FABP, obesity, 666
FABP7, glioblastomas, 527
Fabry's disease, hypertrophic cardiomyopathy, 339
Factor II *see* Prothrombin
Factor V, 375
polymorphism, effects on protein C pathway, 765
sepsis, 765
Factor V Leiden
meningococcal disease susceptibility, 765
patient information needs, 381
targeted screening, 384*t*
thromboembolism risk, 195
Factor VII, 375, 376
genetic variation, 376
thrombosis-associated variants, 380
Factor VIII, 375, 376
genetic variation, 376
Factor IX, 375, 376
Factor X, 375, 376
thrombotic event prediction with anticardiolipin antibodies, 380
Factor XII, 379
genetic variation, 376
Factor XIII, genetic variation, 376, 377*t*
False discovery rate, 118
False negative result
pharmacogenetic tests, 182
False positive rate, 118
False positive result
pharmacogenetics
case–control studies, 178
tests, 182
Familial adenomatous polyposis, 1, 457, 461, 463
allelic variants, 461
APC gene mutations, 461
APC mutation-negative, 461, 463
clinical features, 461
colorectal cancer

Familial adenomatous polyposis, (*Continued*)
chemoprevention, 468
screening, 467
extra-gastrointestinal manifestations, 461
genetic testing, 213, 465, 466
pancreatic cancer association, 502
Familial atypical multiple mole melanoma, pancreatic cancer association, 502
Familial breast and ovarian cancer, pancreatic cancer association, 502
Familial chylomicronemia syndrome, 277, 279
Familial combined hyperlipidemia, 278
Familial Creutzfeldt–Jakob disease, 693, 695
Familial defective apolipoprotein B-100, 272
Familial dysbetalipoproteinemia *see* Hyperlipoproteinemia type III
Familial hyperalphalipoproteinemia, 278
Familial hypercholesterolemia, 272, 278, 282
novel therapeutic targets, 281, 282
prenatal diagnosis, 282
Familial hypertriglyceridemia, 278
Familial hypoalphalipoproteinemia, 278
Familial hypobetalipoproteinemia, 274, 282
Familial hypocalciuric hypercalcemia, pulmonary fibrosis, 617, 620
Familial interstitial pneumonia, 616
clinical features, 620
genetic determinants, 622–623
genome-wide linkage analysis, 623
microarray studies, 620–622
Familial study design, pharmacogenetics, 178–179
Familial surfactant protein C mutation, pulmonary fibrosis, 617
Familial ventricular tachycardia *see* Catecholaminergic polymorphic ventricular tachycardia (CPVT)
Family history
adenomatous polyps, 464
Alzheimer's disease, 689
asthma, 591
breast cancer, 447–448
cardiovascular risk, 289
chronic obstructive pulmonary disease, 604
colorectal cancer, 457, 461, 464, 466
congenital heart disease, 397
electronic health records, 144
lung cancer, 433
sarcoidosis, 617
thromboembolism, 380–381
Family studies
Alzheimer's disease, late onset, 689
infectious diseases susceptibility, 735
inflammatory bowel disease, 576
rheumatoid arthritis, 552
sepsis outcome, 762
FANCF methylation status, tumor treatment response prediction, 67
Fanconi anemia, 512
head and neck cancer risk, 512
Farnesyltransferase inhibitors, leukemia, 422
Fasting-induced adipose factor (FIAF), 101
Fat-soluble vitamins, intestinal absorption, 270
Fatal familial insomnia, 693
Fatty acids
metabolism, 270
FBX07 gene, 708
FCERIB, asthma, 593
Fcγ receptor mutations, host response to *Neisseria meningitidis*, 752
FcγRIIA, host response to *Streptococcus pneumoniae*, 750
FcγRIII
infliximab pharmacogenomics, 582
Fcgr3, glomerulonephritis-related copy number variation, 52
FDA *see* Food and Drug Administration
Fetal dopaminergic neuron transplantation, Parkinson's disease, 706–707
FGF20 gene, 708
FGFR1 mutations, glioblastoma, 33
FHIT, methylation in lung pre-neoplastic lesions, 434
Fibrates (PPARα agonists), 281
pharmacogenomics, 281
FIBRILLIN-1 (Marfan syndrome), congenital heart disease, 393–394
Fibrin, 378–379
variation, 377
Fibrinogen, 377
sepsis, 765
variants, 376, 377, 377*t*
hypertension-related stroke, 263
thrombosis, 380
Fibroblast gene expression profiles, 77–78
tissue of origin differences, 116
Fibronectin
hepatic stellate cell production, 649–650
FIF21A, strabsimus, 713
FKBP5, antidepressant response variation, 727
Flecainide, gene-specific long QT syndrome treatment, 360
Flow cytometry, 82
FLT3
acute myeloid leukemia, 424
prognosis, 423
targeted therapy, 542
leukemias, 421
FLT3 inhibitors, leukemia treatment, 422
Fludarabine, pharmacogenomics, 424
Flunisolide, asthma management, 596
Fluorescence resonance energy transfer (FRET) dyes, 29
Fluorescence spectroscopy, oral cancer screening, 512–513
Fluorescence-activated cell sorting, 75
cell response analysis in multiple sclerosis, 569
Fluorescence *in situ* hybridization (FISH)
congenital heart disease, 396
5-Fluorouracil, 83, 516
colorectal cancer, 468, 469
head and neck cancer, 516
pharmacogenomics, 469
treatment response protein markers, 469–470
Fluticasone, asthma management, 596
FMS, ovarian cancer, 495
FN1, diffuse large B-cell lymphoma prognosis, 411
Folate intake, 465
Folate metabolism
colorectal cancer studies, 465
Follicular lymphoma, 413
diagnosis, 413–414
gene expression profiles, 413
prognosis, 414–415
transformation, 414
treatment, 413
Food and Drug Administration (FDA), 230–231
analyte-specific reagent (ASR) Guidance document, 227
approved diagnostic platforms, 195
genetic testing regulation, 213
laboratory-developed tests, 226, 227
genomic medicine regulation, 223
Guidance documents, 223
in vitro diagnostics regulation, 225
Guidance documents, 225, 226
Modernization Act, 225
pharmacogenetics
drug development, 228
drug labels, 230
test policy, 183, 184*t*
Voluntary Genomic Data Submission, 228
Formalin-fixed paraffin-embedded samples
DNA methylation assessment, 62
gene expression profiling, 81
genotyping technologies, 43
Formoterol, chronic obstructive pulmonary disease, 610
Fos, head and neck cancer, 517
FOSB, prostate cancer, 484
Foundations of genomic and personalized medicine, 1, 5–6
Fourier transform ion cyclotron mass spectrometers (FT-ICR-MS), 100
FOXC1, ocular development, 712
Fra-2, head and neck cancer, 517
Fragile X syndrome, 143
newborn screening, 193
Framingham Heart Study, 377, 379
Framingham risk score, 303
Frataxin (*FXN*) mutations, hypertrophic cardiomyopathy, 339
Friedrich ataxia (*FXN* mutations), hypertrophic cardiomyopathy, 339
Frontotemporal dementia, 687, 691, 706
clinical heterogeneity, 691

gene mutations, 691, 692*f*
inheritance patterns, 691
FTO
diabetes type 2, 678
Functional genomics, 81
Fungal infection
pathogen recognition, 747
sepsis, 760
FUSION study, 678, 679
Future, of personalized medical care, 7–8
FXN (frataxin) mutations, hypertrophic cardiomyopathy, 339

G

G3139, 543
G-proteins, antihypertensive agent pharmacogenomics, 265
GAB2 haplotype, 695
Galactosemia
newborn screening, 193
GATA-4 mutation, 398, 399
Galectin-2
gene polymorphism, myocardial infarction association, 293, 296
Gardner syndrome, 461
Garrod, Archibald, 13
Gas chromatography-mass spectrometry (GC-MS), metabolic profiling, 98, 100, 102
Gastric cancer, 461, 631, 634
Gastric lymphoma, *Helicobacter pylori* association, 401
Gastric ulcer, 627, 627*f*
susceptibility, 635, 636
see also Peptic ulcer disease
Gastrin, 628
Gastrinoma
peptic ulceration, 627
see also Zollinger-Ellison syndrome
Gastrointestinal stromal tumors, 194, 532
imatinib mesylate, 540, 542
sunitinib, 542
GATA2, coronary artery disease, 292
GATA4, congenital heart disease, 391, 392
Gaucher disease
pulmonary fibrosis, 617, 620
Gefitinib, 426, 542
lung cancer, 542
small-cell, 194
Geisinger Clinic MyCode biorespository, 146–147, 147*f*
Gel electrophoresis, two-dimensional
proteomics
acute coronary syndrome diagnosis, 308
brain tumors, 527
multiple sclerosis, 570
pancreatic cancer diagnosis, 504
see also Polyacrylamide gel electrophoresis, two-dimensional (2D-PAGE)
Gemtuzumab ozogamicin, 423, 538
acute myeloid leukemia, 538
GenBank
liver proteome, 649
Gene expression, 19, 612–613
allelic imbalance, 20
data analysis, 76–77
mining, 76–77
supervised learning, 76, 77*f*
unsupervised learning, 76, 77*f*
data repositories, 119–120
genetic control of level, 20
genomic/epigenomic aspects, 19–20
see also Gene expression profiles; Transcriptomics
Gene Expression Omnibus (GEO), 119, 119*f*, 120
Gene expression profiles
acute lymphoblastic leukemia, 424, 426
acute myeloid leukemia, 424, 425*f*
asthma, 594, 595*t*
brain tumors, 525–527, 526*t*
breast cancer *see* Breast cancer
cancer *see* Cancer
cardiac allograft rejection diagnosis from peripheral blood mononuclear cells, 322, 328–330, 330*f*
chronic lymphocytic leukemia, 424
chronic obstructive pulmonary disease, 608
colorectal cancer, 460, 468
congenital heart disease gene discovery, 395, 396*f*
The Connectivity Map, 83
diffuse parenchymal lung disease, 622–623
drug response prediction, 426
familial pulmonary fibrosis, 620–622, 621*f*, 622*f*
gliomas, 523, 524–525
head and neck cancer, 511, 513, 515, 516–517
heart failure monitoring, 319, 320
Helicobacter pylori, 634
hepatic stellate cells, 651
hepatitis B infection, 779
hepatitis C infection, 779–780
infectious disease responses, 736–739, 748
infectious disease (sepsis), 766
pathogen signatures, 766–767, 768*f*
leukemias, 423–424
liver tissue, 648
lung cancer *see* Lung cancer
lymphoma, 409–410
Burkitt's, 415–416
diffuse large B-cell, 411
follicular, 413–414
Hodgkin/Reed-Sternberg cells, 412–413
mantle cell, 415
primary mediastinal large B-cell, 411–412
multiple sclerosis, 568, 569
ovarian cancer, 495
pancreatic mucinous cystic neoplasms, 506
prostate cancer *see* Prostate cancer
rheumatoid arthritis, 78, 556, 569
thrombotic event prediction with anticardiolipin antibodies, 380, 380*f*, 382*f*
Gene expression-based high-throughput screening, drug discovery, 83
Gene identifier (NCBI), 121
Gene Ontology (GO), 121
Gene patenting, 196, 217
Gene regulatory networks
systems biology approach, 132
Gene set enrichment analysis (GSEA), 76
Gene silencing
cancer, 496
Gene–environment interactions *see* Environment–gene interactions
Gene polymorphism studies, 612
GENECARD, 290, 292
GeneCards, 124
GeneChip studies *see* Affymetrix microarrays
Genentech, 545
GenePattern, 122, 123*f*
GENEQUEST, 291
Genes, 4
coding/non-coding, 16
human genome, 14–16
Genetic Analysis of Idiopathic Thrombosis (GAIT), 379
Genetic Association Information Network (GAIN), 171, 205
genetic association studies (GAS), 365
Genetic counseling, 197, 339, 397–398
Genetic counselors, 196, 197
Genetic exceptionalism, 246–247
Genetic Information Nondiscrimination Act, 216
Genetic medicine
definition, 2
see also Genomic medicine
Genetic reductionism, 246
Genetic testing, 191
consent, 196, 197
current levels, 223
economic evaluation, 235
evaluation framework, 239, 239*t*
family members, 235
integration into clinical practice, 164, 168, 169*t*
regulation, 225–227
validation, 235
value-based reimbursement, 239–240, 240*f*
Genetic variation *see* Variation
Genetics
definition, 2
GenMAPP, 124, 124*f*
Genome, 4
composition, 16
definition, 2
systems biology approach, 131, 132

Genome assembler software, 30
Genome segment arrays (tiling paths), 14, 15*f*
Genome information, tools and catalogs of, 170
Genome sequencing
  electronic medical records, 143
  systems medicine applications, 135
  whole genome shotgun method *see* Whole genome shotgun sequencing
  see also DNA sequencing
Genome-based clinical trials, 172
Genome-wide association studies, 4–5, 38–40, 191, 125, 252, 284, 365, 583, 623
  Alzheimer's disease (late onset), 691
  asthma, 594
  blood lipids, 279, 280*t*
  cardiovascular disease, 279–280, 280*t*
    chromosome 9p21 region of interest, 294
  case–control studies, 38–39
    control genotyping databases, 38
  colorectal cancer, 465
  complex disease, 164
  coronary artery disease/myocardial infarction, 290–291
  diabetes type 1, 677, 680–681
  diabetes type 2, 677–679, 678*t*, 679*t*
    enrichment analysis, 680
  electronic medical records, 145–146
    accessing populations, 146
  experimental design, 38–40
    continuous phenotype studies, 39
    disease phenotype characterization, 39–40
    genotyping technologies, 40
    population stratification, 40
    power calulations, 39
    sample size, 39, 39*t*
    study focus, 38–39
  infectious diseases susceptibility, 735, 748
  inflammatory bowel disease, 578
  pharmacogenetics, 179–180
  prostate cancer, 478
  raw data quality control, 42–43
  sample collection/processing, 43
  viral hepatitis susceptibility, 777*t*
Genomes, 2, 4
Genomic Health, 194
Genomic imprinting, 60–61
Genomic literacy, 197–198
  health professionals, 249
Genomic medicine
  complex disease, 165–167, 167*f*
  definition, 2
  direct-to-consumer marketing, 214
  economic evaluation, 233
  educational strategies
    core competencies, 197
    healthcare professionals, 197
  electronic medical records, 143–144, 144*t*
    genomic research, 145–148, 148*t*
  federal regulation, 223
  health professional delivery, 197, 215–216, 216*t*
  policy issues
    access/reimbursement, 214–215
    services integration, 213–217
  privacy issues, 166
  translation into medical practice *see* Translational genomics
Genomic signatures, 3
Genomic testing *see* Genetic testing
Genomics, 2
  applications, 246
  personalized medicine approach, 155–156
  research policy issues, 211
    race as analytical variable, 210–211
    results disclosure to participants, 212–213
Genome Research for Human Health, 210
Genotyping technologies, 40
  comparative aspects, 42
  copy number variation (CNVs) detection, 42, 50, 51
  formalin-fixed paraffin-embedded tissue samples, 43
  raw data quality control, 42–43
  sample processing/throughput, 43
GEP, 428
Gerstmann–Sträussler–Scheinker syndrome, 693
Ghrelin
  obesity, 667
GJA5, atrial fibrillation, 365
GKCR gene, common variants, 277
GLA, corneal disease, 713
Glaucoma, 716
Gleevec, 545
  see also Imatinib mesylate
Glial fibrillary acidic protein (GFAP)
  glioma prognosis, 526
  hepatic stellate cells, 649
  multiple sclerosis biomarker, 570
Glial-derived neurotrophic factor (GDNF)
  hepatic stellate cells, 649
Glioblastoma, 522
  alkylating agent chemoresistance, 527
  chromosomal alterations, 524–525
  de novo/primary tumors, 523
  FGFR1 gene mutation detection, 33, 34
  gene expression profiles, 525
  genetic alterations, 523, 524*f*
  mir-21 alterations, 527
  proteomics, 527
  risk factors, 523
  secondary, 523
Gliomas, 522–527
  chromosomal alterations, 524–525, 525*t*
  classification/grading, 522, 523*t*
  DNA methylation profiles, treatment response prediction, 67
  epidermal growth factor receptor kinase inhibitors, 527
  gene expression profiles, 523, 525–527, 526*t*
  molecular alterations, 522–523
  pharmacogenomics, 527
  predisposition, 523–524
  prognostic factors, 524
  proteomics, 527
Gln279Arg, 612
Global Registry of Acute Coronary Events (GRACE), 303–304
Glomerulonephritis, *Fcgr3* copy number variation, 52
Glucocorticoid resistant asthma, 596
Glucocorticoid-remediable aldosteronism, hypertension, 263
Glucokinase mutations, maturity onset diabetes of the young type 2 (MODY2), 192
Glucokinase regulatory protein (*GCKR*) polymorphism, 279
Glycoproteomics, liver function, 655
Glucose-stimulated insulin secretion, 102–103, 103*f*
Glutathione peroxidase, 655
Glutathione *S*-transferase M1 (*GSTM1*), 251, 533
Glutathione *S*-transferase P1 (*GSP1*)
  aberrant DNA methylation
    head and neck cancer monitoring, 516
    lung cancer, 434
    prostate cancer, 67, 478
Glutathione *S*-transferases, cigarette smoke metabolism, 433
Glycopeptide capture technique, blood protein biomarker analysis, 136
Glycoprotein IIb/IIIa inhibitors, pharmacogenomics, 297
GNB3
  atrial fibrillation, 365
  obesity, 666
Google, 155
Gordon's syndrome *see* Pseudohypoaldosteronism type 2
GP73 (golgi protein 73), hepatocellular carcinoma, 655
GPD1L, Brugada syndrome, 356
GPNMB, glioma prognosis, 526
GPR154/NPSR1 (G protein-coupled receptor for asthma susceptibility), 593
GPRA, chronic obstructive pulmonary disease, 607
Gram-negative bacteria
  gene expression response to infection, 767
  host detection, 763
  host genetic susceptibility, 749*t*, 751–753, 763
  sepsis, 760
Gram-positive bacteria
  gene expression response to infection, 767
  host genetic susceptibility, 748–751, 749*t*, 763
  host recognition, 748, 763
  sepsis, 760

Granulocyte colony-stimulating factor,
multiple sclerosis lesion expression, 568
Grateful Med, 153
GRN (progranulin) mutations, frontotemporal
dementia, 691
GroEL, *Helicobacter pylori* infection marker, 634
Groenouw's (granular) type 1 corneal
dystrophy, 714, 715
GSDML, 597
GSN, corneal disease, 713
GSTM1 *see* Glutathione *S*-transferase M1
GSTP1 *see* Glutathione *S*-transferase P1
GSTT1 *see* Glutathione *S*-transferase T1
Guillain-Barre syndrome, 570
GUSTO-IV, 309
Gut microbiota
contribution to metabolic profile in obesity,
101
GWAS approach, 283
GWAS report, 5

H

H19, 22
Haemophilus influenzae
chronic obstructive pulmonary disease
exacerbations, 606
Hamartomatous syndromes, colorectal cancer
risk, 463–464
HAND2, ventricular development regulation,
395
Haplotypes, 17
HapMap project *see* HapMap project
HapMap project, 27, 38, 180, 181, 205, 209,
211, 252, 706, 735
coronary artery disease studies, 290
pharmacogenetic marker selection, 181–182
Haptoglobin, hepatitis B infection proteomics,
654
HDL cholesterol, 284
HDM2 mutations, gliomas, 523
Head and neck cancer, 511
applications of genomics, 517–518
bevacizumab treatment, 541
biomarkers, 513
distant metastases, 515
cytotoxic agent resistance, 516
diagnosis, 513
environmental risk factors, 511–512
field cancerization, 513
gene expression profiles, 511
clinical outcome associations, 515
locoregional recurrence prediction,
516–517
lymph node metastasis, 513
genomic instability/DNA repair defects, 512
human papilloma virus association, 511, 512
metabolomics, 517
monitoring, 516–517
pharmacogenomics, 516
pre-malignant lesions, 512
prognosis, 513–515, 514*t*
proteomics, 517
screening, 512–513
cytology with DNA-image cytometry,
512
fluorescence spectroscopy, 512–513
TNM, 513
Health care information technology (HIT), 171
Health economics, 233
Health insurance discrimination, 216, 217
Health Level 7 (HL7)
standards
data, 167
Health professional and public education, 7
Health professionals
education, 215, 249
genomic literacy, 249
genomic medicine delivery, 215–215, 216*t*
policy issues, 215
Heart failure, 313–323
BiDil adjunct treatment, population
stratification by race, 175, 319
biomarkers/biosignatures, 320, 322*f*
cardiac transplantation, 326
cardiomyocyte necrosis/apoptosis, 316–317,
318*f*
chronic obstructive pulmonary disease, 610
classification, 314–315
clinico-genomic, 320, 321*t*
diagnosis, 317–318
diastolic, 313
environment–gene interactions, 314
etiology, 313
genetic basis, 313–314, 314*f*, 317*t*
genomic profiling, 319
hypertrophic cardiomyopathy, 337
monitoring, 319
myocardial infarction, 296
neurohumoral mechanisms, 315–316
pharmacogenomics, 318–319
with preserved ejection fraction, 313, 318
prognosis, 318
right ventricular, 314
systolic, 313
transcriptional profiling, 319–320
ventricular remodeling, 313, 316, 316*f*
Heat shock proteins, multiple sclerosis
proteomics, 570
HeavyMethyl (HM) assays, methylation
assessment in disease screening samples, 65
Hedgehog pathway genes, primary biliary
cirrhosis, 652
Helicobacter pylori, 636
adhesins, 633
BabA polymorphism, 633
SabA, 633
clarithromycin resistance, 635
colonization rates, 630
cytotoxin associated gene product (CagA),
631, 632–633
cag PAI-encoded Type IV secretion
system, 632–633
geographic differences, 633
polymorphisms, 633, 633*f*
eradication, 628, 635
host polymorphisms influence, 636
pharmacogenomics, 635
flagella, 633
gastric cancer association, 631
lymphomas, 401
genome, 629–635
allelic diversity, 630–635
pathogenicity islands, 629
plasticity zone, 629
structure, 629–630
molecular diagnostic techniques, 636
peptic ulcer disease, 627
genomic markers, 634
geographic strain differences, 634
proteomic markers, 634
serological markers, 634–635
ulcer formation, 629
primers for polymorphic regions, 632*t*
urease, 633
vacuolating cytotoxin (VacA), 631
vacA polymorphism, 631, 631*f*, 634
virulence determinants, 631
Type I/Type II strains, 631
whole genome expression studies, 633–634
Helper T cells
pathogen-specific activation, 738
see also CD4 T cells
Hematologic malignancies
monoclonal antibody therapy, 538–540
small molecule-targeted therapy, 540,
542–543
Hematopoietic progenitor cell antigen (CD34)
precursor, thrombotic event prediction
with anticardiolipin antibodies, 380
Heme oxygenase-1 (HMOX1), chronic
obstructive pulmonary disease, 604
Hemochromatosis gene polymorphism, 289
Hemoglobinopathies, malaria outcome, 735
Hemostasis, 374–387
circulating cellular/protein influences,
378–379
coagulation protein genotype–phenotype
relationships, 376–377
gene–environment interactions, 377–378
instrumentation research, 384
personalized approach, 380
pharmacogenomics, 384–386
twin studies, 376
vascular bed-specific, 377–378
Hepatic lipase (*LIPC*)
deficiency, 277
lipoprotein metabolism, 270, 271
genome-wide associations, 279
obesity association, 668, 671
Hepatic steatosis, metabolic profiling, 101

Hepatic stellate cells, 645
activation, 649–650
genomic studies, 651
liver fibrosis, 649–651
proteomics, 653
Hepatitis A, 774
diagnosis, 777
predisposition, 776
transmission, 776
virology, 774, 775*f*
Hepatitis B, 774
adaptive immune response/antibody production, 779
chronic persistent infection, 774
diagnosis, 777
DNA quantitation, 778
serology, 777, 778*t*
environmental factors, 776
gene expression profiles, 651
genetic susceptibility, 776–777, 777*t*
HLA Class II genes, 736
genome, 775, 775*f*
genomic studies, 779*t*
genotypes A-H, 775
hepatitis D (delta hepatitis) co-infection, 776
hepatocellular carcinoma, 653, 654–655, 774, 779*t*
hepatocyte entry, 778
HIV co-infection, 776
innate immune response, 778, 779
pathogenesis, 776, 778, 779
proteomics, 654, 779–780
screening, 777, 778
transmission, 776
vertical, 776
treatment, 781
interferon, 782
nucleoside analogs, 781
pharmacogenomics, 782
virology, 775, 775*f*
Hepatitis C, 774
chronic persistent infection, 774, 778
diagnosis, 777
RNA quantitation, 778
serology, 777
environmental factors, 776
gene expression profiles, 651, 780
genetic susceptibility, 776–777, 777*t*
HLA Class II genes, 736
genome, 775–776, 775*f*
genomic studies, 779*t*
genotypes/subtypes, 775, 780
hepatocellular carcinoma, 653, 774, 779*t*
hepatocyte entry, 778
HIV co-infection, 776
immune response evasion, 780
innate immune response, 778, 780
liver fibrosis, 780
assessment, 778
susceptibility genes, 646
pathogenesis, 776
prognosis, 778
proteomics, 654, 655, 780
receptors, 776
recurrence in liver transplants, 780
screening, 777, 778
transmission, 776–777
vertical, 776
treatment, 782
interferon, 781
resistance, 781
virology, 775–776, 775*f*
Hepatitis D (delta hepatitis)
diagnosis, 777
hepatitis B co-infection, 776
virology, 775*f*, 776
Hepatitis E, 774
diagnosis, 777
genome, 776
predisposition, 776
transmission, 776
virology, 775*f*, 776
Hepatobiliary tumors, hereditary nonpolyposis colorectal cancer, 463
Hepatocellular carcinoma, 774
cirrhosis relationship, 645, 653
gene expression profiles, 653
hepatitis B, 654, 778
hepatitis C, 778, 780
proteomics, 780
proteomics, 654, 655
Hepatocytes, 645
hepatitis B/C virus entry, 776, 778
hepatitis virus infection, 774
interferon-α response, functional genomic studies, 782
proteomics, 653
Hepsin, gene expression in prostate cancer, 483
HER2 testing, 545
HER2-NEU marker, 455
HER2-positive tumors, 455
HER-2/neu
breast cancer *see* Breast cancer, HER2/neu positive
head and neck cancer, 515
monoclonal antibody treatment *see* Trastuzumab
Oncotype Dx assay, 533
ovarian cancer, 495
pancreatic cancer, 501
Her3, head and neck cancer, 515
HERA, 446
Herceptin, 455
see also Trastuzumab
Hereditary hemochromatosis (*HFE*)
population screening, 193, 249
Hereditary nonpolyposis colorectal cancer (Lynch syndrome), 457, 463
clinical variants, 463
colorectal cancer screening, 467
gene mutations, 463
genetic testing, 164, 465, 466–467
indications, 463
non-colorectal malignancy risk, 463
pancreatic cancer risk, 502
Heredity, definition, 2
HERG/KCNH2 (*LQT2*), 353, 355, 359
short QT interval syndrome, 361
sudden infant death syndrome, 352
Hermansky-Pudlak syndrome, pulmonary fibrosis, 616, 617, 620
Heterocyclic amine-metabolising enzyme polymorphisms, colorectal cancer risk, 464–465
Heterotaxy (*ZIC3*), congenital heart disease, 392–393
Heterozygosity, 18
HuRef genome analysis, 18
HFE mutations *see* Hereditary hemochromatosis
Hierarchical clustering, 118
B-cell mediated kidney transplant rejection, 116
disease sub-type discovery, 119
diffuse large B-cell lymphoma, 114
familial pulmonary fibrosis candidate genes, 620–622
gene expression profiles
data analysis, 76
heart failure, 319
lung cancer, 434
High-density lipoprotein cholesterol, 271
cardiovascular risk, 271, 278–279
genetic disorders
elevation, 276
reduction, 274, 276
genetic influences, 271
common variants, 275*t*, 276–277
novel therapeutic targets, 282
screening, 272
High-density lipoproteins (HDL), 269
apolipoproteins, 269
metabolism, 270–271, 271*f*
remodeling, 271
High-throughput techniques, 111
DNA methylation patterns, lung cancer early diagnosis, 434, 435
gene discovery
colorectal cancer, 460
Histone deacetylase 2 (HDAC2)
chronic obstructive pulmonary disease, 608
Histone deacetylase inhibitors, 422
Histones, 19, 20
Acetylation
chronic obstructive pulmonary disease, 608
prostate cancer, 478
methylation
prostate cancer, 478

N-terminal modifications
ovarian cancer, 496
Histopathology, bioinformatics applications, 116
HIV disease, 739
HLA alleles
diabetes type 1, 676, 677, 680, 681
infectious disease susceptibility, 736
hepatitis B, 776
hepatitis C, 776
HIV infection/AIDS, 762
mycobacterial disease, 753, 754
pharmacogenomics, 182
sarcoidosis, 617
HLA-B*5701, 172, 242
HLA-B*1502, 172, 186
HLA Class I genes
HIV infection
susceptibility, 736
malaria susceptibility, 736
tuberculosis susceptibility, 736
HLA Class II genes
hepatitis B persistence, 776
hepatitis B/C susceptibility, 736
leprosy susceptibility, 736
typhoid fever susceptibility, 736
HLA-A1, sarcoidosis, 617
HLA-B*5701, abacavir hypersensitivity reactions, 182
HLA-B8
sarcoidosis, 617
HLA-B13, sarcoidosis, 617
HLA-B22, sarcoidosis, 617
HLA-B27
spondyloarthropathies
inflammatory bowel disease association, 581
HLA-B35
arthropathy, inflammatory bowel disease association, 581
HLA-B44, arthropathy, inflammatory bowel disease association, 581
HLA-CRB1*14, multiple sclerosis association, 566
HLA-CRB1*17, multiple sclerosis association, 566
HLA-DP, 552
HLA-DQ, 552
HLA-DQA1, diabetes type 1, 676
HLA-DQB1
asthma, 593
diabetes type 1, 676
sarcoidosis, 617
HLA-DQB1*0301, hepatitis C, 776
HLA-DR, 289, 552
HLA-DR3, sarcoidosis, 617
HLA-DR15, sarcoidosis, 617
HLA-DR16, sarcoidosis, 617
HLA-DR17, sarcoidosis, 617
HLA-DRB1
asthma, 593, 594
diabetes type 1, 676
rheumatoid arthritis, 551, 552–554, 556, 557
shared epitope alleles, 553, 553*t*, 554
sarcoidosis, 617
HLA-DRB1*0103, ulcerative colitis, 580
HLA-DRB1*03, Crohn's disease, 581
HLA-DRB1*1302, hepatitis B clearance, 776
HLA-DRB1*1502, ulcerative colitis, 580
HLA-DRw52, sarcoidosis, 617
HLA-G, asthma, 593
HMG-CoA lyase defect, newborn screening, 101–102
HMG-CoA reductase inhibitors *see* Statins
HMGCR, low-density lipoprotein cholesterol interindividual variation, 274
HMP database, 100
HNF4A, diabetes type 2, 677
HNPCC, 1
HNSCC, 518
Hodgkin lymphoma, 401, 412–413
gene expression profiles, 412, 413
prognosis, 413
treatment, 534
Holt-Oram syndrome (*TBX5*), congenital heart disease, 391, 396
HopQ, *Helicobacter pylori* virulence factors, 633
HopZ, *Helicobacter pylori* adhesins, 633
Host polymorphisms, 636–637
HpA, *Helicobacter pylori* adhesins, 633
HPV-positive cancers, 518
Hsp27, hepatitis C proteomics, 780
Hsp70
head and neck cancer, 517
polymorphism, 289
Hsp90
androgen receptor regulatory activity, 485–486
head and neck cancer, 517
prostate cancer, 486
Hsp90 inhibitors
prostate cancer treatment, 83
HSPA1A, *Mycobacterium leprae* susceptibility, 753
HTR2C, obesity, 671
huJ-591 (anti-PSMA$_{EXT}$), 539–540, 540*f*
Human Epigenome Project, 62, 63*t*
Human genome, 14–16
conserved regions, 16
expression, 19
genes, 14–16
"reference sequence", 16, 52
sequencing, 27
variation, 17–19
HuRef genome analysis, 18
Human Genome Epidemiology Network (HuGE Net), 248
Human immunodeficiency virus (HIV) *see* Human immunodeficiency virus-1 (HIV-1); Human immunodeficiency virus-2 (HIV-2); Human immunodeficiency virus (HIV) infection/AIDS
Human Genome Project, 1
Human immunodeficiency virus-1 (HIV-1)
drug resistance mutations, 224
Human immunodeficiency virus (HIV) infection/AIDS
abacavir hypersensitivity reaction
*HLA-B*5701* genotyoing, 182
CCL3L1 copy number variation, 53–54
hepatitis B/C co-infection, 776
lymphomas association, 401
secondary dementia, 687
susceptibility genetics, 762
CCR2 mutations, 736
CCR5 deletion-related resistance, 735, 736
HLA Class I gene, 736
treatment
viral resistance mutations in selection, 224
Human papillomavirus
head and neck cancer, 511, 512
Human parology database, 48
Human T-cell leukemia virus-1 (HTLV1), adult T-cell leukemia/lymphoma, 401
Human Tissue Bill, 249
Human tissues use, public health genomics, 249
HumGen database, 248
Huntington's disease, 246
HuRef genome, 5, 6
analysis of variation, 18
Hyaluronan, hepatic stellate cell production, 650
Hydralazine, pharmacogenomics, 319
5-Hydroxytryptamine transporter (*5-HTT*; SLC6A4)
antidepressant drug response, 723, 727*t*
Hyperforin metabolism, 727
Hyperlipoproteinemia type III (familial dysbetalipoproteinemia), 277, 278, 279
genetic testing, 282–283
Hypertension, 19, 259–266
age associations, 260, 262*t*
cardiovascular risk, 264
diagnosis, 262–263
end-organ damage, 260, 263
environment–gene interactions, 259, 260, 260*f*
epidemiology, 259
essential (common form), 259, 260
heart failure, 313
monogenic forms, 259, 263
nephrosclerosis, 264
obesity association, 663
pharmacogenomics, 264–266
predisposition, 260–262
candidate gene approach, 262
salt-sensitivity, 260–261

Hypertension, (*Continued*)
primary/secondary, 262
prognosis, 263–264
stroke, 263
vascular-occlusive dementia, 687
Hypertrophic cardiomyopathy, 313, 336–345
calcium-handling, 338–339
clinical presentation, 337–338
septal morphology, 337, 337*t*, 338*f*, 341
definition, 336
diagnostic imaging, 337–338
dual-chamber pacing, 345
family screening, 342–343, 342*f*
follow-up, 343
genetic couseling/echocardiography-guided genetic testing, 341
genotype–phenotype relationships, 339–341
genotypic/phenotypic heterogeneity, 336, 337
implantable cardiverter defibrillators, 344–345
indications, 344*f*
Maze procedure, 345
metabolic/syndromal, 339, 341*t*
Noonan syndrome (*PTPM11*), 393
molecular genetics, 338–342, 339*t*, 340*f*
MYH7 (beta myosin heavy chain) mutations, 313, 336, 338, 340, 341
nomenclature, 336–337
pharmacogenomics, 343
pharmacological therapy, 343
sarcomeric/myofilament, 337, 338, 340*f*
septal ablation, 344
septal myectomy surgery, 344
sports participation, 343
transcriptomics, 341–342
Z-disk, 336, 338–339, 341
Hypothalamus, body weight regulation, 662
Hypothalamus-pituitary-adrenal axis, depression, 724, 727, 728

I

I-ELCAP study, 434
^{131}I-tositumomab, 538
non-Hodgkin lymphoma treatment, 538
IBD1, 576
IBD1-IBD9, 576
IBD2, 576
IBD3, 576
IBD5, 576, 578
infliximab pharmacogenomics, 582
IBD6, 576, 578
IBDCHip, 581–582
ICAM-1 (intercelluler cell adhesion molecule 1)
chronic obstructive pulmonary disease, 607
thrombosis, 379
ICD *see* International Classification of Diseases (ICD)
*ice*A, *Helicobacter pylori* virulence factors, 634
Iceland
biobanking policy issues, 212
ICF disorder, *DNMT3B* mutations, 60
Idiopathic dilated cardiomyopathy, 313
biomarkers from serial samples, 320
infectious etiologies, 314
Idiopathic generalized eplilepsy, 57
Idiopathic hypertrophic subaortic stenosis, 337
Idiopathic interstitial pneumonia, 616
Idiopathic pulmonary fibrosis
surfactant protein A variants, 619
surfactant protein B variants, 619
Idiopathic ST-elevation, 361
see also Brugada syndrome
IFIH1, diabetes type 1, 680
IFIH1 gene, 681
IgE receptor expression, multiple sclerosis lesions, 568
IGF-1 (insulin-like growth factor)
colorectal cancer, 465
obesity, 666, 668
IGF-2
genomic imprinting, 61
head and neck cancer, 517
ovarian cancer, 496
IGF-binding protein 1, sepsis, 767
IGF-binding protein 3
colorectal cancer, 465
ovarian cancer, 496
IκB
host response to *Streptococcus pneumoniae*, 750
IκB kinase inhibitors, diffuse large B-cell lymphoma, 411
IKBL+738, ulcerative colitis severity marker, 580
IKZF1 deletions, 428
IKZF1 gene, 428
IL1RL1, 597
IL-2 receptor (IL-2Ra) gene, 571
IL-7 receptor (IL-7Ra), 571
IL10, 583
IL12B, 583
IL17RB, 597
IL18R1, 597
IL18RAP, 597
IL23R, 198, 583
Illumina
control genotyping databases, 38
Illumina arrays, 40–41, 40*f*, 41*f*
BeadChip, 81
comparative aspects, 42
copy number variation (CNVs) detection, 42, 50–52
diabetes type 1, 680
probe length, 42
whole genome amplified DNA, 43
Illumina Human1M-Duo BeadChip, 43
Illumina HumanHap300 array, 40, 42
Illumina HumanHap550 BeadChip, 42
Illumina Infinium assay, 40–41, 40*f*, 41*f*
BeadChip, 40, 40*f*, 42
processing steps, 41
type I, 40
type II, 40, 41
Illumina-Solexa GA (Genome Analyzer), 34, 35
IMAGE (Invasive Monitoring Attenuation through Gene Expression), 334
Imaging investigations
congenital heart disease, 397
head and neck cancer, 513
prostate cancer, 482–483
IMAGE study, 334
Imatinib mesylate, 194, 224, 229
BCR-ABL tyrosine kinase inhibition, 540
chronic myelogenous leukemia, 532, 534, 540
diagnostic test of patient eligibility, 532
gastrointestinal stromal tumors, 542
pharmacogenomics, 424
prostate cancer trials, 485
Immune response
follicular lymphoma gene signatures, 414
multiple sclerosis high-throughput analysis, 569–570
Immunophenotyping, lymphoma, 409
Immunosuppression
head and neck cancer risk, 511
lymphoma risk, 401
Implantable cardioverter defibrillators
hypertrophic cardiomyopathy, 344–345
long QT syndromes, 359, 360
short QT interval syndrome, 361
In utero exposure *see* Intrauterine environmental exposure
In vitro diagnostic multivariate index assay (IVDMIA), 334
In vitro diagnostics, 223
clinical studies, 226
definition, 225
investigational device exemptions (IDEs), 226
multivariate index assays (IVDMIAs), 226, 227
Premarket Approval Application, 225–226
Premarket Notification Submission/510(k) application, 225, 226
regulatory issues, 225–226, 225*f*
regulation as medical devices, 225
risk-related classification, 225–226
In vivo videomicroscopy, head and neck cancer screening, 513
Inborn errors of metabolism
neonatal hyperbilirubinemia, 653
Incidental genomic findings, 155
Incremental cost-effectiveness ratio (ICER), 234
Infectious agents, rheumatoid arthritis, 556
Infectious disease, 733
adaptive immune response, 733, 745
cytokine polymorphisms, 764–765

gene expression profiling, 736–739
common host response, 737–738, 737*f*, 745*f*
outcome of infection, 738–739
pathogen-specific responses, 738
progression signatures, 767
host genetic susceptibility, 734–736, 735*t*, 749*t*
candidate gene studies, 736
Gram-negative organisms, 751–753
Gram-positive organisms, 748–751
mouse models, 734
mycobacteria, 753–754
host susceptibility study methods, 748
innate immune response, 733, 745, 763, 766
pathogen recognition/signaling, 763–764, 766
CD14, 764
intracellular signaling molecules, 764
mannose-binding lectin, 764
NOD-like receptors, 763–764
RIG-like receptors, 763–764
Toll-like receptors (TLRs), 733, 734, 763
see also Bacterial infections; Sepsis; Viral infections
INFGR1 (interferon-gamma receptor type 1), *Leishmania* infection susceptibility, 736
Inflammatory bowel disease (IBD), 191, 575, 583
anti-microbial antibodies, 578, 580
biomarkers, 578, 580
diagnosis, 579–580
environmental factors, 576
epidemiology, 575
genetic factors, 576, 578
extra-intestinal manifestations, 581–582
genome-wide association studies, 578
intestinal barrier dysfunction, 578
lymphoma association, 580
pharmacogenomics, 582
prognosis, 580–582
screening, 578
treatment
infliximab, 582
see also Crohn's disease; Ulcerative colitis
Infliximab
inflammatory bowel disease, 582
pharmacogenomics, 582
rheumatoid arthritis, 558
Influenza
host gene expression in response to infection, 767
Ingenuity, 94
Inhibitors of apoptosis (IAP genes), autoimmune hepatitis, 652
Inkjet printing
oligonucleotide microarray fabrication, 75
Innate immune response, 763
hepatitis B, 779
hepatitis C, 779
infectious disease, 733, 745, 763, 766
Toll-like receptor (TLR) signaling, 763
INS (insulin gene)
diabetes type 1, 676, 680, 681
obesity, 666
INS-1 cells, glucose-stimulated insulin secretion, 102–103
Insertion/deletion (in/del) polymorphism, 17
complex disease association studies, 27
HuRef genome analysis, 18
resequencing process, 33
whole genome shotgun sequencing, 31, 31*f*
Insertions, 17
see also Insertion/deletion (in/del) polymorphism
InSNP, 33
Institute for Genome Sciences & Policy, 172
Institute of Medicine Committee for the Study of the Future of Public Health, 248
Insulin gene *see* INS
Insulin receptor substrate 1/2 (IRS-1/2), obesity, 668
Insulin resistance
metabolic profiling
nuclear magnetic resonance spectroscopy, 101
obesity, 103
Insulin-like growth factor receptors (IGFRs), ovarian cancer metastasis, 496
Insulin-like growth factors *see* IGF-1; IGF-2
Intellectual property rights, 238
biobanking, 212
data-sharing policies, 213
14-13-3 Interacting gene, hepatitis B infection, 651
Interactome (protein–protein interactions)
systems biology approach, 137
Intercelluler cell adhesion molecule 1 *see* ICAM-1
InterDom, 121
Interferon response genes
hepatitis C infection, 651, 654
Interferon therapy
genomics studies, 782
hepatitis B, 781
hepatitis C, 782
pharmacogenomics, 782–783
proteomics, 782
Interferon-stimulating genes (ISGs)
common host response to infection, 737, 738
hepatitis B/C infection response, 779
pathogen-specific activation, 738
Interferonγ, 747
hepatitis B response, 779
hepatitis C response, 779–780
mycobacterial disease, 754
Interferons
hepatitis B response, 779
hepatitis C response, 779–781
infectious disease response, 747
Interleukin-1 (IL-1)
Gram-positive organism infection response, 748
polymorphism
sepsis, 764
rheumatoid arthritis pathogenesis, 557
Interleukin-1β
chronic obstructive pulmonary disease, 607
gastic cancer susceptibility, 635
gastric acid secretion regulation, 628
peptic ulcer disease, 635
polymorphism-related *Helicobacter pylori* eradication rate, 636
Interleukin-1 receptor (IL-R) polymorphism
gastic cancer susceptibility, 635
sepsis, 765
Interleukin-2 (IL-2) therapy, 327, 413, 538
Interleukin-2 receptor (IL-2R)
follicular lymphoma, 413
sepsis, 767, 768
Interleukin-4 (IL-4)
host response to pathogens, 747
polymorphism
asthma, 593, 594
Leishmania susceptibility, 736
Interleukin-4 receptor (IL-4R)
asthma-related polymorphism, 593, 594
brain tumor associations, 523
follicular lymphoma, 413
Interleukin-6 (IL-6)
chronic obstructive pulmonary disease, 607
host response to pathogens, 747, 748
obesity, 671
polymorphism
inflammatory bowel disease, 581–582
sepsis, 764
sepsis severity biomarker, 767, 768
thrombosis gene linkage studies, 379
Interleukin-6 receptor (IL-6R), obesity, 666
Interleukin-7 (IL-7), 571
Interleukin-7 receptor (IL-7R) polymorphism, diabetes type 1, 571
Interleukin-8 (IL-8)
alcoholic liver disease, 652
chronic obstructive pulmonary disease, 607
Gram-positive organism response, 748
Helicobacter pylori infection, 632, 633
ovarian cancer angiogenesis, 497
peptic ulcer disease, 634
sepsis, 764
severity biomarker, 767, 768
Interleukin-10 (IL-10)
host response to pathogens, 747
polymorphism
sepsis, 764–765
Interleukin-10 receptor (IL-10R) polymorphism
multiple sclerosis, 567
mycobacterial disease susceptibility, 754

Interleukin-12 (IL-12)
host response to pathogens, 747
polymorphism, *Salmonella* infection susceptibility, 736
Interleukin-12 receptor (IL-12R) polymorphism, mycobacterial disease susceptibility, 749
Interleukin-13 (IL-13)
asthma-related polymorphism, 591, 593, 594
brain tumor associations, 523
Interleukin-18 (IL-18) polymorphism, multiple sclerosis, 567
Interleukin-22 receptor (IL-22R) polymorphism, thrombotic event prediction with anticardiolipin antibodies, 380
Interleukin-23 receptor (IL-23R) polymorphism, Crohn's disease, 578
Indeterminate colitis, 575, 582
see also Inflammatory bowel disease
Intermediate-density lipoproteins (IDL), 269
apolipoproteins, 270
metabolism, 270
International Classification of Diseases (ICD)
with Clinical Modification, 121
ICD-9, 152
ICD-9-CM, 121
ICD-10-CM, 121
International HapMap Project, 252
International Prognostic Index (IPI), diffuse large B-cell lymphoma, 409
Internet health information, 152
consumer health vocabularies, 152
consumer search characteristics, 151–152
information sites, 152–154
personalized genomics, 154–155
quality assurance, 154
InterPro, 121
Interrupted aortic arch, 391
Interstitial cystitis, diagnostic marker, 94
Interstitial lung disease *see* Diffuse parenchymal lung disease
Intraocular pressure elevation, 715–716
Intrauterine environmental exposure
cigarette smoke
asthma associations, 591
inv(16), acute myeloid leukemia, 421, 424
Investigational device exemptions (IDEs), 226
Ion channel function, multiple sclerosis/Guillain-Barre syndrome, 570
IP10 *see* CXCL10
IPF, 623
IRAK4 (IL-1 receptor-associated kinase-4)
pyogenic bacterial infection susceptibility, 736
sepsis-related intracellular signaling, 764
IRF-6, hepatitis B infection, 651
IRGM, 583
Irinotecan
adverse reactions, 469
colorectal cancer, 469
pharmacogenetics, 469
drug labels, 230
genotype-guided clinical cancer trial, 182
Iris, 712
genetic disorders, 712
Ischemia modified albumin, acute coronary syndrome biomarker, 306
ISG15, hepatitis C infection, 651
ISG16, hepatitis C infection, 651
ISGs *see* Interferon-stimulating genes
ISL1, obesity, 666
Isochores, 16
Isosorbide dinitrate pharmacogenomics, 319
Isotope-coded affinity tags (ICAT), quantitative proteomics, 91, 91*f*
Isotopic tagging for relative and absolute quantification (iTRAQ), blood protein biomarkers, 92*t*, 136

J

JAG1 (Alagille syndrome), congenital heart disease, 393
Jak/stat kinases
lung cancer, 438
ovarian cancer metastasis, 496–497
JAK2, 583
Jervell and Lange-Nielsen syndrome, 352, 357–358
clinical features, 357
sensorineural deafness, 353, 357
genetics, 353, 354*t*, 358
KCNE1 mutations, 358
KCNQ1 mutations, 358
join technique for database management, 115
JPH2 mutations (junctophilin type 2), hypertrophic cardiomyopathy, 339
Jun, head and neck cancer, 517
Junctophilin type 2 (*JPH2*) mutations, hypertrophic cardiomyopathy, 339
Juvenile polyposis syndrome, colorectal cancer risk, 464

K

k-means clustering, 118
gene expression data analysis, 76
k-nearest neighbors
microarray data analysis, 76
K-ras, 543
colorectal cancer, 65, 458, 460, 467
lung cancer, 438, 440
ovarian cancer, 496–497
pancreatic cancer, 501, 503
pancreatic cystic neoplasms, 505
Kaplan–Meier analysis, 114, 119
Kaposi's sarcoma human herpes virus-8
primary effusion lymphoma (PEL) association, 401
Karyotyping, 14, 15*f*
congenital heart disease, 381
KCNA5, atrial fibrillation, 363, 364
KCNE1
atrial fibrillation, 363
Jervell and Lange-Nielsen syndrome, 358
KCNE5, atrial fibrillation, 365
KCNH2, atrial fibrillation, 363, 364
KCNJ2
atrial fibrillation, 363, 364
LQT7, 358
KCNJ11
diabetes type 2, 677
permanent neonatal diabetes, 681
KCNQ1, 359
atrial fibrillation, 364
Jervell and Lange-Nielsen syndrome, 358
KEGG *see* Kyoto Encyclopedia of Genes and Genomes
KEL, obesity, 666
KIF6 719Arg allele, 298
KIF6 gene, 298
KIF21A mutations, congenital fibrosis of extraocular muscles type 1 (CFEOM1), 713
Kir 2.1/KCNJ2 mutations
atrial fibrillation, 364
LQT7, 358
short QT interval syndrome, 361
KIR 6.2, 198
KIRs (killing inhibitory receptors)
hepatitis C clearance, 776
c-kit therapeutic targets
gastrointestinal stromal tumors, imatinib mesylate responsiveness, 540–542
Kjer's autosomal dominant optic atrophy, OPA1 mutations, 716
Kolmogorov–Smirnov test statistic, 76, 116
KORA study, 104
Kosmix, 155
Krabbe disease, newborn screening, 193
KRAS testing, 186
KRT3, corneal disease, 713
Kuppfer cells
therapeutic inhibition in cirrhosis, 648
Kveim reaction, 619
KvLQT1/KCNQI mutations
Jervell and Lange-Nielsen syndrome, 358
LQT1, 353, 354, 359
short QT interval syndrome, 361
Kyoto Encyclopedia of Genes and Genomes (KEGG), 100, 124

L

L-myc, Hodgkin lymphoma/Reed-Sternberg cells, 413
L-type calcium channel (Ca(V)1.2) mutations, Timothy syndrome, 358
Laboratory-developed tests
regulation, 223, 226–227
Laboratory standards, 194–195
Lactate levels, sepsis staging/severity assessment, 767

Lamin mutations
atrial fibrillation, 364
Laminin
hepatic stellate cell production, 650
Lamivudine
hepatitis B treatment, 781
LAMP2 mutations, hypertrophic cardiomyopathy, 339
Laryngeal squamous cell carcinoma
diagnosis, 513
protein expression profiles, 517
see also Head and neck cancer
Laser capture microdissection
head and neck cancer proteomics, 517
tissue sampling
DNA methylation assessment, 62
microarray experiments, 75
LATS1/LATS2 DNA methylation, breast cancer prognosis, 67
Lattice type 1 corneal dystrophy, 714, 715
LCAT gene mutations, 276, 283
LDB3 mutations, hypertrophic cardiomyopathy, 338, 341
Lead optimization, 227
Learning disability *see* Mental retardation
Leber's congenital amaurosis, RPE65 mutations, 716
Leber's hereditary optic neuropathy, 716
Lecithin-cholesterol acyltransferase (LCAT), 271
deficiency, 276, 279
therapeutic upregulation, 282
Leflunomide
rheumatoid arthritis, 558
Left cervicothoracic sympathetic ganglionectomy, long QT syndromes, 359
Left ventricular dysfunction (LVD), 104
Legal issues
genetic discrimination, 216
genomics research policy, 210
public health genomics, 246, 249
Legionella
genetics of host response, 752, 763
infection susceptibility, TLR5 mutations, 736
Leiden Thrombophilia Study, 376
Leishmania infection susceptibility, 736
Lenalidomide, 428, 543
Lens, 712
genetic disorders, 715
LEP *see* Leptin
Leprosy susceptibility, 736
HLA Class II genes, 736
TLR2 mutations, 736
TNF gene mutations, 736
Leptin (LEP)
body weight regulation, 662
obesity, 664, 665, 666, 667, 669
Leptin receptor (LEPR) mutations, obesity, 664, 665
let-7 microRNAs, lung cancer, 440
Letrozole, 533
Leukemia, 421–428
biomarkers identification, 426–427
classification, 113, 421, 422*t*
copy number variation (CNVs), 428
cytogenetics, 421–422, 422*t*
diagnostic-therapeutic combinations, 533–534
gene expression profiling, 78, 79*f*, 423–424, 425*f*, 428
genomics applications
class prediction, 423–424, 426*f*
molecular genetics, 421–422, 422*t*
pharmacogenomics, 424–426
drug discovery, 426
drug response prediction, 426
prognostic factors, 423, 423*t*, 426–427
treatment, 422–423
Leukemia and Lymphoma Molecular Profiling Project (LLMPP), 409
Leukemia inhibitory factor (LIF), glioma prognosis, 526
Leukotriene A4 hydrolase (LTA4H)
myocardial infarction risk, 297
HapK haplotype association, 294, 304
Leukotriene B4, myocardial infarction risk, 294
Leukotriene inhibitors
asthma, 596
chronic obstructive pulmonary disease, 611
myocardial infarction, 297
pharmacogenomics, 596
Leukotriene pathway, myocardial infarction, 294
Lev-Lenegre progressive cardiac conduction disease
clinical features, 326
genetics, 326, 327
management, 326
Levalbuterol, asthma, 596
Levodopa
Parkinson's disease, 700, 706
pharmacogenomics, 707
Lewy bodies, 687
α-synuclein, 692, 693*f*, 702
Parkinson's disease, 700, 703
LGALS2 polymorphism, myocardial infarction, 293, 296
LGALS4 polymorphism, ovarian cancer, 495
Li-Fraumeni syndrome
brain tumors, 523
breast cancer, 448
p53 gene mutations, 458
Licensing fees, 196
Liddle's syndrome, hypertension, 263
LIF (leukemia inhibitory factor), glioma prognosis, 526
Life insurance discrimination, 216
LIM domain binding 3 (*LDB3* mutations), hypertrophic cardiomyopathy), 338, 341
LIM protein (*CSRP3* mutations), hypertrophic cardiomyopathy, 338
Linkage disequilibrium, 17
pharmacogenetic candidate gene approaches, 180
prostate cancer studies, 477
sarcoidosis, 618
thrombosis-related genes, 380
Linkage studies (genome scans)
Alzheimer's disease (late onset), 690
asthma, 591–593, 592*t*
chronic obstructive pulmonary disease, 608
depression, 724
diabetes type 1, 676
diabetes type 2, 677
familial colorectal cancer, 464
familial interstitial pneumonia, 623
infectious diseases susceptibility, 735, 748
inflammatory bowel disease, 576
multiple sclerosis, 566
obesity, 664, 666
rheumatoid arthritis, 552
sarcoidosis, 618–619
thrombosis, 379–380
LIPC see Hepatic lipase
LIPG gene, 284
Lipid disorders, 269
cardiovascular risk, 269, 272, 274, 277, 278–279
common genetic variants, 274, 275*t*, 276–277, 278
genetic disorders, 273*t*
genetic testing, 282–283
genome-wide association studies, 279, 280*t*
high-density lipoprotein, 274, 276–277
low-density lipoprotein, 272–274
screening, 272
triglycerides, 277–278
Lipid-modulating therapies, pharmacogenomics, 281
Lipocalin 2, pancreatic cancer, 506
Lipopolysaccharide
gene expression response, 767
host Gram-negative bacteria recognition, 751
TLR4 response, 738
sepsis-related signaling, 764
Lipopolysaccharide-binding protein (LBP), host response to infection
Enterobacteriacea, 751
Streptococcus pneumoniae, 750
Lipoprotein lipase (LPL)
chylomicrons metabolism, 270
mutations, familial chylomicronemia syndrome, 277, 279
polymorphism, 276
blood lipid genome-wide association studies, 279

Lipoprotein lipase (LPL) (*Continued*)
common variants, 277
obesity association, 668
very-low-density lipoprotein metabolism, 270
Lipoprotein(a), 270
Lipoproteins
classes, 269
genetic factors in phenotypic variance, 271–272
metabolism, 269–271
endogenous pathway, 270, 270*f*
exogenous pathway, 270, 270*f*
therapeutic targets, 281–282
Lipoteichoic acid, 748
5-Lipoxygenase activator protein (FLAP), cardiovascular risk, 304
5-Lipoxygenase pathway genes, cardiovascular risk, 304
Liquid chromatography–mass spectrometry
acute coronary syndrome metabolomics, 306
quantitative proteomics, 92
Liquid chromatography–tandem mass spectrometry (LC-MS/MS), proteomics
head and neck cancer, 517
ovarian cancer, 495
Listeria monocytogenes
genetics of host response, 751
Literature mining, prostate cancer, 484
Liver
aging process, 655
lipopolysaccharide-induced injury, 655
lipoprotein metabolism, 270, 271
proteome, 649
transcriptome expression, 648–649
Liver cancer
see also Hepatocellular carcinoma
Liver disease
biomarkers, 653, 654
proteomics, 653–654
metalloproteome, 655
transcriptomics, 651–655
Liver fibrosis, 645
development, 646, 649–651
hepatitis C, 781
proteomic index development, 655
see also Cirrhosis
Liver function tests, proteomics, 655
LKB1 pathway mutations, Peutz-Jeghers syndrome, 464
LKX1, 597
LM02, diffuse large B-cell lymphoma prognosis, 411
LMNA mutations *see* Lamin mutations
LMX1B, ocular development, 712
Long-chain L-3-hydroxyacyl CoA dehydrogenase (LCHAD) defect, newborn screening, 102
Long QT syndromes, 350–357
acquired/drug-induced forms, 352
clinical features, 352–353, 353*f*
complex forms, 357
costs of testing, 196
diagnostic criteria, 352, 352*t*
genetics, 353–356, 354*t*, 355*f*
see also Romano-Ward syndrome
genotype–phenotype correlations, 359
inherited forms, 352
management, 359–360
gene-specific approach, 192, 360
torsade de pointes, 350, 351–352, 351*f*
Longitudinal Study of Aging of Danish Twins, 376
Lou Gehrig's disease *see* Amyotrophic lateral sclerosis
Low birth weight, chronic obstructive pulmonary disease association, 604
Low-density lipoprotein cholesterol
cardiovascular risk, 271
genetic factors modulating, 278, 294
genetic common variants, 274, 275*t*
genetic disorders, 278
elevation, 272–274
reduction, 274
genetic influences, 271
PCSK9 polymorphism, 294, 295, 295*f*
lipid-modulating therapies, 281
novel therapeutic targets, 281–282
screening, 272
Low-density lipoprotein receptor (LDLR)
loss of function mutations, 282
familial hypercholesterolemia, 272
low-density lipoprotein cholesterol interindividual variation, 274
obesity, 666
therapeutic targeting, 281
Low-density lipoprotein receptor-mediated endocytosis, 270, 271
Low-density lipoproteins (LDL), 269
apolipoproteins, 270
metabolism, 270
oxidation, 377
LOXL1 gene, 198, 717
LPL *see* Lipoprotein lipase
LQT1 (KVLQT1/KCNQ1), 353, 354, 359, 360
LQT2 (HERG/KCNH2), 353, 355, 359, 360
LQT3 see Cardiac sodium channel gene (SCN5A) mutations
LQT4 (ankryn-β), 353, 356
LQT5 (minK/KCNE1), 353, 354, 356
LQT6 (MiRP1), 353, 355, 356
LQT7 (Kir 2.1; KCNJ2), 358
LQT9 (caveolin-3), 356–357
LQT10 (SCN4B), 357
LRRK2 (*PARK8*; dardarin)
diffuse Lewy body dementia, 692–693, 694
Parkinson's disease, 704–705, 708
LTA4H *see* Leukotriene A4 hydrolase
LTBP4 (latent transforming growth factor-β binding protein-4), chronic obstructive pulmonary disease, 604
LTC4S, leukotriene inhibitor pharmacogenomics, 596
Lumera Nanocapture Gold, 135
Lung cancer, 432–442
biomarkers, 434, 435
cigarette smoking association, 432
at-risk smoker identification, 434–435
genetic influences, 432, 433–435
classification, 435–438
histological, 435–436
molecular, 436–437
early diagnosis, 434, 441
epidemiology, 432, 433, 434
never smokers, 433
family history, 433
field cancerization, 434, 436
gene expression profiles, 434, 435, 436, 438, 440–441
"lung metagene" model, 437
prognostic value, 114, 436–437, 437*f*
subset identification, 436, 437*f*
gene polymorphism studies, 433–434
growth factors/growth factor receptor alterations, 438–439, 439*f*
microRNA alterations, 439–440
molecular alterations, 438–439, 438*t*
oncogene activation, 438
pathway signatures, 439–440
pathogenesis, 438–441
pre-malignant lesions, 436
prevention/chemoprevention, 435
screening, 434, 435
serum protein antibody microarrays, 135
small molecule-targeted therapy
erlotinib, 542
gefitinib, 542
stage at diagnosis, 432, 434
staging
TNM, 436
therapeutic targets, 438–439, 440–441
tumor suppressor gene mutations, 439
Lymph node status
breast cancer gene expression profiles, 452, 453*f*
LymphoChip, 78, 409, 410, 411, 412
Lymphocyte cytosolic protein-1 (*LCP1*), NSC 624004 anti-cancer agent susceptibility, 115
Lymphoma, 401
chromosome aberrations, 408*t*, 409
diagnosis, 409
diagnostic-therapeutic combinations
gene expression profiling, 409–410
data analysis, 76
inflammatory bowel disease association, 580
novel therapeutics, 409

predisposition, 407
prognosis, 409
types, 401, 408*t*
Lymphoscintigraphy, head and neck cancer diagnosis, 513
Lymphotoxin-α
galectin-2 binding, 293, 296
polymorphism, 289
Lynch syndrome, 1
See also Hereditary nonpolyposis colorectal cancer
Lysophosphatidic acid (LPA), ovarian cancer metastasis, 497
Lysosome-associated membrane protein 2 (*LAMP2*) mutations, hypertrophic cardiomyopathy, 339

M
M1S1, corneal disease, 713
Mac-2BP (Mac-2-binding protein), cholangiocarcinoma, 654
Machine learning, 118
see also Supervised learning; Unsupervised learning
Macrophage migration inhibitory factor (MIF), sepsis, 764
Macrophages
chronic obstructive pulmonary disease, 606–607
pathogen responses, 733
gene expression profiles, 737, 738, 766
innate immune response, 745, 763
sarcoidosis, 617–618
TLR-mediated responses, 738
Macular degeneration, 27, 191, 716
MAGI2, 584
Magnetic resonance imaging (MRI)
Alzheimer's disease, 689
Crohn's disease, 579–580
head and neck cancer diagnosis, 513
heart failure, 318
hypertrophic cardiomyopathy, 337–338
pancreatic mucinous cystic neoplasms, 505
prostate cancer, 482
Magnetic resonance spectroscopy (MRS)
head and neck cancer, 517
prostate cancer, 483
Major histocompatibility region, 552–554
microarrays, T cell response analysis in multiple sclerosis, 569
Malaria
role of hemoglobinopathies in outcome, 735
susceptibility
HLA Class I genes, 736
TNF gene mutation, 736
Malmo Diet and Cancer Study–Cardiovascular Cohort, 274, 275*t*, 277, 278, 279
MammaHealth search service, 155
MammaPrint, 82, 226, 236, 451*f*, 452, 454
economic evaluation, 236
Managed care, 235
Mannose-binding lectin (MBL), 745
complement activation, 745
mycobacterial disease susceptibility, 754
pathogen recognition/signaling, 764
response to *Neisseria meningitidis*, 752
response to *Streptococcus pneumoniae*, 750
Mantle cell lymphoma, 415
diagnosis, 415
gene expression profiles, 415
prognosis, 415
Manual microdissection, tissue sampling for microarray experiments, 75
MAOA *see* Monoamine oxidase
MAOB gene, 708
MAPK *see* Mitogen-activated protein kinase
Maple syrup urine disease, newborn screening, 102
MAPT see Microtubular-associated protein tau
Marfan syndrome (*FIBRILLIN-1*), congenital heart disease, 393–394
Marimastat, 544
MASCOT, 92
MASP2 gene, 769
Mass spectrometry
biomarker identification
glycopeptide capture, 136
isotopic tagging for relative and absolute quantification (iTRAQ), 136
protein mixture analysis, 135
instruments, 92, 93*t*
metabolic profiling, 98, 99–100, 99*t*
cultured cell regulatory/signaling mechanisms, 102–103
disease mechanisms, 101–102
targeted profiling methods, 100, 101–105
newborn screening, 193
proteomics, 89, 90, 91–92, 93*f*, 94
brain tumors, 527
colorectal cancer, 467
ovarian cancer, 495
pancreatic cancer, 504–505
prostate cancer, 481, 482
rheumatoid arthritis, 557
software packages, 92–93
modified peptide identification, 93–94
see also Tandem mass spectrometry (MS/MS)
Massively parallel sequencing-by-synthesis, 33–34
Mastocytosis, peptic ulceration, 627
MAT1, hepatic stellate cells, 651
Matrix-assisted laser desorption ionization (MALDI), proteomics
brain tumors, 527
ovarian cancer, 495
Matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS)
proteomics
colorectal cancer, 467
Helicobacter pylori infection markers, 634
prostate cancer, 481, 482
sarcoid bronchoalveolar fluid, 619
tumor-specific blood biomarkers, 94
Matrix-comparative genomic hybridization *see* Array comparative genomic hybridization
Matrix metalloproteinase 2 (MMP-2), ovarian cancer metastasis, 497
Matrix metalloproteinase 7 (MMP-7; matrilysin)
head and neck cancer, 517
peptic ulcer disease, 635
Matrix metalloproteinase 9 (MMP-9)
chronic obstructive pulmonary disease, 607, 608
glioblastoma, 527
peptic ulcer disease, 635
Matrix metalloproteinase 11 (MMP-11), Oncotype Dx assay, 533
Matrix metalloproteinase 19 (MMP-19), thrombotic event prediction with anticardiolipin antibodies, 380
Matrix metalloproteinases, 612
atherosclerosis
plaque rupture risk, 290
head and neck cancer, 513
Maturity onset diabetes of the young (MODY), molecular subtype diagnosis, 192
Maturity onset diabetes of the young type 2 (MODY2), glucokinase mutations, 192
Maytansanoid (DM-1), 538
Maze procedure, hypertrophic cardiomyopathy, 345
MCH, obesity, 666
MCP-1 (monocyte chemoattractant protein-1; CCL1)
chronic obstructive pulmonary disease, 607
thrombosis gene linkage studies, 379–380
MDA5, 764
MDR-1 (multidrug-resistance gene 1), corticosteroid pharmacogenomics, 582
MDR1 C3435T genotype, 637
Medical device regulations
analyte-specific reagents (ASR), 227
in vitro diagnostics, 225
laboratory-developed tests, 226
risk-related classification, 225–226
Medical geneticists, 196, 197
Medical genetics
definition, 2
Medical history-taking, 112
Medical Sequencing Project, 210
Medical Subject Headings (MeSH), 121
Medium-chain acyl-CoA dehydrogenase deficiency
newborn screening, 193
MEDLINE searches, 153

MedlinePlus, 153
MedStory, 155
MEF2A (myocyte-enhancing factor 2A), myocardial infarction, 292
Melanoma
  gene expression profiles, 80, 114
Membrame-type-1 matrix metalloproteinase (MT1-MMP), ovarian cancer metastasis, 497
Meningococcus *see* *Neisseria meningitidis*
Mental retardation (learning disability)
  copy number variation (CNVs) associations, 194
  DNA sequencing, 34
  public health genomics, 248
6-Mercaptopurine
  acute lymphoblastic leukemia, 185
  inflammatory bowel disease, 582
  pharmacogenomics, 185, 582
    dose adjustment, 224
    drug labels, 230
    response prediction, 426
Mesothelin, pancreatic mucinous cystic neoplasms, 505
Messenger RNA (mRNA), 16
  amplification for microarray experiments, 75–76
Mestizo population, 211
c-Met, ovarian cancer metastasis, 497
Metabolic disorders
  see also Inborn errors of metabolism
Metabolic profiling, 97–105
  targeted, 98
    disease mechanisms, 101–104
    human disease pathogenesis, 103–104
    mass spectrometry, 100, 101–104
  unbiased, 98
    comparison of methods, 99–100
  see also Metabolomics
Metabolites, 98
  targeted mass spectrometry-based assay modules, 102
Metabolomics, 2, 104
  acute coronary syndromes, 306, 307*f*
  advantages in disease research, 97
  complex disease, 166
  coronary artery disease, 304
  cultured cell regulatory/signaling mechanisms, 102–103
  diagnostic applications, 114
  electronic medical records, 143
  head and neck cancer, 517
  impact on health–disease continuum, 165
  mass spectrometry, 98, 99–100, 99*t*
  measurement technology, 98–99, 99*t*
  newborn screening, 193
  nuclear magnetic resonance spectroscopy, 98, 99–100
    disease research, 101
  public health genomics, 246
  stable isotope-labeled internal standards, 98–99, 98*f*
  see also Metabolic profiling
Metagenomics, 4
Metalloproteome, liver disease, 655
Metaproterenol, asthma management, 596
Metastasis
  colorectal cancer, 458
  genomics
    gene expression profiles, 80
    17-gene expression signature, 484
  ovarian cancer, 496–497
  primary diagnosis, 116, 117*f*
Metavinculin (*VCL*) mutations, hypertrophic cardiomyopathy, 338
Metformin, pharmacogenomics, 681
Methotrexate
  pharmacogenomics, 582
    response prediction, 426
1-Methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) exposure, 700
Methyl-CpG-binding domain (MBD) proteins, 60
Methylation-sensitive restriction enzymes, 61
  DNA methylation marker discovery, 62
Methylation-specific antibodies, 61–62
Methylation-specific polymerase chain reaction (MSP), methylation assessment in tissue samples, 64
Methylation-specific single nucleotide primer extension (MS-SNuPE), tissue sample DNA methylation assessment, 63–64
Methylenetetrahydrofolate reductase (*MTHFR*) polymorphism
  colorectal cancer, 465
  hypertension/nephrosclerosis, 264
  pharmacogenomics
    5-fluorouracil, 469
O6-Methylguanine-DNA methyltransferase (*MGMT*), 69, 528
  cancer-related hypermethylation
    head and neck cancer, 516
    lung cancer, 434, 435
    protate cancer, 67
    tumor treatment response prediction, 68
  glioblastoma
    chemoresistance, 527
MethyLight assay, methylation assessment
  disease screening samples, 65
  protate cancer, 67
  tissue samples, 64, 64*f*
Methylthioadenosine phosphorylase gene, cardiovascular risk, 294
Metronidazole, *Helicobacter pylori* eradication, 628, 635
Mexican Genome Project, 211
Mexiletine, gene-specific long QT syndrome treatment, 360
MGMT see O6-Methylguanine-DNA methyltransferase
MIAT (myocardial infarction associated transcript), 294
MicroArray Quality Control (MAQC) project, 81
Microarrays, 73–84, 449–450
  brain tumor genomic mapping, 524–525
  clinical expression profiling, 82–83
    see also Gene expression profiles
  data analysis, 76–77
    bioinformatics, 113, 114
    mining, 76–77
    supervised/unsupervised learning, 76, 77*f*
  data repositories, 76, 119–120
  diagnostic applications, 78, 112, 113–114
  diffuse large B-cell lymphoma subtype differentiation, 410
  disease classification, 78, 114, 117
  DNA methylation marker discovery, 62
  DNA sequencing by hybridization, 34
  drug target discovery, 83
  heart failure monitoring, 319
  Helicobacter pylori genomics, 629
  hepatic stellate cell studies, 651
  histopathology applications, 116
  hypertrophic cardiomyopathy, 341–342
  limitations, 80
  MIAME data standards, 76
  normal tissue taxonomy, 77–78
  platforms, 73–75, 74*f*
    new technologies, 81
  prognostic applications, 78–79, 114
  prostate cancer, 479
  reproducibility, 81
  sample preparation, 75–76
    RNA isolation, 75–76
    tissue sampling, 75
  transcriptomics *see* Transcriptomics
Microbial polymorphisms, 636
Microbiome, 4
Microdissection-based genotyping, pancreatic cancer diagnosis, 504
Microfluidic cell sorting techniques, 134
Microfluidic chips, 135
MicroRNAs (miRNAs), 16, 428, 470, 497–498, 507, 528, 655
  brain tumors, 527
  chronic obstructive pulmonary disease, 611–612
  expression profiling, 82
  hypertrophic cardiomyopathy, 342
  leukemia, 428
  lung cancer, 439–440
Microsatellite instability
  colorectal cancer, 65, 458, 459–460, 468
    hereditary nonpolyposis colorectal cancer, 463, 467
  pancreatic cancer, 501
  pancreatic endocrine tumors, 506
Microsatellite markers, 266, 464, 504
  head and neck cancer screening, 512

rheumatoid arthritis, 552
Microsoft Live Search, 155
Microsomal expoxide hydrolase polymorphism, chronic obstructive pulmonary disease, 604
Microsomal transfer protein (MTP)
inhibitors, 282
mutations, abetalipoproteinemia, 274
Microsomal triglyceride transfer protein (MTO), 270
Microtubular-associated protein tau (*MAPT*)
frontotemporal dementia, 691, 692*f*
MIF (macrophage migration inhibitory factor), sepsis, 764, 767
MIG (CXCL9), hepatitis C infection, 780
MINDACT (Microarray in Node negative Disease may Avoid ChemoTherapy), 2, 82, 450, 454
Mineral oil exposure, rheumatoid arthritis risk, 555
Minimal cognitive impairment, 687
Minimum Information About a Microarray Experiment (MIAME) standards, 76, 319
Mining, gene expression data analysis, 76–77
minK/KCNE1 (*LQT5*), 353, 354, 356
MiR-7, 528
MiR-21, 528
MiR124, 528
MiR-128, 528
MiR-296, 528
MiRNA-21, 470
MiRP1 (*LQT6*), 355, 356
Mitochips, mitochondrial genome sequencing, 194
Mitochondrial cytochrome c, cardiomyocyte apoptosis in heart failure, 316
Mitochondrial DNA mutations, optic nerve disorders, 716
Mitochondrial genome
damage, Parkinson's disease, 705
sequencing, 194
variation, 194
Mitochondrial trifunctional protein (MTP) defect, newborn screening, 102
Mitogen-activated protein kinase (MAPK)
cardiomyocytes in heart failure, 320
chronic obstructive pulmonary disease, 608
hepatic stellate cells, 650
lung cancer, 438
see also p38 MAPK
Mitral valve prolapse/regurgitation, 393
Mitrazapine, pharmacogenetic studies, 179
MLH1
colorectal cancer, 459
methylation aberrations, 65, 460
mutations, 65
hereditary nonpolyposis colorectal cancer, 463, 467
lung cancer, promoter hypermethylation, 434
promotor methylation status, tumor treatment response prediction, 67
ulcerative colitis severity associations, 580–581
MLL
acute lymphoblastic leukemia, 424
acute myeloid leukemia, 423, 424
MMP-9, 612
MMRPRO, 465
Modification of Diet in Renal Disease (MDRD), 264
Modifiers, pharmacogenetic test clinical validity, 182
Modular data analysis, 76
Molecular concept map, prostate cancer, 484
Molecular function, Gene Ontology (GO), 121
Molecular signature classifiers
systems medicine, 137
Monoallelic gene expression, 20
Monoamine oxidase (MAOA), 706, 707
depression, 724
prostate cancer, 484
antidepressant drug response effects, 727*t*
Monoclonal antibody therapy, cancer, 535–540, 536*t*
hematologic malignancies, 538
solid tumors, 538–540
Monocrystalline iron-oxide nanoparticle (MION) contrast agents
cell surface receptor imaging, 227
Monocyte chemoattractant protein-1 *see* MCP-1
Montelukast, asthma management, 596
MPTP exposure, Parkinson's disease, 704
MRP8 (calgranulin A)
follicular lymphoma, 413
rheumatoid arthritis, 557
MRP14
follicular lymphoma, 413
MRT2C, obesity, 666
MSH2
colorectal cancer, 459
hereditary non-polyposis colorectal cancer, 463, 467
promoter hypermethylation in lung cancer, 434
MSH6
colorectal cancer, 459
hereditary nonpolyposis colorectal cancer, 463, 467
MSR1, prostate cancer, 477, 478
MST1, 583
MTHFR see Methylenetetrahydrofolate reductase polymorphism
MUC1, prostate cancer, 484
MudPIT (multidimensional protein identification technology), quantitative proteomics, 91
Muir-Torre syndrome, 463
Multicenter Automatic Defibrillator Implantation Trial (MADIT), 360
Multidimensional protein identification technology (MudPIT), 95
Multiexperiment Viewer (MeV), 122, 122*f*
Multifactor Dimensionality Reduction (MDR), 181
Multiparameter blood-borne biomarkers, 137
Multiple endocrine neoplasia type 1 (MEN1), 484
peptic ulceration, 627
Multiple myeloma, 401
bortezomib therapy, 543
Multiple Risk Factor Intervention Trial (MRFIT), 264
Multiple sclerosis, 566–572
biomarkers
identification, 570–571
Epstein–Barr virus association, 570
gene expression profiles, 78, 568–569
autoimmune disorder comparisons, 569
lesion characterization, 568
relapse, 569
reponse to therapy, 569
genomics, 566
admixture mapping, 567
association studies, 567
linkage mapping, 567
whole genome scans, 567
immune response high-throughput analysis, 568, 569–570
B cell response, 570
T cell response, 569–570
ion channel function, 570
pharmacogenetic research, 177
proteomics, 570–571
transcriptomics, 568–569
Multiplex ligation-dependent probe amplication (MLPA), congenital heart disease, 396
Muscle
lipoprotein metabolism, 270
Muscular dystrophies
newborn screening, 193
Muscular subaortic stenosis, 337
Mut Y Homologue gene (MYH)-associated polyposis, 461, 463
Mutagenesis and phenotypic screens, congenital heart disease gene discovery, 395–396
Mutation Surveyor, 32
Mx-1, hepatitis C infection, 651–652
MYBPC3 mutations (myosin-binding protein C), hypertrophic cardiomyopathy, 338, 340, 341
MYC
acute lymphoblastic leukemia, 421
lung cancer, 438
ovarian cancer, 496
prostate cancer, 480

Mycobacteria
genetics of host response, 753–754
host genetic susceptibility, 749*t*, 762
host recognition, 747
Mycobacterium leprae infection *see* Leprosy susceptibility
Mycobacterium tuberculosis
catalase-peroxidase, sarcoidosis involvement, 619
see also Tuberculosis
MyCode biorespository, 146–147, 147*f*
Mycophenolate mofetil, 327
Mycoplasma genitalium genome, 33
MyD88, 747, 749
host infection response
Legionella, 752
Listeria monocytogenes, 751
Pseudomonas aeruginosa, 751
Staphylococcus aureus, 750
Streptococcus pneumoniae, 749, 750
Myeloid-related protein 8 *see* MRP8 (calgranulin A)
Myeloperoxidase
acute coronary syndromes
diagnostic biomarker, 306
risk stratification, 309
peptic ulcer disease, 634
MYH6 mutations (alpha myosin heavy chain), hypertrophic cardiomyopathy, 338
MYH7 (beta myosin heavy chain) mutations, hypertrophic cardiomyopathy, 313, 336, 338, 340, 341
MYL2 mutations (regulatory myosin light chain), hypertrophic cardiomyopathy, 338
MYL3 mutations (essential myosin light chain), hypertrophic cardiomyopathy, 338
Mylotarg *see* Gemtuzumab ozogamicin
MYO9B, inflammatory bowel disease, 578
MYOC (myocilin), glaucoma/intraocular pressure elevation, 716
Myocardial infarction, 289–290
association studies, 291–294
genome-wide scans, 290–291, 290*t*
leukotriene pathway, 294
proprotein convertase subtilisin/kexin type 9 serine protease (PCSK9), 294
specific genes, 291, 291*t*
chromosome 9p21 region of interest, 294, 304
diagnosis
protein biomarkers, 305–306, 306*t*
ST-segment/non-ST-segment elevation, 305
drug-eluting intracoronary stents, 297
environmental risk factors, 295
genetic influence on risk, 289–290
heart failure, 296, 314, 315*f*
leukotriene inhibitor therapy,297
pharmacogenomics, 296
post-infarction therapy response, 296
prognostic implications, 296
risk factors *see* Cardiovascular risk
risk stratification, 296
ST segment elevation, 305, 314
ventricular remodeling, 296, 316, 316*f*
see also Acute coronary syndromes
Myocardial infarction associated transcript (MIAT), 294
Myocyte-enhancing factor 2A (MEF2A), myocardial infarction, 292
Myocytes, fatty acids metabolism, 270
Myofilament-associated gene mutations, hypertrophic cardiomyopathy, 338
Myosin-binding protein C (*MYBPC3*) mutations, hypertrophic cardiomyopathy, 338, 340, 341
Myozenin 2 (*MYOZ2*) mutations, hypertrophic cardiomyopathy, 338

N
NAHR, 56
Naip5, *Legionella* susceptibility, 752
Nanoparticle biosensors
head and neck cancer screening, 513
Nanosphere Verigene Warfarin Metabolism Nucleic Acid Test, 183
National Center for Biotechnology Information (NCBI)
Gene Expression Omnibus (GEO), 119, 119*f*, 120
Gene identifier, 121
Taxonomy, 121
UniGene identifier, 121, 648
National Coalition for Health Professional Education in Genetics (NCHPEG), 197, 216
National Genetics Education and Development Centre, 216
National Institute for Health and Clinical Excellence (NICE), 235, 239
National Lung Cancer Screening Trial (NLST), 434
National Office of Public Health Genomics, 249
Natural killer cells
hepatitis C innate immune response, 780
Navigenics, 198
NCBI *see* National Center for Biotechnology Information (NCBI)
NCI-60 cancer cell lines, 441–442
NCI60 database, gene microarray measurements–drug susceptibility relationships, 114–115, 115*f*
Neimann-Pick disease, pulmonary fibrosis, 617, 620
Neisseria gonorrhoea
genome mapping, 744
Neisseria meningitidis (meningococcus), 744
genetics of host response, 752–753
sepsis, 765
susceptibility
genetic factors, 762, 763
Neonatal hyperbilirubinemia, 653
Neonatal screening *see* Newborn screening
Nestin, hepatic stellate cells, 649
NET1, ovarian cancer, 495
Network discovery, 118
Neurodevelopmental disorders, copy number variation (CNVs)
associated syndromes, 52
clinical cytogenetic diagnostics, 54
database, 48
Neurofibromatosis
brain tumor predisposition, 523
pulmonary fibrosis, 617, 620
Neurofilaments, multiple sclerosis biomarker, 570
Neuromodulin, hepatic stellate cells, 649
Neuropeptide Y (NPY)
body weight regulation, 662
obesity, 666
Neutrophil elastase
emphysema inflamatory mechanisms, 607
New and improved human reference genome, development of, 57
New chemical entities (NCEs), 168
see also Drug development
Newborn screening, 112, 193
Newcastle disease virus
glioma cancer vaccine, 527
NF-κB pathway, 559
chronic obstructive pulmonary disease, 607–608
diffuse large B-cell lymphoma therapeutic target, 411–412
primary mediastinal large B-cell lymphoma, 411–412
sepsis-related signaling, 763, 764
NHE9, 56
NHEJ, 56
NHGRI, 170
NHS National Genetics Education and Development Centre, 250
Nicotine addiction, 604
Nicotine replacement therapy, 610
Nicotinic acid (niacin), 281
Nifidipine, hypertrophic cardiomyopathy, 343
NIH, 171
NimbleGen oligonucleotide arrays, 50
Nip3, hepatic stellate cells, 651
Nitric oxide
chronic obstructive pulmonary disease, 607
hepatic stellate cell production, 650
Nitric oxide synthase
chronic obstructive pulmonary disease, 607
Nitrosamine carcinogens, head and neck cancer, 512
NKX2-3, 583
NKX2-5

congenital heart disease, 391, 392
microarray analysis, 395
NKX3.1, prostate cancer, 478
NLGN4, 56
Nod1/Nod2, host response to *Streptococcus pneumoniae*, 750
NOD2
host infection response
Listeria monocytogenes, 751
Staphylococcus aureus, 751
NOD-like receptors (NLRs), 745, 747
caspase recruitment domains (CARD), 763
pathogen recognition/signaling, 763–764
Nomenclature standards, 167
Non-alcoholic fatty liver disease, 652
proteomics, 653
transcriptomics, 652
Non-coding RNAs (ncRNAs), 16, 22
Non-Hodgkin lymphoma, 401
^{131}I-tositumomab treatment, 538
Non-homologous end joining, copy number variation generation, 47–48, 49*f*
Non-parametric tests, 118
Non-small-cell lung carcinoma, 435
TNM staging, 436
Non-steroidal anti-inflammatory agents (NSAIDs)
cancer chemoprevention
colorectal cancer, 468
peptic ulcer disease
mechanisms, 628
susceptibility polymorphism, 365
Noonan syndrome (*PTPN11*), congenital heart disease, 393
Norepinephrine neurotransmission depression, 723
Northern blotting, 73
NOS3, atrial fibrillation, 365
Nosology (disease ontology), 116–117, 120–121, 120*t*
NOTCH1 (bicuspid aortic valve), 393
NOTCH2 (Alagille syndrome), congenital heart disease, 393
NOTCH3 mutations, CADASIL, 694
Notch pathway genes
astrocytomas, 526
primary biliary cirrhosis, 652
Novartis, 543
novoSNP, 32
NPM1, acute myeloid leukemia, 423, 424
NPSR1, asthma, 593
NR3C1, obesity, 666
NRAMP1 *see* SLC11A1
NRTK2 mutations, obesity, 665
NRXN1, 56
NSABP B-13, 446
NSABP B-14, 446
NSABP B-21, 446
NSAIDS, 637
NSC 624004 anti-cancer agent, lymphocyte cytosolic protein-1 (*LCP1*) susceptibility association, 115
NSCLC cell lines, 441
NTN1, multiple sclerosis, 567
Nuclear erythroid-2 related factor 2 (Nrf2), 613
Nuclear magnetic resonance spectroscopy, metabolic profiling, 98, 99–100
cultured cell regulatory/signaling mechanisms, 102–103, 103*f*
disease research, 101
Nuclear myocardial perfusion imaging, acute coronary syndrome diagnosis, 306
Nuclear receptor coactivator-2 (NCOA3) *see* AIB1 (SRC3)
NUGENOB, 671
Number of genes, 14
Nurse geneticists, 216
Nurses' Health Study, 295
Nutritional genomics, 671

O

Obesity, 191, 661, 672
animal models, 663
causes, 661–663
colorectal cancer risk, 464
definition, 661
diabetes type 2 association, 676, 678
diagnosis, 668–669, 669*t*
environmental factors, 662, 663, 666
family studies, 663
gene-based treatments, 669–671
genetic factors, 661–663, 663*f*
adipocyte metabolism, 668
appetite regulation, 667
association studies, 664, 666
candidate genes, 665*t*, 666, 666*t*
characterization, 664–667
chromosomal aberrations, 666
energy expenditure regulation, 667
food intake control, 671
gene map, 664, 666
identification, 663–664
linkage studies, 666, 667
single-gene disorders, 664, 665
syndromic associations, 667*t*
weight reduction program response, 669, 670*t*
health consequences, 663–664
measurement, 664
metabolic profiling, 103
nuclear magnetic resonance spectroscopy, 101
prognosis, 669–671
regional distribution of weight gain, 661
rheumatoid arthritis risk, 556
screening, 668–669
thrifty gene theory, 661, 663
OCT1, metformin pharmacogenomics, 681
Ocular disorders, 714*t*
Oculomotor nerve disorders, 713
Office of Genomics and Disease Prevention, 381
Office of *In vitro* Diagnostic Device Evaluation and Safety (OIVD), 225
OipA, *Helicobacter pylori* virulence factors, 633
Okhiro syndrome (*SALL4*), congenital heart disease, 392
2′-5′ Oligoadenylate synthetase, hepatitis C infection induction, 780
Oligodendroglioma, 522, 523*t*
chromosomal alterations, 525
gene expression profiles, 525, 526
genetic alterations, 523, 524*f*
pharmacogenomics, 525
progression, 522, 523
Oligonucleotide microarrays, 73, 74–75, 74*f*
copy number variation (CNVs) evaluation, 194
disease diagnosis/classification, 78
hybridization detection, 74–75
ovarian cancer, 495
Omega 3 fatty acids, 281
Omeprazole, *Helicobacter pylori* eradication, 635
"Omics", 2
Oncogenes
cancer
expression signatures, 117, 117*f*
head and neck cancer, 511
heart failure, 320
lung cancer, 434, 438, 438*t*
ovarian cancer, 495–496
pancreatic cancer, 501
Oncotype Dx assay, 82, 215, 235, 453, 454, 533
economic evaluation, 236, 237
TAILORx, 82, 210, 236, 453
tamoxifen response prediction, 533
Online health information
consumer retrieval, 151–168
information sites, 152–154, 153*t*, 154*t*
quality assurance, 154
search characteristics, 151–152
personalized genomics, 154–155, 155*t*
Online medical libraries, 153
Online Mendelian Inheritance in Man (OMIM), 205, 390, 665
Ontologies
consumer on-line health information retrieval, 152
diseases, 120–121, 120*t*
genomic, 121
OPA1 mutations, Kjer's autosomal dominant optic atrophy, 716
Open Biomedical Ontologies (OBO), 121
Opt-out model, 148–149
Ophthalmology, 712
Optic nerve, 712
genetic disorders, 716
Oral cancer *see* Head and neck cancer
Oral leukoplakia, 512

Orbitrap mass spectrometer with electron transfer dissociation (ETD), 95
Orbitz, 156
Orexins
  body weight regulation, 662
  obesity, 667
Organic acids, targeted metabolic profiling, 102
ORMDL3, 597
Osaka Acute Coronary Insufficiency Study, 296
Osteopontin
  alcoholic liver disease, 652
  hepatic stellate cells, 650
  multiple sclerosis lesions, 568
Osteoporosis, 57
Ovarian cancer, 493–498
  angiogenesis, 497
  BRCA1/2, 191–192, 196, 493
  chemotherapy response prediction, 495
  DNA repair defects, 494
  epigenetics, 496
  gene expression profiling, 495
  genetic alterations, 494, 495*t*
  genomic testing, 164
  growth factors, 495–496
  hereditary nonpolyposis colorectal cancer (Lynch syndrome), 463
  high-risk criteria, 494, 494*t*
  inherited syndromes, 493–494
  metastasis, 496–497
    spheroids, 497
  oncogenes, 495–496
  origin/pre-malignant lesions, 493
  prevention, 494
  prognostic molecular signatures, 137
  proteomics, 495
  screening, 494
  somatic mutations, 494–495
  tumor suppressor genes, 496
Oxygen therapy, chronic obstructive pulmonary disease, 611

P

1p chromosome, 528
P2Y12 *see* Platelet ADP receptor polymorphism
P3/P4 medicine, 132, 201
  see also Prospective health care
P3G Consortium, 248
P3G Observatory, 248
9p13/*CCL21*, 558
P13 kinase/Akt pathway
  gliomas, 523
  lung cancer pathogenesis, 438
  ovarian cancer, 495, 496
  therapeutic targeting
    brain tumors, 527
p14^{ARF}
  gliomas, 523, 524
  mantle cell lymphoma, 415
  methylation in colorectal cancer, 460
10p15/*PRKCQ*, 558
p16/INK4A
  cancer-related aberrant methylation, 460, 496, 516
  cardiovascular risk associations, 294, 304
  colorectal cancer, 460
  gliomas, 523, 524
  head and neck cancer, 511, 515, 516
  lung cancer, 439
  mantle cell lymphoma, 415
  ovarian cancer, 496
  pancreatic cancer, 501, 502
p21-ARC, laryngeal carcinoma, 517
9p21 chromosome, 298, 365
p21SNFT, Hodgkin lymphoma/Reed-Sternberg cells, 413
p27, hepatitis C infection, 651
1p36/*TNFRSF14*, 558
p38 MAPK
  follicular lymphoma therapeutic target, 414
p53
  breast cancer, 448
  colorectal cancer, 65, 458, 460, 467, 468
  fludarabine pharmacogenomics, 424–425
  gliomas, 523, 524
  head and neck cancer, 511, 515
    cisplatin resistance, 516
  hepatitis B infection, 651
  Li-Fraumeni syndrome, 458
  lung cancer, 439
  ovarian cancer, 494, 496
  pancreatic cancer, 501, 502
  prostate cancer, 480
  PTEN cooperative interactions, 480
Pacemaker implantation
  Lev-Lenegre progressive cardiac conduction disease, 362
  long QT syndromes, 359
  sinus node dysfunction/sick sinus syndrome, 362
Paclitaxel
  drug-eluting intracoronary stents, 297
PAD14, rheumatoid arthritis, 554, 556
Paget's disease, heart failure, 314
PAI-1 *see* Plasminogen activator inhibitor-1
Pancreatic β-cells, glucose-stimulated insulin secretion, 102–103
Pancreatic cancer, 501–507
  antibody microarray serum protein profiling, 135
  carcinogenesis/progression, 501, 502*f*
  development, 507
  diagnosis, 503–505, 507
    microdissection-based genotyping, 504
    proteomics, 504–505
  DNA repair gene expression, 506
  erlotinib therapy, 542
  familial, 502
  hereditary syndromes, 502
  molecular analysis, 501
  precursor lesions, 501
  prognosis, 505–506, 507
  risk factors, 502
  screening, 502–503
    mutational analysis, 503
Pancreatic carcinoma *in-situ*, 501
Pancreatic cystic neoplasms, 501
  diagnosis, 505
  gene expression profiles, 505
  malignant potential, 504
Pancreatic endocrine tumors, 507
  prognosis, 506
Pancreatic intraepithelial neoplasia (PaIN), 501
Pancreatitis, chronic, pancreatic cancer risk, 502
Panitumab, 543
Panitumumab, 231
Papillary thyroid carcinoma, 506
Paralogs
  Helicobacter pylori genome, 629
Parametric tests, 118
Paraquat exposure, Parkinson's disease, 704
PARD3, 584
PARK1 (*SNCA* α-synuclein), 702
PARK2 (*PRKN*; PARKIN), 702–704, 703*f*
PARK4, 702
PARK6 (*PINK1*; Pten induced kinase 1), 704
PARK7 (*DJ-1*), 704
PARK8 (*LRRK2*; dardarin), 704, 708
Parkinson disease with dementia, 692–693, 702
  α-synuclein aggregates, 692
Parkinsonian syndromes, 700
Parkinson's disease, 700–701
  causes, 700
  cell-based therapy (fetal dopaminergic neuron transplantation), 706–707
  clinical features, 700–702, 701*t*
  dopamine replacement therapy
    levodopa, 700, 706
    metabolic pathways, 706, 706*f*
  environmental factors, 700, 702, 704
  genetic approaches to, 708
  genetics, 700, 701*t*, 702–705
    candidate genes, 705
    DJ-1 (*PARK7*), 704
    LRRK2 (*PARK8*; dardarin), 704–705, 708
    microarray experiments, 34
    PINK1 (*PARK6*; Pten induced kinase 1), 704
    PRKN (*PARK2*; PARKIN), 702–704, 703*f*
    SNCA (*PARK1; PARK4; α-synuclein*), 702
    sporadic cases (typical Parkinson's disease), 705–706
  genomic approaches to, 708
  Lewy bodies, 700, 703
  surgical interventions, 706
    basal ganglion neuronal pathways, 707*f*
    deep brain stimulation, 706
  treatment, 706–707
    future role of genetics, 707–708
Parkinsonian-pyramidal syndrome (PPS), 708

Paroxetine, pharmacogenetics, 181
Parvovirus, idiopathic dilated cardiomyopathy, 314
Patent ductus arteriosus, 393
Patenting
  genes/genomes, 217
    public health genomics, 248
  genomic technology, 248
Pathogen-associated molecular patterns (PAMPs) *see* Pathogen recognition receptors
Pathogen–host interactions, 733–734
  host genetic susceptibility *see* Infectious disease
Pathogen recognition receptors (PRRs; pathogen-associated molecular patterns), 745, 763, 766
Pathogen recognition/signaling, 763–764, 766
  CD14, 764
  mannose-binding lectin, 764
  NOD-like receptors, 763–764
  RIG-like receptors, 763–764
  Toll-like receptors, 763
Pathogen virulence, 733–734
Pathogenicity islands
  Helicobacter pylori, 629, 630
PAX3
  primary biliary cirrhosis, 652
PAX5, 428
  follicular lymphoma, 413
  Hodgkin lymphoma, 413
PAX6, ocular development, 712
PCA3, prostate cancer urine biomarkers, 481
PCDH10, 56
PCSK1 mutations, obesity, 665
PCSK9 *see* Proprotein convertase subtilisin/kexin type 9
PCV chemotherapy, brain tumor pharmacogenomics, 525, 527
PECAM-1 (platelet endothelial cell adhesion molecule 1)
  astrocytomas, 526
Penetrance, pharmacogenetic test clinical validity, 182
Peptic ulcer disease, 627–636
  applications of genomics, 633–635
  Helicobacter pylori, 627, 628, 628*f*, 629
    eradication, 628, 635
    genomic markers, 634
    geographic strain differences, 634
    proteomic markers, 634
    serological markers, 634–635
  historical management, 627–628
  molecular diagnostic techniques, 636
  non-steroidal anti-inflammatory agent (NSAID)-induced, 629
  pathophysiology, 628
    ulcer formation, 628–629
  susceptibility polymorphisms, 635
PeptideAtlas, 120
Peptidoglycans, TLR2 recognition of pathogens, 750
Perinuclear antineutrophil cytoplasmic antibodies (pANCA)
  inflammatory bowel disease, 578, 579, 560
  primary sclerosing cholangitis, 560
Peripheral blood mononuclear cells, microarray analyses
  cardiac allograft rejection, 322, 328–330, 330*f*
Peripheral T-cell lymphomas (PTCL), 416
Permanent neonatal diabetes, 681
Peroxisome proliferator-activated receptor α (PPARα) agonists *see* Fibrates
Peroxisome proliferator-activated receptors (PPARs)
  pancreatic cancer, 505, 506
Personal genomics, 6, 170, 172
Personalized genomics for consumers, 154–155
  direct-to-consumer initiatives, 154
  incidental findings, 155
  online information, 155*t*
Personalized Health Care, 205, 217
Personalized medicine, 2
  colorectal cancer risk assessment, 465, 466*f*
  genetic stratification of disease, 191–192
  health risk assessment, 203
  hemostasis, 380
  implementation issues *see* Translational genomics
  thrombosis, 380
  see also Prospective health care
Pescadillo, head and neck cancer, 517
Peter's anomaly, PAX6 mutation, 712
Peutz-Jeghers syndrome
  breast cancer, 448
  LKB1 pathway mutations, 464
  pancreatic cancer, 502
PGx testing, 186–187
Pharmaceutical Benefits Advisory Committee (PBAC), 235
Pharmacodynamics
  pharmacogenetic markers, 180–181
Pharmacogenetics, 114, 175–176, 728
  candidate gene approaches, 179–180
  case–control studies, 177–178, 179*f*
  clinical utility of analytic testing, 182–183
  cohort multiple dosing studies, 179
  copy number variation (CNVs), 54
  definition, 168, 175
  epistatic associations, 181
  familial studies, 178–179
  FDA approved biomarkers, 184*t*
  future developments, 185–186, 186*t*
  HapMap markers, 181–182
  health professionals' role, 215
  healthy volunteer challenge studies, 179
  objectives, 175, 177
  pathway-based analyses, 181
  pharmacokinetic versus pharmacodynamic markers, 180–181
  placebo response effects, 179
  prospective analyses, 182
  regulatory requirements, 183
  research study design, 177–179, 178*f*
    phenotype definition, 177
    research question definition, 177
  research targets, 176–177, 176*t*
  test validity, 182
  validation of associations, 182
  whole-genome association studies, 179–180
  see also Pharmacogenomics
Pharmacogenomics, 168, 175–176
  antidepressants, 726–727, 727*t*
    drug interactions, 726
  asthma, 594, 596–597
  brain tumors, 527
  chronic obstructive pulmonary disease, 611–612
  colorectal cancer, 468–469
  corticosteroids, 597
  definition, 168
  diabetes, 681
  dopamine metabolism, 707
  drug-diagnostic test combinations
    clinical trial design, 229
    diagnostic test development, 229
    dosing adjustments, 229–230
    drug labels, 229
    economic value, 238
    patient selection/monitoring, 228–229
    safety/effectiveness demonstration, 229, 230
  economic evaluation, 224, 235, 236
  economic incentives, 237–239
  gliomas, 525
  head and neck cancer, 516
  heart failure, 318–319
  Helicobacter pylori eradication, 635
  hemostasis, 384
  hypertension, 264–266
  hypertrophic cardiomyopathy, 343
  impact on health–disease continuum, 164
  inflammatory bowel disease, 582
  leukemias, 424–426
  myocardial infarction, 296–297
  oligodendrogliomas, 525
  population genomic screening, 193
  principles, 175–176
  prostate cancer, 484–485
  regulatory issues, 183, 224, 227–230, 238
  rheumatoid arthritis, 557
  thrombosis, 381
  translational barriers, 228
  viral hepatitis, 782
  Voluntary Genomic Data Submission (VGDS), 183, 224
  see also Pharmacogenetics

Pharmacokinetics
pharmacogenetics
markers, 180–181
Phase II trials
drug-diagnostic test combinations, 229
Phase III trials
drug-diagnostic test combinations, 229
Phenotypic plasticity, 20
Phenylalanine hydroxylase deficiency, newborn screening, 102
Phenylketonuria, 246
newborn screening, 102, 193
Phenytoin
metabolism, 726
pharmacogenomics, 181
PHF11 (PDH finger protein 11)
asthma, 593
chronic obstructive pulmonary disease, 607
Philadelphia chromosome, 421
Phosphatidylethanolamine *N*-transferase, alcoholic liver disease, 652
Phosphatidylinositol 3-kinase/AKT pathway, 442
Phospholamban (*PLN*) mutations, hypertrophic cardiomyopathy, 338
Phospholipase A2, hepatocellular carcinoma, 653
Phospholipid transfer protein (PLTP), 271
Phospholipids, targeted metabolic profiling, 102
Photolithography, 41
microarray generation
oligonucleotide microarrays, 74, 74*f*
PHOX2A, strabismus, 713
PHOX2B, 583
Phusion, 91
Phylogenetic analysis, blood coagulation proteins, 375*f*, 376*f*
Pilot trials, pharmacogenetic tests, 183
PiMZ/PiMM, chronic obstructive pulmonary disease, 604
PINK1 (*PARK6*; Pten induced kinase 1), 704
Pirbuterol, asthma, 596
PITX2
DNA methylation, breast cancer prognosis, 67
ocular development, 712
PLA2 polymorphism, aspirin resistance, 297
Placebo response, pharmacogenetic studies, 179
Placental growth factor, acute coronary syndromes risk stratification, 309
Plasminogen activator, genetic variation, 376
Plasminogen activator inhibitor-1 (PAI-1)
polymorphism
obesity, 668, 671
sepsis, 765
Platelet ADP receptor (P2Y12) polymorphism
aspirin resistance, 296, 297
clopidogrel resistance, 297
Platelet-derived growth factor (PDGF)
hepatic stellate cell production, 650
ovarian cancer angiogenesis, 497
Platelet-derived growth factor (PDGF) antagonists, cirrhosis, 648
Platelet-derived growth factor receptor (PDGFR)
gliomas, 522, 524
sorafenib inhibition, 542
Platelet endothelial cell adhesion molecule 1 *see* PECAM-1
Platelet glycoprotein IIb–IIIa receptor polymorphism
aspirin resistance, 296, 297
Platelets
activation, 378
biomarkers
arterial thrombosis risk (Bloodomics project), 383
RNA expression profiling, acute coronary syndromes, 308
Platinum compounds, DNA methylation profiles in tumor response prediction, 67
PLAU-promoter DNA methylation, breast cancer prognosis, 67
PLN mutations (phospholamban), hypertrophic cardiomyopathy, 338
PML-RAR fusion protein
acute promyelocytic leukemia, 422
PMS1
colorectal cancer, 459
hereditary nonpolyposis colorectal cancer (Lynch syndrome), 463, 467
PMS2
colorectal cancer, 459
hereditary nonpolyposis colorectal cancer (Lynch syndrome), 463, 467
POLG2, cirrhosis, 647
Policy issues, 209–217, 209*f*
biobanking, 212
data disclosure, 212–213
direct-to-consumer marketing, 214
genetic discrimination, 216–217
genetic exceptionalism avoidance, 246
genomic database access, 213
genomic medicine
access, 214–215
integration into healthcare, 213–217
health professional education, 215
intellectual property, 217
medical privacy, 216
public health genomics, 249
service development/evaluation, 248–249
reimbursement, 214–215
research, 210–211
allocation/prioritization, 210
large-scale studies, 211–213
race as analytical variable, 210–211
results disclosure to participants, 212
PolicyDB database, 248
Polony multiplex analysis of gene expression (PMAGE)
hypertrophic cardiomyopathy, 342
Polyacrylamide gel electrophoresis, two-dimensional (2D-PAGE)
proteomics
ovarian cancer, 495
prostate cancer, 482
quantitative, 90, 90*f*
Polymerase chain reaction
amplicon re-sequencing methodology, 31–33
DNA sequencing by synthesis, 33
hepatitis B DNA quantitation, 778*t*
lymphoma, 409
microdissection-based genotyping, pancreatic cancer diagnosis, 504
mRNA amplification for microarray experiments, 75–76
Reed–Sternberg cells, 412
see also Real-time polymerase chain reaction; Reverse transcription–polymerase chain reaction
Polyphred, 33
Population-based registries
see also Biobanks
Population-based studies
genomic screening, 193
Population genomics, 18, 19
public health approaches, 245
Population sampling
electronic medical records utilization, 146
Population structure, 146
genome-wide association studies, 40
Positional cloning
diabetes type 2, 677
inflammatory bowel disease, 576
Positron emission tomography (PET)
head and neck cancer diagnosis, 513
in vivo molecular diagnostics, 137
Parkinson's disease, 702, 707–708
Positron emission tomography-computerized tomography (PET-CT)
head and neck cancer
diagnosis, 513
post-treatment neck dissection planning, 517
Post-translational modification, 89, 377
Posterior chamber, 712
Power analysis
genome-wide association studies, 39
PPARα agonists *see* Fibrates
PPARG
diabetes type 2, 677
obesity, 666
PPARs *see* Peroxisome proliferator-activated receptors
PPMS, 571
Prader–Willi syndrome, 664
Pravastatin, 296
Pre-analytical variation
genetic tests, 182

Predictive factors
prospective health care, 203–205
models, 203, 204*f*
Predictive healthcare, 165
see also Prospective health care
Predictive value
pharmacogenetic tests, 182
Prednisolone, pharmacogenomics, 426
Premarket Approval Application, *in vitro* diagnostics, 225–226
Premarket Notification Submission, *in vitro* diagnostics, 225, 226
Prenatal testing
current levels, 213
Presenilin 1/2 (*PSEN1/2*) mutations, Alzheimer's disease, 689
Prevalence of specific phenotype, pharmacogenetics test clinical validity, 182
Prevention, 165, 201
genotypic versus phenotypic interventions, 246
population genomic screening, 193
public health approaches, 245
see also Prospective health care
Primary biliary cirrhosis, 652
proteomics, 654
transcriptomics, 652
Primary effusion lymphoma
Kaposi's sarcoma human herpes virus-8 association, 401
Primary mediastinal large B-cell lymphoma, 411
diagnosis, 412
therapeutic targets, 412
Primary pulmonary hypertension, 314
Primary sclerosing cholangitis, 652
perinuclear antineutrophil cytoplasmic antibodies (pANCA), 580
proteomics, 653
transcriptomics, 652
PRIMER3, 32
Principal component analysis, 118
cancer gene expression profiles, 117
metabolic profiling
cardiovascular disease, 103
mass spectrometry, 100
nuclear magnetic resonance spectroscopy, 100
obesity, 103
PRINTS, 121
Prion disease (spongiform encephalopathies), 133, 134, 687, 693
Privacy fears, of genetic information, 7
Privacy issues, 166
biobank data, 166, 212
genomic information
databases, 213
genetic exceptionalism avoidance, 246
policy, 216
PRKAG2 mutations, hypertrophic cardiomyopathy, 339
PRKN (*PARK2*; PARKIN), Parkinson's disease, 702–704, 703*f*
Procarbazine, oligodendrogliomas, 525
Progesterone receptor assay, breast cancer targeted therapy, 533
Progranulin (*GRN*) mutations, frontotemporal dementia, 691
Progressive cardiac conduction defect *see* Lev-Lenegre progressive cardiac conduction disease
Progressive supranuclear palsy, 700
Proliferating cell nuclear antigen (PCNA), astrocytomas, 526
Promoters, 19
aberrant methylation
head and neck cancer, 516
ovarian cancer, 496
pancreatic cancer, 503
local regulatory variants, 20
Pro-opiomelanocortin (POMC)
body weight regulation, 662
gene mutations, obesity, 664, 665, 666
Properidin gene mutation, *Neisseria meningitidis* susceptibility, 762
Proprotein convertase subtilisin/kexin type 9 (PCSK9)
deficiency, 274
mutation
autosomal dominant hypercholesterolemia, 273
familial hypobetalipoproteinemia, 274
polymorphism
low-density lipoprotein cholesterol reduction, 274, 294, 295 295*f*
myocardial infarction risk, 279, 294, 295*f*
nonsense/missense variants, 294, 295*f*
therapeutic inhibition, 282
Prospective health care, 201, 202*f*, 203*f*
baseline risk assessment, 205
key features, 203
predictive factors, 203–205
predictive models, 203, 204*f*
Prostaglandin E2
ovarian cancer angiogenesis, 497
Prostaglandins, targeted metabolic profiling, 102
Prostate cancer, 19, 477–486
androgen ablation therapy, 486
androgen signaling, 485
biomarkers, 481*f*
circulating tumor cells, 482
serum, 482
urine, 481
chromosomal rearrangements, 483
detection, 480–483
systems medicine approach, 132, 134
epigenetic changes, 478–479
histones acetylation/demethylation, 478
promoter methylation, 67, 478
gene expression profiles
androgen receptor target-genes, 486
biochemical relapse after surgery, 483–484
cancer initiation, 483
correlates with pathology/aggressive disease, 484
literature mining, 484
metastatic progression, 484
response to treatment, 484–485
targeted cytotoxic therapy, 485
genome-wide association studies, 478
genomic changes, 483–484
integrative analysis, 484
response to treatment, 484–485
germline polymorphisms, 478
highly penetrant genes, 477–478
hormone-refractory disease, 485–486
imaging, 482–483
mitochondrial gene mutations, 478
pharmacogenomics, clinical trials, 484–485
prognostic molecular signatures, 137
proteomics, 484
serum protein profiling, 135
screening, 480–481
somatic DNA alterations, 479–480, 479*f*
Myc, 480
p53, 480
PTEN/P13K/mTOR, 479–480
Rb, 480
therapeutic HSP90 inhibition, 83
TMPR22–ERG/ETV1 gene fusion, 77
UGT2B17 copy number variation, 52–53, 53*f*
xenotropic murine leukemia virus-related virus (XMRV) association, 478
Prostate-specific antigen (PSA), 480–481
prostate cancer, 482
Prostate-specific membrane antigen (PSMA), monoclonal antibody targeted therapy, 539, 540*f*
Prostatic intraepithelial neoplasia (PIN), 67
urine biomarkers, 481
Proteasome subunit α type 6 (PSMA6) polymorphism, myocardial infarction, 293
Protein aggregation cardiomyopathy (PAC; desmin-related myopathy), 321–322
Protein C, 375
factor V polymorphism effects, 765
sepsis, 765
Protein convertase 1 (PC1) mutations, obesity, 664
Protein microarrays
blood biomarker analysis, 135
head and neck cancer, 517
multiple sclerosis antibody reactivity, 570
Protein–protein interactions
systems biology approach, 131
Protein Tyrosine Phosphatase 2 (PTPN2) gene, 583

Proteins
  molecular function, Gene Ontology (GO), 121
  post-translational modification, 377
Proteomics, 89–90
  acute coronary syndromes
    diagnosis, 307–308, 307*f*, 308*f*
  alcoholic liver disease, 654
  autoimmune hepatitis, 654
  bioinformatics, 92–94
    data display, 94
    modified peptide identification, 93
  biomarkers, 94
  blood coagulation proteins, 377
  brain tumors, 527
  cholangiocarcinoma, 654
  chronic obstructive pulmonary disease, 609
  colorectal cancer
    screening/surveillance, 467
    treatment response prediction, 469–470
  complex disease, 166
  coronary artery disease screening, 304
  data repositories, 120
  depression, 727–728
  diagnostic applications, 114
  electronic medical records, 143
  gel-based techniques, 90–91
  head and neck cancer, 517
  hepatitis B, 654
  hepatitis C, 654, 655, 781
  hepatocellular carcinoma, 654, 781
  liver, 649
    disease, 653–655
    interferon therapy specific response, 781
  mass spectrometry, 89, 90, 91–92, 93*f*
  multiple sclerosis, 570–571
  non-alcoholic fatty liver disease, 656
  ovarian cancer, 495
  pancreatic cancer diagnosis, 504–505
  primary biliary cirrhosis, 654
  primary sclerosing cholangitis, 654
  prognostic applications, 114
  prostate cancer, 484
    serum biomarkers, 482
    urine biomarkers, 481
  public health genomics, 245
  rheumatoid arthritis, 556
  sample preparation methods, 92*t*
  sarcoid bronchoalveolar fluid, 619
  solution-based techniques, 90–91
  systems medicine applications, 135–136
  undersampling issues, 90
Proteomics Identifications Database (PRIDE), 120
Prothrombin (factor II), 375
  polymorphism, 379
    cerebral venous thrombosis risk, 380
    targeted screening, 384*t*
Protocadherin-2 precursor, sarcoid bronchoalveolar fluid, 619
Proton pump inhibitors
  Helicobacter pylori eradication, 628, 635
  pharmacogenomics, 628, 635
PROWESS, 767, 768
PSEN1/2 (presenilin 1/2) mutations, Alzheimer's disease, 689, 694
Pseudoexfoliation glaucoma, 717
Pseudohypoaldosteronism type 2 (Gordon's syndrome), hypertension, 263
Pseudomonas aeruginosa
  genetics of host response, 751
  genome mapping, 66
Pseudomonas fluorescens antibodies, inflammatory bowel disease, 578
PSMA6 (proteasome subunit α type 6) polymorphism, myocardial infarction, 293
Psoriasis, 56–57
PTEN
  Bannyan-Ruvalcabe-Riley syndrome, 464
  breast cancer, 448
  Cowden syndrome, 464
  gliomas, 523, 524
  ovarian cancer, 496
  p53 cooperative interactions, 480
  prostate cancer, 479–480
PTP4A, Hodgkin lymphoma/Reed-Sternberg cells, 413
PTPN11 (Noonan syndrome), congenital heart disease, 393
PTPN22, 583
  diabetes type 1, 676, 680, 681
  rheumatoid arthritis, 551, 552, 554–555, 556, 557
Public Health Genetics Unit, 248, 250
Public health genomics, 245
  communication, 248
  core activities, 248–250
  definition, 245
  gene–environment interactions as deteminants of health, 246, 247*f*
  genetic exceptionalism avoidance, 247
  health professional education, 249
  information sharing, 249
  key concepts, 246–247
  knowledge base
    basic research, 247
  knowledge integration, 248
  policy issues, 249
  service development/evaluation, 248–249
  societal attitudes, 248
  stakeholder engagement, 248
  translational strategy, 247, 247*f*, 248
  workforce training, 249
Public health professionals
  genomics specialist training, 249
Public Population Project In Genomics (P3G) Consortium, 248
PubMed, 153
PUD, 637
Pulmonary embolism, 382
Pulmonary fibrosis, 616–617
  genetic determinants in inherited disorders, 620
  see also Diffuse parenchymal lung disease
Pulmonary hypertension, chronic obstructive pulmonary disease, 605
Pulmonary stenosis, 393
Pulmonary surfactant proteins, 619
Pupil, developmental abnormalities, 712
Pyrosequencing
  tissue sample DNA methylation assessment, 64
Pyruvate cycling pathways, glucose-stimulated insulin secretion, 102–103, 103*f*
PYY, obesity, 666

Q

19q chromosome, 528
20q13/*CD40*, 558
22q13/*IL2RB*, 558
12q13/*PIP4K2C*, 558
10q21, 583
1q22/*CD244*, 558
6q23/*TNFAIP3*, 558
16q24, 583
4q27/*IL2-IL21*, 558
2q32/*STAT4*, 558
2q33/*CTLA4*, 558
9q33/*TRAF1-C5*, 558
Quality-adjusted life years (QALYs), 242
  cost-utility analysis, 234, 235
Quality assurance
  diagnostic assays, 194, 195
  pharmacogenetics test validity, 182
Quantitative proteomics, 95
Quantitative trait loci (QTLs)
  obesity, 664
  salt-sensitivity, 261
Quinidine
  Brugada syndrome, 363*f*
  drug-induced long QT syndromes, 352
  short QT interval syndrome, 361

R

R, 122–123
RA genetics, 558
RAB, promoter hypermethylation in lung cancer, 434
Race *see* Ethnic/racial variation
RAD001 (Everolimus)
  prostate cancer targeted therapy, 485
Radioimmunotherapy, 537
RANTES (CCL5), 739
  chronic obstructive pulmonary disease, 607
Rapamycin (FK506; sirolimus), 83
  acute lymphoblastic leukemia, 116
  drug-eluting intracoronary stents, 297
RARB methylation status, tumor treatment response prediction, 67

Rare variant hypothesis, 4
Rare variants search, medical resequencing in, 5
Ras
  colorectal cancer, 458
  leukemias, 421
  lung cancer, 438, 440
  ovarian cancer, 496
RASSF1 methylation status
  lung cancer, 434
  tumor treatment response prediction, 67
RASSF1A methylation status
  breast cancer, 67
  ovarian cancer, 496
  prostate cancer, 478
Rb
  mutations
    gliomas, 523
    lung cancer, 439
    prostate cancer, 480
RBP1 methylation status, tumor treatment response prediction, 67
Reactive oxygen species
  chronic obstructive pulmonary disease, 607
  endothelial cells, 762
  head and neck cancer, 512
  prostate cancer associations, 478
Real-time polymerase chain reaction
  cardiac allograft rejection gene expression profiles, 329, 330
  Helicobacter pylori diagnosis/monitoring, 636
Recombination
  non-allelic homologous, copy number variation (CNV) generation, 46–47, 48, 48*f*
Rectal cancer
  irinotecan treatment, genotype-guided clinical trial, 182
  see also Colorectal cancer
Reductionist approach to disease, 201, 202*f*
Reed-Sternberg cells, 836
Regulatory issues, 223
  genomic tests, 225–227
  laboratory-developed tests, 226–227
  pharmacogenetic tests, 183
  public health genomics, 248
  targeted therapies, co-approval of diagnostic test, 532
Regulatory myosin light chain (*MYL2*) mutations, hypertrophic cardiomyopathy, 338
Regulatory T cell (Tregs), 569, 571
Reimbursement
  genomic tests, 239–240, 240*f*
  policy issues, 215
Reis–Bückler's corneal dystrophy, 714, 715
REL
  diffuse large B-cell lymphoma, 410
  primary mediastinal large B-cell lymphoma, 411, 412
Relative expression reversals, 137
Relevance networks, 118
Renin-angiotensin system
  hypertension, 263
    pharmacogenomics, 264
  hypertrophic cardiomyopathy, 343
  therapeutic targeting in cirrhosis, 646
Renal cell carcinoma
  sorafenib therapy, 542
Renal disease
  hypertension association, 263
Renal transplant rejection, 116
Repeatability *see* Reproducibility
Reproducibility
  association studies, 763
    pharmacogenetics, 182
Research policy, 210–211
Re-sequencing
  polymorphisms detection, 30
  process, 31, 31*f*
    electrophoresis, 32–33
    PCR set-up and clean-up, 32
    primer design, 31–32, 32*f*
    sequencing reaction set-up and clean-up, 32
    trace analysis, 32–33, 33*f*
Resistin, obesity, 668
Respiratory distress syndrome, 619
Respiratory syncytial virus, 606
Restriction fragment length polymorphisms (RFLPs), 636
Restrictive cardiomyopathy, 337
Retina, 712–713, 713*f*
  disorders, 716
Retinitis pigentosa, 716
Retinoblastoma, 716
Retinoic acid metabolism
Retinoids
  DNA methylation profiles, tumor response prediction, 67
RETN, obesity, 666
Rett syndrome, 66*t*
Reverse phase arrays, 569
  prostate cancer serum proteomics, 482
  T cell response in multiple sclerosis, 569
Reverse transcription–polymerase chain reaction
  breast cancer, 452, 453
  expression profiling clinical applications, 82
  follicular lymphoma, 413, 414
  lymphoma diagnosis, 409
  mantle cell lymphoma, 415
Reversed phase high-performance liquid chromatography (RPLC), quantitative proteomics, 91
  mass spectrometry coupling, 92
Rheumatic heart disease, 314
Rheumatoid arthritis, 551
  auto-antibodies, 551, 552, 556
    anti-cyclic citrullinated peptide (anti-CCP), 551, 552, 556
    rheumatoid factor (RF), 551, 552, 556
  clinical features, 551*t*, 552
  diagnosis, 558
  diagnostic criteria, 552
  disease-modifying anti-rheumatic drugs (DMARDS), 558
  environmental risk factors, 552, 555
  gender-related risk/hormonal factors, 555–556
  gene expression profiles, 78, 557
    comparison with other autoimmune disorders, 569
  genetic factors, 551, 552
  HLA-DRB1 association, 552–554, 556, 557
    shared epitope alleles, 552, 553*t*, 555
  infectious agent associations, 556
  monitoring, 558
  non-*HLA-DRB1* MHC genes, 556
  non-MHC genes, 554, 554*t*
    candidate genes, 554
  predisposition model, 556
  prognosis, 558
  proteomics, 557
  PTPN22 associations, 551, 552, 554, 556
  screening, 557–558
  tumor necrosis factor-α inhibitors in management, 552, 557
Rheumatoid factor (RF), 553, 556
RhoA, ovarian cancer metastasis, 497
RhoC
  Hodgkin lymphoma/Reed-Sternberg cells, 413
Rhodopsin mutations
  congenital night blindness, 716
  retinitis pigentosa, 716
Ribavirin
  hepatitis C, 782
Ribosomal RNAs (rRNAs), 16
RIG-like receptors, pathogen recognition/signaling, 763–764
RigI-helicases, 745
Rituximab, 538
  Hodgkin's lymphoma, 534
  rheumatoid arthritis, 558
RNA interference
  gene expression profiling studies, 81
RNA interference (RNAi) screen, 455
RNA isolation, microarray experiments, 75–76
RNA-seq, 83
RNAseL polymorphism, prostate cancer, 477, 478
Roadmap Epigenomics Initiative program, 68
ROBO3, strabsimus, 713
Roche, 545
Roche-454 GS (Genome Sequencer), 34, 35
Romano-Ward syndrome, 352, 353
  genetics, 353
    ion channel mutations, 354–356
    LQT1 (KVLQT1/KCNQ1), 353, 354
    LQT2 (HERG/KCNH2), 353, 355
    LQT3 (SCN5A), 353, 355–356

Romano-Ward syndrome, (*Continued*)
LQT4 (ankryn-β), 353, 356
LQT5 (minK/KCNE1), 353, 354, 356
LQT6 (MiRP1), 355, 356
LQT9 (caveolin-3), 356–357
LQT10 (SCN4B), 357
non-ion channel-encoding genes, 356–357
RP16, hepatic stellate cells, 651
RPE65 mutations, Leber's congenital amaurosis, 716
RRMS, 571
Rubenstein–Taybi syndrome, 713
RyR2 mutations (cardiac ryanodine receptor)
catecholaminergic polymorphic ventricular tachycardia (CPVT), 362
hypertrophic cardiomyopathy, 339

S

S, 122, 123
S-100 proteins
glioma prognosis, 526
head and neck cancer, 517
multiple sclerosis, 570
pancreatic mucinous cystic neoplasms, 505
rheumatoid arthritis, 557
SabA, *Helicobacter pylori* adhesins, 633
Salivary biomarkers, head and neck cancer locoregional recurrence, 516–517
SALL1, Towne–Brocke syndrome, 713
SALL4 mutations
congenital heart disease, 391, 392
Duane's syndrome, 713
strabismus, 713
Salmeterol, chronic obstructive pulmonary disease, 610
Salmonella infection, susceptibility genes, 736
Salmonella typhimurium, 583
Salt-sensitivity/resistance, 261
genetic basis, 260–261
hypertension predisposition, 260
Sanger sequencing technique, 28
resequencing, 31
whole genome shotgun sequencing, 28
Sarcoidosis, 617–619
candidate gene studies, 617–618, 618*t*
familial clustering, 617, 618
genetic factors, 617
Kveim reaction, 619
linkage analysis, 618–619
Mycobacterium tuberculosis catalase-peroxidase involvement, 619
prognostic applications of genomics, 619
proteomics, 619
Sarcomas, systems medicine approach to diagnosis, 133
SARS virus, 739
Scandanavian Simvastatin Survival Study (4S), 296
Scavenger receptor class BI (SR-BI), 271
Schizophrenia
microarray experiments, 78
SCN1A polymorphism
phenytoin response influence, 181
SCN4B (*LQT10*), 357
SCN5A (*LQT3*) *see* Cardiac sodium channel gene mutations
Scoring ALgorithm for Spectral Analysis (SALSA), 93
Screening, 165
acute coronary syndromes, 304
asthma, 594
brain tumors, 524
breast cancer, 447
cardiovascular risk, 296–297
chronic obstructive pulmonary disease, 609–610
colorectal cancer, 457, 467
genomic, 193
germline, 193
head and neck cancer, 512–513
hepatitis B, 777–778
hepatitis C, 777–778
hereditary hemochromatosis, 249
hypertension, 262
hypertrophic cardiomyopathy, 342–343, 342*f*
inflammatory bowel disease, 579
intraocular pressure elevation, 716
lipid disorders, 272
lung cancer, 434, 435
lymphoma, 409
newborn *see* Newborn screening
obesity, 668
ovarian cancer, 494
pancreatic cancer, 502–503
prostate cancer, 480–481
public health genomics, 248
rheumatoid arthritis, 557
thrombophilias, 380–381, 384*f*
population-based, 381, 385*f*
viral hepatitis, 777–778
SCYA3, diffuse large B-cell lymphoma prognosis, 411
SELDI ProteinChip, hepatitis B infection proteomics, 654
SELDI-TOF MS, 695
Selective serotonin reuptake inhibitors, pharmacogenomics, 727
Self-organizing maps, 118
disease sub-type discovery, 119
Sensorineural deafness, Jervell and Lange-Nielsen syndrome, 353, 357
Sepsis, 760–761
applications of genomics, 766
associated genetic polymorphisms, 761–765
association studies, 762–763
candidate gene studies, 762
clinical critera, 760*t*
clinical staging/severity indices, 767
coagulation, 765
cytokine polymorphisms, 764–765
diagnosis, 767
gene expression profiles
pathogen signatures, 766–767, 768*f*
progression signatures, 767–769
host response, 761, 762*f*
multiple serum biomarkers, 765, 766*f*
pathogen recognition/signaling, 763–764
CD14, 764
intracellular signaling molecules, 764
mannose-binding lectin, 764
NOD-like receptors, 763–764
RIG-like receptors, 763–764
Toll-like receptors, 763
severe (with acute organ dysfunction), 760
therapeutics
activated protein C, 765, 767
Septic shock, TLR4 polymorphism influence, 736
Sequential organ failure assessment (SOFA), 767
SEQUEST, 93
SeqWright, 156
Serial analysis of gene expression (SAGE), 73
liver tissue, 648
Serine/threonine kinase 15 (STK15), hepatocellular carcinoma biomarker, 653
Serotonin (5-HT) neurotransmission
depression, 723, 724–725, 725*t*
emotional behavior regulation, 725
stress hormone interactions, 724
Serotonin transporter gene polymorphism *see* 5-Hydroxytryptamine transporter (*5-HTT*; SLC6A4)
SERPINE2, 612
Serum biomarkers
head and neck cancer recurrence prediction, 516–517
prostate cancer, 482
SFTPB, 612
SH2B3, diabetes type 1, 680
SHANK3, 56
Short interfering RNAs (siRNA), 497, 498
therapy
apoB targeting, 282
chronic obstructive pulmonary disease, 611
Short QT interval syndrome, 361
clinical features, 361
genetics, 362, 362*t*
treatment, 361
Shprintzen syndrome, 391
Shrimp alkaline phosphatase/exonuclease I (SAP/Exo I) mix, 32
Sickle-cell anemia, 201
14-13-3Sigma promoter methylation, prostate cancer, 478
Significance Analysis of Microarrays (SAM), 621

familial pulmonary fibrosis, 620
Significance measures, 118
Silica exposure, rheumatoid arthritis, 555
SIM1 mutations, obesity, 665
Simvastatin, 296
Single-base extension (SBE), Illumina Infinium II assay, 40
Single cell analysis
DNA-encoded antibody libraries (DEAL), 136
systems medicine applications, 135
Single nucleotide polymorphisms (SNPs), 4–5, 17, 21–22, 38, 194, 297, 365, 528, 612, 682
blood lipid variation, 279
high-density lipoprotein cholesterol, 275*t*
low-density lipoprotein cholesterol, 274, 275*t*
triglycerides, 275*t*
complex disease association studies, 27–28, 164
coronary artery disease/myocardial infarction studies, 290, 291, 291*t*
genetic susceptibility biomarkers, 205
genome-wide association studies, 40
genotyping technologies, 38, 40
raw data quality control, 42–43
HuRef genome analysis, 18
local regulatory variants, 20
population studies
heterozygosity estimates, 18
rheumatoid arthritis, 552
tagging, 38, 40
copy number variation (CNVs) characterization, 54
HapMap project, 181
pharmacogenetic studies, 180
thrombosis-related gene studies, 380
whole genome shotgun sequencing, 29–30, 30*f*
Sinus node dysfunction/sick sinus syndrome, 350, 362
clinical features, 362–363
genetics, 363
Sirolimus *see* Rapamycin
SIRT2, glioblastomas, 527
Sitosterolemia, 273–274
Skills for Health Service, 216
SLC6A3, nicotine addiction, 604
SLC6A4 *see* 5-Hydroxytryptamine transporter
SLC11A1 (NRAMP)
Leishmania infection susceptibility, 736
mycobacterial infection susceptibility, 736, 754, 762
SLC22A4/OCTN1
inflammatory bowel disease, 576, 578, 581
rheumatoid arthritis, 554, 555
SLC22A5/OCTN2, inflammatory bowel disease, 576, 578, 581
SLC30A8, diabetes type 2, 678, 679
SLCO1B1 gene, 186
Slow acetylator phenotype *see* *N*-Acetyltransferase 2 (NAT2) polymorphism
SMAD2, ovarian cancer, 495
SMAD7, colorectal cancer, 465
Small-cell lung cancer, 435
gefitinib treatment, 194
Small interfering RNAs *see* Short interfering RNAs (siRNA)
Small nucleolar RNAs (snoRNAs), 16
SNCA (*PARK1; PARK4*; α-synuclein), Parkinson's disease, 702, 705–706
SNOMED Clinical terms, 121
SNPdetector, 32
SNPedia, 158
SNTA1 (LQT12), 365
Social issues
genomics research policy, 210
public health genomics, 246, 249
Software, bioinformatics
free tools, 121–124
SOLAR, metabolic profiling in cardiovascular disease, 104
Solid tumors
monoclonal antibody therapy, 540–542
small molecule-targeted therapy, 544–545
Somatic mutation, 5, 194
Somatostatin, gastric acid secretion regulation, 628
Somatostatin receptor type 2 (SSTR2), positron emission tomography (PET) reporter Sorafenib (BAY 43-9006), 544
prostate cancer trials, 485
Sorafenib, 545
SORL1, Alzheimer's disease, 691, 695
Sotalol
Brugada syndrome, 363
drug-induced long QT syndromes, 352
Sphingolipids, targeted metabolic profiling, 102
Spinal muscular atrophy
newborn screening, 193
SPIO, cell surface receptor imaging, 227
Spliceosome
RNAs, 16
SPMS, 571
Spongiform encephalopathies *see* Prion disease
Spotted cDNA microarrays, 73, 75
SRBP-1, hepatitis C receptor, 776
SRC3 *see* AIB1
Stanford Microarray Database (SMD), 119–120
Staphylococcus aureus, 4
genetics of host response, 748, 750–751, 763, 767
genome mapping, 744
Starch diet, 57
STAT signaling pathways, 57
STAT1
hepatitis C persistent infection, 779
mycobacterial disease susceptibility, 754
STAT3, 583
head and neck cancer, 517
STAT4, 559
Stathmin, laryngeal carcinoma, 517
Statins, 281
familial hypercholesterolemia, 272
pharmacogenomics, 281, 296–297
post-myocardial infarction ventricular remodeling, 296
STATS, interferon-α activation, 782
Stevens-Johnson syndrome, 230
STK11 mutations
breast cancer, 448
STK15, hepatocellular carcinoma, 653
Strabsimus, 713
Strategic health planning, 165
Stratifin, head and neck cancer, 517
Streptococcus pneumoniae
chronic obstructive pulmonary disease exacerbations, 606
genetics of host susceptibility, 749–750, 767
genome, 33
mapping, 744
Stress cardiomyopathy, 314
Stroke
hypertension association, 263
Strong cation exchange (SCX), quantitative proteomics, 91
Subtractive hybridization/differential display, 73
congenital heart disease gene discovery, 395
Sudden infant death syndrome, long QT syndrome-like, 352, 357
Sudden and Unexpected Death Syndrome (SUDS), 361
Sudden Unexpected Nocturnal Death Syndrome (SUNDS), 361
Sulfasalazine
rheumatoid arthritis, 558
Sulfotransferases, cigarette smoke metabolism, 433
Sulindac, colorectal cancer chemoprevention, 468
Sunitinib, 545
gastrointestinal stromal tumors, 542
prostate cancer trials, 485
Superoxide dismutase (SOD3), 612
Supervised learning, 118
acute leukemias classification, 113
gene expression data analysis, 76, 77*f*
acute lymphoblastic leukemia, 424
heart failure, 319
methods, 118
prognostic marker identification in leukemia, 427, 427*f*
Support vector machines, 137, 118
cancer metastates primary diagnosis, 116, 117*f*
microarray data analysis, 76

Supravalvular aortic valve stenosis, 394
Surface-enhanced laser desorption/ionizing time-of-flight mass spectrometry (SELDI-TOF-MS)
  colorectal cancer, 467
  head and neck cancer, 517
  non-alcoholic fatty liver disease, 655
  ovarian cancer, 495
  pancreatic cancer, 504
  prostate cancer, 482
  sarcoid bronchoalveolar fluid, 619
Surface plasmon resonance imaging
  blood biomarker analysis, 135–136
  head and neck cancer screening, 513
Surfactant protein A variants
  idiopathic pulmonary fibrosis, 619
  mycobacterial disease susceptibility, 754
Surfactant protein B deficiency, neonatal respiratory distress, 619–620
Surfactant protein B variants, 619
Surfactant protein C variants, 619
  diffuse parenchymal lung disease, 620
  familial pulmonary fibrosis, 620
  family interstitial pneumonia/neonatal respiratory distress, 619
Surfactant protein D variants, 619
  mycobacterial disease susceptibility, 754
Susceptibility genes
  asthma, 593
  infectious disease, 734–736, 735*t*, 762
  public health genomics, 246, 247*f*
SWOG 99-16, 485
Syndrome of apparent mineralocorticoid excess, hypertension, 263
Syndrome X, colorectal cancer, 464
Systemic inflammatory response syndrome (SIRS), 760
  clinical critera, 760*t*
Systemic lupus erythematosus
  complement component C4 gene copy number variation, 53
  DNA methylation aberrations, 65
  gene expression profiles
    comparison with other autoimmune disorders, 569
Systemized Nomenclature of Medical-Clinical Term (SNOMED-CT), 120, 121, 153
Systems biology, 131–138
  biological networks, 159, 166
  blood protein diagnostic markers, 133–134
    organ-specific fingerprints, 134
  computational/mathematical challenges, 137
  defining features, 131
  emerging technologies, 134–136
  in vitro measurements, 134–135
  infectious disease, 739
  medical applications, 132–133
  molecular signature classifiers, 137

T
t(8;21), acute myeloid leukemia, 421, 424
t(9;22), chronic myelogenous leukemia, 421
t(11;14)(q12;q32), mantle cell lymphoma, 415
t(11;14)(q13;q32), mantle cell lymphoma, 414
t(11;q23), acute myeloid leukemia, 424
t(14;18)(q32;q21), diffuse large B-cell lymphoma, 410
t(14;18)(q32;q21), follicular lymphoma, 413, 414
t(15;17)
  acute myeloid leukemia, 424
  acute promyelocytic leukemia, 422, 542
T cells
  cardiac allograft acute rejection, 327
  chronic obstructive pulmonary disease, 606
  cytotoxic *see* Cytotoxic T cells
  helper *see* Helper T cells
  multiple sclerosis immune response, 569–570
  sarcoidosis, 617, 619
  see also CD4 T cells, CD8 T cells
t-tests, 118
Tacrolimus, 327
TAFII, primary biliary cirrhosis, 652
TAILORx (Trial Assigning Individualized Options for Treatment), 2, 83, 82, 210, 236, 453, 454
Tamoxifen, 67, 454–455
  biomarkers of response, 535
  breast cancer targeted therapy, 535
Tandem mass spectrometry (MS/MS)
  metabolic profiling, 98, 100, 102, 103, 104
  newborn screening, 101–102, 738
  proteomics, 89, 90, 92, 93*f*
Tangier disease, 276, 279, 282
TAP2
  Mycobacterium leprae susceptibility, 754
TAS2R50 gene polymorphism, myocardial infarction, 293
Tau (microtubule associated protein)
  multiple sclerosis, 570
  Parkinson's disease, 706
TAX 327, 485
Taxanes, DNA methylation profiles in tumor response prediction, 67
Taxonomy, 116–117
TBX1, congenital heart disease, 391–392
TBX5
  cardiac gene expression interactions, 392
  congenital heart disease, 391
  microarray analysis, 395
  SALL4 regulation, 392
TBX20, 398
TCAP mutations (telothonin), hypertrophic cardiomyopathy, 338, 341
TCF2, 198
TCF7L2, 682
  diabetes type 2 susceptibility, 677, 678, 679
TCGA, 528

TEL-AML1, acute lymphoblastic leukemia, 424
Telomerase
  activated hepatic stellate cells, 651
  pancreatic cancer, 501
  pancreatic cystic neoplasms, 505
  pancreatic endocrine tumors, 506
Telothonin (*TCAP*) mutations, hypertrophic cardiomyopathy, 338, 341
Temozolomide
  oligodendrogliomas, 525
  tumor response prediction, 67
Tenascin
  hepatic stellate cell production, 649, 651
Tenofovir, hepatitis B treatment, 781
N-Terminal pro-B-type natriuretic peptide, acute coronary syndrome risk stratification, 308
TERT/TERTC, 623
Tetralogy of Fallot, 391, 392, 393, 395
TFAP2B (Char syndrome), congenital heart disease, 393
TFF1, ovarian cancer, 495
TGFB1/BIGH3 (keratoepithelin), corneal dystrophies, 713, 714, 715
Theophylline, chronic obstructive pulmonary disease, 610
Therapeutic index, 176
Thiazide diuretics, pharmacogenomics, 265
Thiopurine methyltransferase (*TPMT*)
  haplotypes, 181–182
  polymorphism, thiopurines metabolism, 230
    acute lymphoblastic leukemia, 185
    inflammatory bowel disease, 582
Thiopurines
  pharmacogenomics
    inflammatory bowel disease, 582
Third-party reimbursement for genetic/genomic tests, 196
Thrombin, 374, 375, 378, 379
  activation, 377
  sepsis, 765
Thrombomodulin, 375
  oxidation, 377
  sepsis, 765
Thrombophilias
  acquired, 381
  coagulation protein posttranslational modification, 377
  prognosis, 382–383
  screening, 380–381, 384*f*
    population-based, 381, 385*f*
  vascular bed-specific thrombosis, 377–378
Thrombosis, 374–387
  circulating cellular/protein influences, 378–379
  family history, 380–381
  gene association studies, 380
  gene linkage studies, 379–380
  heritability, 380

instrumentation research, 384
personalized approach, 380
gene expression profiles, 380, 381*f*, 382*f*
pharmacogenomics, 384–386
screening, 380–381
population-based, 381
targeted approach, 380–381, 384*f*, 384*t*
vascular bed-specific, 377–378, 378*t*
Thrombospondin
head and neck cancer, 517
polymorphism, myocardial infarction, 291
Thromboxane, thrombotic event prediction with anticardiolipin antibodies, 380
Thymidine phosphorylase, 5-fluorouracil pharmacogenomics, 469, 469–470
Thymidylate synthase polymorphism
colorectal cancer, 465
5-fluorouracil pharmacogenomics, 469, 470
Thyroid cancer
radioiodine therapy
see also Papillary thyroid carcinoma
Thyrotoxicosis, heart failure, 314
Timothy syndrome, 352, 353, 358
clinical features, 358
genetics, 353, 358, 359*t*
tinman, 391
Tiotropium bromide, chronic obstructive pulmonary disease, 610
Tipifarnib, 422
TIR-associated protein (TIRAP; Mal), 747
Tissue factor, 374, 378
Tissue factor pathway inhibitor (TFPI), 374
Tissue inhibitor of metalloproteinases 1 (TIMP-1)
atherosclerotic plaque, 290
Tissue inhibitor of metalloproteinases 2 (TIMP-2)
atherosclerotic plaque, 290
ovarian cancer metastasis, 497
promoter methylation, prostate cancer, 478
Tissue inhibitor of metalloproteinases 3 (TIMP-3), ovarian cancer metastasis, 497
Tissue inhibitors of metalloproteinases (TIMPs)
atherosclerotic plaque, 290
cancer-related hypermethylation, 478
Tissue-type plasminogen activator, pancreatic cancer, 506
Titin (*TTN* mutation), hypertrophic cardiomyopathy, 338, 341
TLR gene, 739
TLR1, 745
host response to infection
Enterobacteriacea, 751
mycobacteria, 753–754
Staphylococcus aureus, 751
Streptococcus pneumoniae, 749
TLR2, 738, 745
genetics of infection susceptibility, 763
leprosy, 736
Gram-negative bacteria detection, 763
Gram-positive bacteria detection, 748, 763
host response to infection
Enterobacteriacea, 751
Legionella, 752
Listeria monocytogenes, 751
mycobacteria, 753, 762
Neisseria meningitidis, 752
Staphylococcus aureus, 750, 751
Streptococcus pneumoniae, 749
ligand
lipoteichoic acid, 748
TLR3, 738, 747
TLR4, 738, 739, 745, 747
genetics of infection susceptibility, 763
lipopolysaccharide response variation, 735–736
Gram-negative bacteria detection, 763
host response to infection
Enterobacteriacea, 751–752
Legionella, 752
Neisseria meningitidis, 752
Streptococcus pneumoniae, 749
lipopolysaccharide co-receptor, 764
TLR5, 745
genetics of infection susceptibility, 736, 763
Gram-negative bacteria detection, 763
Gram-positive bacteria detection, 763
host response to *Legionella*, 736, 752
TLR6, 745
host response to infection
Staphylococcus aureus, 751
Streptococcus pneumoniae, 749
TLR7, 738
TLR8
TLR9, 745
host response to *Streptococcus pneumoniae* infection, 749
TLRs (Toll-like receptors), 745
common host response to infection, 738
infectious diseases susceptibility associations, 735–736
ligands, 746*t*
pathogen recognition, 733, 734, 745, 747, 763, 766
pathogen-specific response to infection, 738
signaling, 747, 747*f*
TMPRSS2/ERG translocation
prostate cancer, 483
urine biomarker, 481
TNF see Tumor necrosis factor-α, polymorphism
TNFAIP3, 416
TNNT2 mutations (cardiac troponin T), hypertrophic cardiomyopathy, 338, 341
TNNT13 mutations (cardiac troponin I), hypertrophic cardiomyopathy, 338
TOB1 gene, 571
Toll-like receptors *see* TLRs
Topoisomerase 2a (TOP2A), astrocytomas, 526
Topoisomerase 2beta, hepatitis B, 654
Torsades de pointes
Andersen-Tawil syndrome, 358
Jervell and Lange-Nielsen syndrome, 357
long QT syndromes, 350, 351, 351*f*
Towne–Brocke syndrome, *SALL1*, 713
Toxicogenomics, 224, 230
TP53 *see* p53
TPH2 see Tryptophan hydrolase
TPM1 mutations (alpha-tropomyosin), hypertrophic cardiomyopathy, 338, 341, 342
TPMT polymorphism *see* Thiopurine methyltransferase
Trabecular meshwork, 715–716
TRAF-3, hepatitis C, 654
TRAG3 methylation status, tumor treatment response prediction, 67
TRAM (TRIF-related adaptor molecules), 747
Transcription factors
congenital heart disease susceptibility genes, 391–393
follicular lymphoma, 413
Hodgkin lymphoma/Reed-Sternberg cells, 412
prostate cancer-related rearrangements, 483
Transcriptomics, 2, 81–82
alcoholic liver disease, 652
applications, 135
autoimmune hepatitis, 652
biliary disease, 652–653
congenital heart disease gene discovery, 395, 396*f*
gene expression microarrays, 81
heart failure monitoring, 319–320
hepatitis B infection, 651
hepatitis C infection, 651
hepatocellular carcinoma, 653
hypertrophic cardiomyopathy, 341–342
liver, 648
disease, 651–652
multiple sclerosis, 568–569
non-alcoholic fatty liver disease, 652
primary biliary cirrhosis, 652
primary sclerosing cholangitis, 652
systems medicine, 135
Transfer RNAs (tRNAs), 16
Transforming growth factor-β
chronic obstructive pulmonary disease, 604, 607
colorectal cancer, 460
head and neck cancer, 517
hepatic stellate cell production, 650
ovarian cancer, 495–496
Transforming growth factor-β antagonists
cirrhosis, 646

Transforming growth factor-β receptor mutations
colorectal cancer, 459, 460
ovarian cancer, 495
Translational diagnostics, 168
definition, 168
Translational genomics, 163–172
candidate health applications, 164
challenges, 165
clinical implementation, 191–198, 192*f*
economic incentives, 237–239
enabling competencies/capabilities, 165–169
evidence-based guidelines development, 164
genomic testing, 164
impact on health-disease continuum, 164–165, 165*f*
implementation strategies, 169–170, 170*f*
pharmacogenomics, 228
public health aspects, 164
research framework, 164, 164*t*
Translational research
access/reimbursement issues, 214–215
policy issues, 211
public health genomics, 247
Translocations, 14
Transthyretin, hepatitis B infection proteomics, 654
Trastuzumab, 224, 229, 455, 534, 541
adverse effects, 449
breast cancer treatment, 194, 448, 536
pharmacogenetics, 183, 185
diagnostic tests of patient eligibility, 534, 537
ovarian cancer treatment, 495
Tregs *see* Regulatory T cells
Tricuspid atresia, 391
Tricyclic antidepressants, drug-induced long QT syndromes, 352
TRIF (TIR domain containing adaptor protein inducing interferonβ), 747
Triglycerides
cardiovascular risk, 271
genetic disorders, 277
genetic variants, 275*t*, 277–278, 279
intestinal absorption, 270
lipid-modulating therapies, 281
lipoprotein transport, 269
metabolism, 270, 271
plasma levels
genetic factors, 271
screening, 272
transport/clearance, 270
Trisomy 13, congential heart disease, 394
Trisomy 18, congential heart disease, 394
Trisomy 21 (Down syndrome)
APP (amyloid precursor protein gene) copy number variation, 689
congenital heart disease, 390, 394
obesity association, 665
strabismus, 713
TRITON Study, 387
trkB *see* Brain-derived neurotrophic factor receptor
Troponin I
acute coronary syndromes
diagnosis, 305
risk stratification, 308, 309
TNNT13 mutations, hypertrophic cardiomyopathy, 338
Troponin T
acute coronary syndromes
diagnosis, 305
risk stratification, 308, 309
TNNT2 mutations, hypertrophic cardiomyopathy, 338, 341
Trp719Arg polymorphism298
Truncus arteriosus, 391
Tryptophan hydrolase (*TPH2*), psychiatric disorder associations
antidepressant drug response effects, 727, 727*t*
depression, 724, 725
TSC1/2 DNA methylation, breast cancer prognosis, 67
TSP-1 polymorphism, myocardial infarction, 291–292, 292*f*
TSP-2 polymorphism, myocardial infarction, 291
TSP-4 polymorphism, myocardial infarction, 291, 292, 292*f*
TTN mutation (titin), hypertrophic cardiomyopathy, 338, 341
Tuberculosis, 201
genetic susceptibility, 753–754
HLA Class I genes, 736
NRAMP1/Slc11a1, 736
see also Mycobacterium tuberculosis
Tuberous sclerosis, pulmonary fibrosis, 617, 620
Tubulin, multiple sclerosis biomarker, 570
Tumor DNA sequencing, 33
Tumor necrosis factor-α
asthma, 593
cardiomyocyte apoptosis in heart failure, 316
chronic obstructive pulmonary disease, 607, 608
follicular lymphoma, 413
hepatitis C infection, 780
host response to pathogens, 747
Gram-positive organisms, 748
polymorphism, 289
asthma, 593, 594
chronic obstructive pulmonary disease, 604
hepatitis B clearance, 776
malaria/leprosy susceptibility, 736
multiple sclerosis studies, 568
peptic ulcer disease, 635
sepsis, 764
rheumatoid arthritis, 557, 558
sarcoidosis, 617
Tumor necrosis factor-α inhibitors
cirrhosis, 647
rheumatoid arthritis, 552, 558
see also Infliximab
Tumor necrosis factor-α receptor
host response to *Staphylococcus aureus*, 751
mycobacterial disease susceptibility, 754
head and neck cancer, 511, 516
local CpG island hypermethylation, 61
lung cancer, 434, 439
ovarian cancer, 496
pancreatic cancer, 501
Tumor-specific biomarkers, 94
Tumor-specific EGFRvIII mutation, 137
Turcot syndrome, 461, 463
brain tumor predisposition, 523
Turner syndrome
congential heart disease, 394
obesity association, 665
23andMe, 156, 198
Twin studies
Alzheimer's disease, late onset, 689
hemostasis, 376
hepatitis B/C susceptibility, 776
infectious diseases susceptibility, 735
inflammatory bowel disease, 576
major depression, 724
obesity, 661, 663
Parkinson's disease, 702
pharmacogenetics, 178–179
pulmonary fibrosis, 616
rheumatoid arthritis, 552
sepsis outcome, 762
tuberculosis susceptibility, 753–754
Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), 95
Type 1 diabetes, genetics of, 681
Type 2 diabetes, genetics of, 681
Typhoid fever susceptibility, HLA Class II genes, 736
Tyrosine kinase inhibitors
leukemia, 422
lung cancer, 438, 440
Tyrosine kinase receptors
gliomas, 522
lung cancer pathogenesis, 438
ovarian cancer, 495

U

UCP1
obesity, 666, 667, 670, 671
UCP2
obesity, 666, 667, 671
UCP3, obesity, 666, 667, 670, 671
UGT1A1 polymorphism
irinotecan metabolism, 230

UGT2B17, 57
  prostate cancer-related copy number variation, 52–53, 52*f*
UK Biobank, 145, 212
UK Genetic Testing Network, 249, 298
Ulcerative colitis, 575
  classification, 579
  clinical features, 579
  colorectal cancer risk, 464, 580
  corticosteroid therapy, 580
  diagnosis, 579
  disease severity genetic markers, 580–581
  genetic factors, 576, 577*t*
    HLA-DRB1 alleles, 580
  incidence, 575
  perinuclear antineutrophil cytoplasmic antibodies (pANCA), 578, 579, 580
  prognosis, 580–581
  surgery
    indications, 581
  twin studies, 576
  see also Inflammatory bowel disease
Ultrasound
  head and neck cancer, 513
  prostate cancer, 482
UMLS Metathesaurus, 153
Undulin, hepatic stellate cell production, 650
Unified Medical Language System (UMLS), 121
UniGene identifier, 121, 648
Unsupervised learning, 118
  disease sub-type discovery, 119
  gene expression data analysis, 76, 77*f*
    acute myeloid leukemia, 424
    heart failure, 319
  methods, 118
Urea cycle intermediates, targeted metabolic profiling, 102
Urinary colecting system tumors, hereditary nonpolyposis colorectal cancer (Lynch syndrome), 463
Urine biomarkers
  prostate cancer, 67, 481
Urokinase-type plasminogen activator, breast cancer metastasis, 67

V

VacA (*Helicobacter pylori* vacuolating cytotoxin), 631, 631*f*
  peptic ulcer disease correlations, 634
VacA s1m1 genotype, 636
VacA s1m2 genotype, 636
Validation
  biomarkers, 238
  diagnostic tests, 195
  gene expression profile data, 264
  genomic tests, 195, 235
  pharmacogenetic associations, 182
Valosin-containing protein (*VCP*) mutations, frontotemporal dementia, 691
Valproic acid, 422
Valvular heart disease, 314
VAMP8 polymorphism, myocardial infarction, 294
Variant detection
  whole genome shotgun sequencing, 30, 30*f*, 31*f*
Variation
  copy number *see* Copy number variation (CNVs)
  human populations, 16–19
  HuRef genome analysis, 18
  types, 17, 18*f*, 17*t*
Vascular endothelial growth factor (VEGF)
  acute coronary syndrome risk stratification, 309
  astrocytomas, 526
  cancer therapeutic targeting, 517
  head and neck cancer, 511, 515, 517
  monoclonal antibody targeted therapy, 541
  ovarian cancer angiogenesis, 497
Vascular endothelial growth factor receptor (VEGFR)
  sorafenin inhibition, 544
Vascular-occlusive dementia, 687, 688
VCL mutations (vinculin/metavinculin), hypertrophic cardiomyopathy, 338, |341
VCP mutations, frontotemporal dementia, 691
Vectibix, 231
Velocardiofacial syndrome, 391, 396, 397
Venous thromboembolism
  family history, 380–381
  prognosis, 382–383
  recurrence, 381, 382
  risk factors, 383*t*
  screening investigations, 380
  see also Deep vein thrombosis; Pulmonary embolism
Ventilator associated pneumonia (VAP), 769
Ventricular development genes, 395
Ventricular remodeling, heart failure, 313, 316, 316*f*
Ventricular septal defects, 391, 395
Ventricular tachyarrhythmias, 350
Verapamil, hypertrophic cardiomyopathy, 343
Very low-density lipoproteins (VLDL), 269
  apolipoproteins, 270
  metabolism, 270, 271
V_H mutational status, chronic lymphocytic leukemia, 424, 427
VKORC1, 186
Vincristine
  oligodendrogliomas, 525
  pharmacogenomics, 426
Vinculin (*VCL*) mutations, hypertrophic cardiomyopathy, 338, 341
Viral hepatitis, 774–775
  diagnosis, 777–778
  immunogenetics, 648, 648*t*
  pathogenesis, 778–781
  pharmacogenomics, 782–783
  predisposition, 776–777
  proteomics, 654
  screening, 777–778
  transmission, 776–777
  treatment, 781–783, 781*t*
    genomics studies, 782
    response prediction markers, 782
  virology, 763*f*, 774–776
Viral infections
  immune response, 733
  TLR3 response, 738
Viral load assessment
  HIV infection/AIDS *see* Human immunodeficiency virus (HIV) infection/AIDS
Viral myocarditis
  diagnosis, 318
  heart failure, 314, 318
Vitamin D receptor (VDR), mycobacterial disease susceptibility, 754
VKORC1 polymorphism, warfarin response variability, 180, 181, 183
Vocabularies, 167
  consumer online health information retrieval, 153
Voluntary Genomic Data Submission, 228
  goals, 228
  pharmacogenomics, 183, 224
Von Willebrand factor
  genetic variation, 376

W

Warfarin
  dose adjustment, 224
    genotype-guided, 182, 230
  pharmacogenomics, 180–181, 183, 384–385, 386*t*
WDR36, glaucoma, 716
Wellcome Trust Case Control Consortium (WTCCC), 678, 679
Western blot analysis, 135
WFS1, 198
Whole genome amplified DNA, genotyping technologies, 43
Whole-genome association studies *see* Genome-wide association studies
Whole genome association study (WGAS), 186
Whole-genome gene expression microarrays, 441
Whole genome microarray analysis
  Helicobacter pylori, 629, 634
Whole-genome sequence comparisons, copy number variation (CNVs) detection, 52
Whole genome shotgun sequencing, 28–31
  colony plating/picking, 28–29
  library construction, 28

Whole genome shotgun sequencing, (*Continued*)
process automation, 28, 29*f*
sequence detection, 29
sequencing reactions, 29
template preparation, 29
variant detection, 30–31, 30*f*, 31*f*
whole genome assembly process, 29–30
Williams-Beuren syndrome *see* Williams syndrome
Williams syndrome
congenital heart disease, 394
cytogenetics, 396
Wilson disease, 655
WNK kinase 1/4 mutations, pseudohypoaldosteonism type 2 (Gordon's syndrome), 263
Wnt pathway genes
colorectal cancer, 458
primary biliary cirrhosis, 652

X
X chromosome inactivation, 60
XIST noncoding genes, 22

Y
^{90}Y-ibirtumomab tiuxetan, 529–530
YKL-40
glioblastoma, 527
glioma prognosis, 526–527, 528

Z
Z-disk hypertrophic cardiomyopathy, 338–339, 341
ZAP-70, chronic lymphocytic leukemia, 424, 426–427
Zebrafish, 571
ZFPM2/FOG2, congenital heart disease, 391, 392
ZIC3 (heterotaxy), congenital heart disease, 392–393
Zollinger-Ellison syndrome (gastrinoma), 627