Kevin M. Sweet
Ron C. Michaelis

The Busy Physician's Guide to Genetics, Genomics and Personalized Medicine

Springer

The Busy Physician's Guide To Genetics, Genomics and Personalized Medicine

Kevin M. Sweet • Ron C. Michaelis

The Busy Physician's Guide To Genetics, Genomics and Personalized Medicine

Springer

Kevin M. Sweet
Clinical Cancer Genetics
and Medical Genetics
The Ohio State University
Columbus, OH
USA
Kevin.Sweet@osumc.edu

Ron C. Michaelis
Biology Department
Western Carolina University
Cullowhee, NC
USA
michaelis@email.wcu.edu

ISBN 978-94-007-1146-4 e-ISBN 978-94-007-1147-1
DOI 10.1007/978-94-007-1147-1
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2011927166

I dedicate this book to my wife, Michele and to my children, Gabriella and Isabel for their inspiration, sustaining love and support. I also dedicate this book to my father and maternal grandmother, who passed away during the course of writing, and to my mother, who each in their own way shaped the person I am today.

Kevin Sweet

I dedicate this book to my children, Matt and Kathryn. My desire to be their role model continues to inspire and motivate me in everything I do.

Ron Michaelis

Finally, we would both like to dedicate this book to Dr. Jack Tarleton. Jack has been a great friend, mentor and role model for us both. We greatly appreciate the role he has played in our development as professionals, and wish him continued success in all his endeavors.

Foreword

Senior Vice President for Research in the Office of Health Sciences; Vice Dean for Research in the College of Medicine; Executive Director of the Center for Personalized Health Care, The Ohio State University; Board Member, Personalized Medicine Coalition

Congratulations on completing this important work that contributes significantly to the dissemination of foundation principles in genetics and genomics education. In a widely read article[1] published in 2009, Keyan Salari, a Stanford scholar, argues that rapid advances in the scientific discovery of human genetics and genomics expose a huge gap in the education of clinicians to fully understand the potential as well as limitations of genetics and genomics in medicine practices. This book will help close this gap.

The ability to understand and translate genetics and genomics into clinical practices are a key to developing and implementing personalized medicine. Salari argues that physicians have long used personalized histories like family history, diet, sleep and exercise in their evaluation of a patient to design preventative health and treatment strategies; using this self-reported information with targeted genetic/genomic tests to create personalized medicine has not been well understood or adopted by most clinicians. Individuals now have choices to acquire this information, including that provided by direct-to-consumer genetic testing companies like 23andMe and Navigenics, which challenges our health care system to respond to these personal genetic data and the attendant questions accompanying these data.

In addition to these commercial platforms, direct sequencing of the human genome is approaching the price point where large communities of individuals will choose to have this done. This additional large amount of data will further test an uncomfortable physician workforce to expand consultations with medical geneticists and genetic/genomic counselors to interpret and act on these data. Already, personalized genetic information is being used to improve outcomes in the area of pharmacogenomics and in the treatment of several diseases, including breast cancer. As personal genetic information becomes an increasingly integral component of

[1]Salari K, 2009 The Dawning Era of Personalized Medicine Exposes a Gap in Medical Education. PLoS Med 6(8): e1000138. doi:10.1371/journal.pmed.1000138

the patient medical record, it is becoming more urgent that practicing physicians as well as medical students be educated to use and interpret this information appropriately and responsibly.

In 2008, during the first Personalized Health Care National Conference hosted by The Ohio State University's Center for Personalized Health Care, Kevin shared with me his vision of writing a book. This idea was sparked after one of the guest speakers mentioned rapid technological advancements in genomics would soon overwhelm the 800,000 practicing physicians in the U.S. without sufficient targeted educational interventions in genetics/genomics. As evident in this book, Kevin and Ron accomplish the mission of educating not only practicing physicians and medical students, but also the general public about genetics and diseases and genetic test options, providing them with tools to help them better manage and become more actively involved in their own healthcare decisions.

The Ohio State University Medical Center has embraced the transformation of healthcare delivery through personalized medicine. We strive to change the current reactive mode of care delivery to proactive, P4™ (predictive, preventive, personalized and participatory) Health Care. P4 Health Care utilizes advances in genomics and molecular diagnostics discoveries and provides ***predictive*** information that is necessary to tailor ***personalized*** disease management approaches for each individual, based on genetic, environmental, behavioral and cultural factors. Therapeutics and health management tools are being developed to help ***prevent*** disease instead of merely treating the symptoms. P4 Medicine also promotes health maintenance and wellness, and engages consumers to actively own and ***participate*** in their healthcare decisions.

As we move to develop and implement P4 medicine at Ohio State, we are very fortunate to have a scholar and educator like Kevin whose passion and enthusiasm for genetics research and education are the driving force leading to the germination and completion of this important work. Kevin has made tremendous strides in the field of clinical genetics and genomics and it is truly my pleasure to write the foreword for his book.

I wish Kevin and his co-author, Dr. Ron Michaelis continued successes in translating the science of genomics discovery to a wide range of audiences so that the vision of personalized medicine can be achieved within the next 10–15 years. The future of medicine becoming predictive, preventive, personalized, and participatory is exciting and exhilarating. Kevin and Ron's book begins to pave the road to this future.

Clay B. Marsh, M.D.

Preface

We wrote this book for several reasons. First, we truly believe that we have entered the age of genetics, genomics and personalized medicine, and despite the difficulties that have been encountered in the early stages of the field's development, these revolutionary advances will ultimately improve health care in all fields of medicine. Surveys of practicing physicians consistently report, however, that many practitioners do not feel they know enough about genetics and genomics to apply these personalized medicine principles to their practice. If personalized medicine is ever to live up to its considerable potential, it is essential to provide health care practitioners with the resources they need to educate (or refresh) themselves regarding the foundational molecular biological principles that underlie personalized medicine, and allow them to critically appraise the new information that they will receive from different sources in the near future. We felt we could provide a reference that would review the foundations of personalized medicine, help physicians appreciate both the potential and the limitations of these tests, describe the clinically useful advances that have been made in the field so far, and in the process help health care practitioners better understand how to evaluate the potential clinical usefulness of the tests that will be developed in the future.

Second, we feel that we are at a time when there is a lot of confusion (among physicians and lay people alike) regarding the benefits and limitations of the personalized medicine tests that are available today. Many discoveries get publicized well before they have been developed into clinically useful tests, and some of the tests that commercial companies now advertise provide little to no actionable information. In addition, genetic testing services are advertised directly to consumers, and this has led to an increase in both informed patients and patients who are misinformed regarding the benefits and limitations of personalized medicine testing. Physicians must know what tests are and are not available at the present time, as well as what the benefits and limitations of the currently available tests are, in order to make the proper recommendations in situations in which personalized medicine testing is available.

We hope this book helps you be the best physician you can be.

We would like to gratefully acknowledge the following people for their help and support during the course of writing this book: Kirk Mykytyn, Ph.D., Kandamurugu Manickam, M.D., Amy Curry Sturm, M.S., CGC and Amanda Toland, Ph.D. We also thank the reviewers for their very helpful comments.

Kevin Sweet
Ron Michaelis

Contents

1 Genetic Variability Provides the Biochemical Basis for Our Individuality, Including Differences in Our Susceptibility to Many Common Diseases 1
1.1 Defining and Differentiating Between Genetics and Genomics 1
1.2 The Structure of DNA, the Variability of the DNA Sequence and the Independent Inheritance of Gene Alleles by Siblings 2
1.2.1 The Structure of DNA 2
1.2.2 The Arrangement of Genes on Chromosomes 5
1.2.3 The Polymorphic Human DNA Sequence: Gene Alleles, Protein Isoforms and Genotypes 8
1.2.4 Each Sibling Inherits a Unique Combination of Gene Alleles from the Parent 10
1.3 A Review of the Process Whereby a Gene Makes Its Protein 12
1.3.1 Coding Sequences and Regulatory Sequences 12
1.3.2 Transcription: Deoxyribonucleic Acid (DNA) Makes Ribonucleic Acid (RNA) 13
1.3.3 Posttranscriptional RNA Processing 13
1.3.4 Translation of mRNA into a Polypeptide 15
1.3.5 The Genetic Code and the Structures of Our Amino Acids 16
1.3.6 Posttranslational Processing of the Polypeptide 17
1.4 A Typical Gene's Sequence and the Level of Activity in the Associated Protein Are as Variable as Any Other Human Trait 17
1.5 Risk-Increasing Alleles Have Variable Frequencies and Variable Levels of Penetrance 21
1.6 Polymorphisms in Promoter Regions and Other Non-coding Sequences Influence the Activity of Our Proteins 22
1.6.1 Polymorphisms in Promoter Regions Alter the Gene's Level of Activity 22

1.6.2 Chromosome Rearrangements Can Cause Promoters to Drive Transcription of the Wrong Sequences 24
1.6.3 Intronic Gene Variants That Influence RNA Splicing or Gene Activity May Be Unrecognized Risk Factors 25
1.6.4 Intronic Polymorphisms Can Influence the Ratio of Protein Isoforms or the Balance of Allelic Expression 25
1.6.5 Interfering RNAs Exert an Important Influence over Gene Activity 26
1.7 Epigenetic Factors Also Control Gene Activity 27
1.8 Common Types of Variants in the Human DNA Sequence 29
1.9 Common Multifactorial Diseases Are Genetic Disorders, Despite Their Non-Mendelian Patterns of Inheritance 31
1.10 Personalized Medicine Testing May Allow You to Better Tailor the Treatment to the Individual, and May Allow the Individual to Make Healthier Choices 32
1.11 Summary 33
Further Readings 34

2 Making the Most of Family History Information, Single Gene Disorders and Mendelian Patterns of Inheritance, and When to Refer to a Genetic Specialist 35
2.1 Maximizing the Use of Family Medical History in Disease Risk Assessment 35
2.2 Single Gene Disorders 37
2.3 Understanding Mendelian Patterns of Inheritance (Single Gene Disorders) 38
2.3.1 Autosomal Dominant Inheritance 39
2.3.2 Autosomal Recessive Inheritance 40
2.3.3 X-linked Recessive 41
2.3.4 X-linked Dominant Inheritance 42
2.3.5 Y-Linked Inheritance 43
2.4 Assessing the Risk of Recurrence in Mendelian Pedigrees 44
2.5 Carrier Frequencies for the More Common Recessive Single-Gene Disorders 45
2.6 Referring to a Genetic Specialist 46
2.7 New Genomic Applications for Complex Disease Will Change Approaches to Genetic Counseling and Personalized Medicine 47
Selected References 49

3 Types of Genetic Tests and Issues Associated with the Interpretation of Their Results 51
3.1 Accessing Current Information on Available Personalized Medicine Tests 51

3.2 Standard Format for Genetic Test Results 53
3.3 Risk-Increasing Gene Alleles Often Have Limited Penetrance 55
3.4 The ACCE and EGAPP Projects Evaluate Emerging Genetic Tests 56
3.5 Assessing the Usefulness of a Genetic Test 57
3.6 Even an Informative Genetic Test May Have Limited Clinical Utility 59
3.7 Single Nucleotide Polymorphisms (SNPs) Are the Most Commonly Tested Polymorphisms 59
3.8 There Are Many Small Deletions and Insertions in Different People's DNA 60
3.9 Repeated Sequence Length Polymorphisms and Microsatellite Analysis 61
3.9.1 The Repeated Sequence Motif Can Vary in Length 61
3.9.2 Microsatellite Instability (MSI) Is Observed in Certain Types of Cancer 62
3.9.3 Loss of Heterozygosity (LOH) of Microsatellites Is Seen in Several Disorders 63
3.10 Chromosome Rearrangements Can Contribute to Some Complex Disorders 64
3.11 Copy Number Variation Is Surprisingly Frequent 64
3.12 It Is Sometimes Necessary to Determine the Level of Activity in Specific Genes 64
3.13 Mitochondrial DNA Variants Are Also Relevant 65
3.14 Many Epigenetic Factors That Influence Gene Activity Are Amenable to Testing 65
3.15 Some Tests Assess Characteristics of the Pathogen 66
3.16 Cancer Analyses Often Must Include Somatic Mutations as well as Germline Mutations 66
3.17 Predictive Algorithms Must Include both Genetic and Nongenetic Factors 66
3.18 Genome-Wide Association (GWA) Studies Provide Insights into the Mechanisms for Disease, But Their Results Are Often Not Clinically Useful 67
3.18.1 Direct-to-Consumer GWA Testing Services May Provide Results That Have Limited Clinical Utility 67
3.18.2 GWA Studies Identify Risk-Increasing Alleles, But Have Their Limitations 67
3.18.3 Measures of the Association Between the Risk-Influencing Allele and the Disorder/ADR 69
3.18.4 Many of the SNPs That Are Used in GWA Studies Are Not Themselves Functional Polymorphisms, But Are Linked to Functional Polymorphisms 70

3.18.5 Haplotype Blocks in the Human Genome Increase the Efficiency of GWA Studies 72
3.18.6 Internet Resources That Summarize Findings from GWA Studies 73
3.19 A Brief Introduction to the Most Important Technological Advances 74
3.19.1 The Polymerase Chain Reaction (PCR) Allows One to Isolate the Sequence of Interest; Many Assays Begin with the PCR 74
3.19.2 Microarray Analyses Have Greatly Accelerated the Pace of Discovery 75
3.19.3 Sequencing Provides Maximum Information, and Will Revolutionize Clinical Diagnostics 76
3.19.4 There Are Several Techniques Commonly Used to Assess Status for a SNP 77
3.19.5 PCR Simplifies Analysis of Indels and Length Polymorphism and Detection of MSI and LOH 78
3.19.6 Hybridization Techniques Detect CNVs 78
3.19.7 Reverse Transcription-PCR Measures the Level of Activity in a Gene 80
3.19.8 Immunohistochemical (IHC) Analyses Allow Direct Visualization of the Protein 80
3.19.9 Expression Arrays Reveal Disease-Associated and Treatment-Associated Changes in Gene Expression 81
3.19.10 Some Epigenetic Factors Can Be Easily Assessed 82
3.19.11 G-Banding Allows the Detection of Chromosome Rearrangements 82
Further Readings 83

4 Toward the Safer and More Effective Use of Prescription Drugs: Pharmacogenetics 85
4.1 Genetic Polymorphisms Affect Both the Pharmacokinetics and Pharmacodynamics of Many Prescription Drugs 86
4.2 Improving on the Disease-Oriented Approach to Prescribing Drugs 88
4.3 Limitations of Genetic Testing 93
4.4 Dose-Calculating Algorithms Must Take Genetic and Nongenetic Factors into Account 95
4.5 Epigenetic Factors Must Be Factored into Many Algorithms as Well 97
4.6 Polymorphisms in the CYP450 Genes Influence the Pharmacokinetics of Many Commonly Prescribed Drugs 98
4.6.1 The CYP450 Enzymes Metabolize Many Commonly Prescribed Drugs 98
4.6.2 Defining the Metabolizer Phenotype by Assessing CYP450 Status 98

4.7 Other Functional Polymorphisms That Affect the Pharmacokinetics of Multiple Drugs 101
4.7.1 N-Acetyltransferase 2 101
4.7.2 Butyrylcholinesterase 102
4.7.3 Functional Polymorphisms in Drug Transporter Genes also Affect the Response to Many Drugs 102
4.8 Polymorphisms in the Genes Encoding Beta-Adrenergic Receptors Influence the Pharmacodynamics of Beta-Blockers 104
4.9 Keeping up to Date with FDA Approvals and the Status of the Field 105
Further Readings 106

5 Taking a Personalized Medicine Approach to Breast and Colon Cancer 107
5.1 Cancer Is a Complex Genetic Disease 107
5.2 Breast Cancer Gene Variants with Low Penetrance 109
5.3 Further Research Will Increase Accuracy and Standardize Risk-Estimating Algorithms 110
5.4 Useful Online Programs to Calculate Breast Cancer Risk 112
5.5 Highly Penetrant Breast Cancer Gene Variants 113
5.6 Hereditary Breast-Ovarian Cancer Syndrome 114
5.7 *BRCA* Gene Testing 116
5.8 Effects of Possessing Risk-Increasing BRCA Alleles 118
5.9 Cancer Screening and Prevention Measures for Female *BRCA* Variant Carriers 119
5.10 *BRCA* Mutation Positive Case Study 120
5.10.1 Mrs. C's Initial Meeting with the Genetic Counselor 122
5.10.2 Interpreting the Test Results and Following Them Up 123
5.10.3 The Plans for Mrs. C and Her Family Members 125
5.10.4 Further Developments 126
5.11 Colon Cancer Gene Variants with Low Penetrance 126
5.12 Highly Penetrant Colorectal Cancer Gene Variants 127
5.13 Lynch Syndrome 127
5.14 Molecular Genomic Testing in Patients Suspected of Having Lynch Syndrome 129
5.14.1 Microsatellite Instability (MSI) Is a Hallmark of Lynch Syndrome 129
5.14.2 Some Labs Offer Complementary MSI and IHC Analyses 130
5.15 Cancer Screening and Prevention Measures for LS Mutation Carriers 131
5.16 Familial Adenomatous Polyposis 131
5.17 *APC* Mutation Screening in Familial Adenomatous Polyposis 132
5.18 Cancer Screening and Prevention Measures for APC Mutation Carriers 133
5.19 Personalizing Drug Therapy for Cancer Patients 134

5.19.1 Gene Expression Assays can Classify Cancers into Molecular Subtypes 134
5.19.2 Several Gene Polymorphisms Affect Activity in Signal Transduction Pathways 135
5.19.3 Glioblastomas with *MSH6* Mutations Resist Alkylating Agents 137
5.19.4 Pharmacokinetically Relevant Gene Polymorphisms Influence Drug Response, Especially the Risk for ADRs 137
5.19.5 HLA Type May Influence the Risk for Hypersensitivity Reactions After Abacavir 139
Further Readings 139

6 Personalizing Risk Assessments and Treatments for Complex Cardiovascular Disease 141
6.1 Cardiovascular Diseases Are Complex, Multifactorial Diseases with Highly Variable Phenotypes 141
6.2 Family History and CVD Risk 143
6.3 Useful Online Programs to Estimate Heart Disease Risk 144
6.4 CVD-Associated Gene Variants Have Been Particularly Difficult to Identify 145
6.5 Most CVD-Associated Gene Variants Have Low Penetrance 146
6.5.1 A Cluster of Linked Markers in 9p21.3 Are Associated with Several CVDs 146
6.5.2 Other Low-Penetrance Variants That Influence Risk for CVD 148
6.6 Low-Penetrance Gene Variants That Affect Predisposing Phenotypes 149
6.6.1 Low-Penetrance Gene Variants That Alter Blood Lipid Levels 149
6.6.2 Sodium/Potassium Regulation and Essential Hypertension 149
6.6.3 Cardiac Channelopathies 150
6.6.4 Variants That Influence the Inflammation Response Influence Risk for CVDs 152
6.6.5 Genetic Variants Influence the Level of Oxidative Stress 152
6.6.6 Plasma Homocysteine Levels Influence Risk for CVD 153
6.6.7 Knowing They Possess Low-Penetrance Variants May Motivate Patients' Behavior 153
6.7 Case Report – Genetic Testing in a Patient with Type 2 Diabetes (T2D) and a Family History of Myocardial Infarction (MI) 153
6.7.1 Jameer's Initial Visit 154
6.7.2 Meeting with the Cardiologist and Genetic Counselor 155

6.7.3 The Plan for Follow-Up 157
6.8 CVD Gene Variants with High Penetrance: Familial Hypercholesterolemia (FH) 157
6.8.1 FH Is Often Underdiagnosed 158
6.8.2 Highly Penetrant FH Gene Variants 158
6.9 Molecular Genomic Testing in Patients Suspected of Having FH 161
6.10 CVD Gene Variants with High Penetrance: Hypertrophic Cardiomyopathy (HCM) 161
6.10.1 The HCM Phenotype Is Highly Variable 161
6.10.2 Highly Penetrant Gene Variants That Increase Risk for HCM 162
6.11 Therapy and Prevention Measures for HCM Mutation Carriers 164
6.12 Genes Influencing the Risk for Other Cardiomyopathies 164
6.13 Cardiovascular Pharmacogenomics 165
6.13.1 The CYP450 Enzymes Metabolize Several Drugs That Are Prescribed for CVDs 165
6.13.2 The Pharmacogenomics of Cholesterol-Lowering Drugs 166
6.13.3 Pharmacogenomic Testing Is Particularly Important for Patients Taking Warfarin 169
6.13.4 The Pharmacogenomics of Antihypertensive Drugs 170
6.13.5 Plasma Homocysteine Levels Influence Pyridoxine Response 171
6.13.6 CYPC19 Testing for Patients Prescribed Clopidogrel 171
Further Readings 171

7 Other Multifactorial Disorders for Which Genetic/Genomic Testing Is Providing Insights 173
7.1 Age-Related Macular Degeneration 173
7.2 Type 2 Diabetes 174
7.2.1 GWA Studies Have Identified Several Genetic Markers That Increase Risk for T2D 175
7.2.2 The Currently Available Predictive Tests Do Not Include Genetic Markers 177
7.2.3 The Pharmacogenomics of Type 2 Diabetes 177
7.3 Personalizing the Approach to Psychiatric Disorders 178
7.3.1 Many Original Findings Must Be Confirmed 178
7.3.2 Testing Genetic Variants That Influence Pharmacokinetics Is Helpful in the Treatment of Some Psychiatric Disorders 179
7.3.3 Some Associations with Attention Deficit-Hyperactivity Disorder (ADHD) Have Been Confirmed 179
7.3.4 Polymorphisms in Genes That Influence Serotonergic and Dopaminergic Function May Influence Personality 180

7.3.5 Copy Number Variation May Influence the Risk for Schizophrenia 180
7.3.6 Several Polymorphisms May Influence the Risk for Schizophrenia and the Response to Antipsychotic Drugs 181
7.3.7 Several Polymorphisms Have Been Associated with Depression and Bipolar Disorder (BPD) 181
7.3.8 Polymorphisms in the Serotonin Transporter (5HTT) Gene May Not Influence One's Risk for Depression After All 183
7.3.9 Polymorphisms in the Monoamine Oxidase a (MAOA) Gene May Not Influence the Propensity toward Antisocial Behavior after All 186
7.3.10 Anxiety Disorders 186
7.4 Personalizing the Treatment for Substance Abuse 186
7.4.1 Most Studies Have Focused on Catecholamine and Opiate Pathways 186
7.4.2 Stimulants 187
7.4.3 Alcohol 189
7.4.4 Opiates 190
7.4.5 Nicotine 191
7.5 Alzheimer Disease and Cognitive Decline in Aging 192
7.5.1 The ApoE Gene Is a Known Risk Factor 193
7.5.2 Other Gene Variants That Influence Risk for AD and Cognitive Decline in the Elderly 193
7.6 Asthma and Other Respiratory Disorders 196
Further Readings 198

Epilogue and List of Resources 199

Index 207

Introduction

It is far more important to know what person the disease has than to know what disease the person has.

—Hippocrates

You Want to Provide Your Patients with the Best Care Possible…

We all know that health care providers want to provide their patients with the safest and most effective treatments possible. Unfortunately, in all fields of medicine a significant percentage of patients either do not improve or suffer adverse effects from their prescribed treatment. Doctors prescribe as wisely as they can, but they often have limited, anecdotal evidence to document their own patients' experiences, and limited time to keep up with the vast and ever-changing body of literature in their field. When deciding how to treat a patient, they often have no choice but to begin with a standard first-line treatment, pursue an iterative trial-and-error strategy and react to the events that unfold.

Personalized medicine involves the capacity to use new molecular biological principles and techniques to identify genetic susceptibilities to common diseases before symptoms appear, and better tailor medical treatments to the individual characteristics of each patient. Personalized medicine is based upon the principles of genetics and genomics, which expands genetics to include studies of DNA, RNA, proteins and other molecules that interact with DNA. In the coming years, personalized medicine tests will be developed that incorporate not only DNA sequence information, but information from RNA and protein tests. The most clinically useful tests will also incorporate clinical data, personal information and family history into their predictive algorithms. These tests will strengthen your ability to actively promote your patients' health and well-being over their entire lifespan.

In order for personalized medicine to be truly personal, however, patients must also become active participants in their own health care. The patient should be made aware of the fact that the vast majority of common disorders are multifactorial

disorders, usually resulting from a combination of genetic and nongenetic factors. The physician must be able to understand genetic risks in the context of the patient's care. As these genetic risks are not modifiable and can sometimes be distressing, these risks must be delivered in a sensitive and compassionate manner in the context of genetic counseling. The patient needs to understand what is at stake and what their genetic risk factors mean not only to their own health but also in the context of their family. In addition, he/she must be able to help the patient understand the importance of the nongenetic factors, and motivate the patient to modify his/her exposure to the critical nongenetic factors if possible.

…And This Book Can Help

Unfortunately, many practicing physicians feel that they do not have a strong background in genetics and genomics. A recent review in the *Journal of the American Medical Association* (*JAMA*)[2] noted that:

> The most important and consistent finding from our literature review is that the primary care workforce, which will be required to be on the front lines of the integration of genomics into the regular practice of medicine, feels woefully underprepared to do so.

This book is intended to enable you to put these principles of personalized medicine into your practice, regardless of how long it has been since you had your education in genetics. For those who feel the need to refresh their basic education in genetics, this book provides you with a thorough review of the principles of genetics that underlie personalized medicine. This includes the basic molecular biology you need to know in order to understand the genetic variability that underlies personalized medicine, the fundamental principles of inheritance that determine genetic risk, and some guidance on how to use family history to better estimate the patient's susceptibility to disease.

The importance of family history is often overlooked, in part because family history can take time to collect. To help make the collection and implementation of this information easier, we list several web-based programs that make it easy to input family history and draw a pedigree. You can share these websites with your patients, as a means of enabling them to arrive at your office with the relevant family history information diagrammed as an annotated pedigree that will allow for your review and analysis.

After reviewing the necessary foundational material, we provide a wide-ranging review, not only of the genetic tests that are available for patients today, but of the discoveries that will give rise to the next wave of personalized medicine tests. We hope that, after finishing this book, each reader will feel confident that he/she understands what can be done today to improve the level of patient care, and can critically appraise the new scientific information that will be available in increasing abundance in the near future.

---

[2]Scheuner, et al., 2008. JAMA 299(11):1320–1334.

A Great Deal of Work Remains, but the Principle Has Clearly Been Proven

Medical journal editorials frequently emphasize the limited progress that has been made in personalized medicine, and the limited predictive utility of the newer personalized medicine tests. These criticisms are largely valid at the present time; even the most staunch advocates of personalized medicine admit that very few basic research discoveries have been translated into tests with true clinical utility. The human DNA sequence has turned out to be even more variable than most scientists thought it would be, and the sheer number of gene variants that influence the activity of proteins, and therefore influence risk for disease or response to treatments, has exceeded most people's estimates. In addition, recent research has revealed that the activity of our genes is regulated by genomic factors other than the sequence of the DNA, including interfering RNAs and epigenetic factors such as the methylation of DNA. At the present time there have been many useful discoveries made regarding the effects of DNA sequence variations on one's risk for diseases or response to treatments. As genomic researchers uncover more of the factors that influence the activity of our proteins, however, personalized medicine tests will expand to include assessments of RNA, proteins and chemical modifications of the DNA.

It is also clear, however, that properly designed personalized medicine testing can improve diagnostic accuracy as well as the safety and efficacy of treatments. Several of the personalized medicine tests that are currently available are already personalizing diagnoses and treatments in fields like oncology and cardiology, and informing medication dosing recommendations in many fields of medicine. In addition, researchers are providing new discoveries every day, from the foundational molecular biology to "translational research," which emphasizes the best way to translate molecular biological discoveries into clinically useful tests.

The pace of this research will only increase in the future. The cost of these procedures is declining rapidly, and as it does, researchers will collect ever-increasing amounts of data. In addition, as personalized medicine testing becomes increasingly cost-effective, it will also become more widespread, further accelerating the pace of clinical research.

There are no insurmountable obstacles to the further progress of personalized medicine. The fact that the best predictive algorithms will often require collecting data on thousands of contributory gene variants can easily be accommodated by microarray-based SNP analyses, which enable the analyst to test hundreds of thousands of gene variants in one assay, or whole-genome sequencing, which provides information on an individual's entire complement of genes. In addition, a number of dietary, environmental and lifestyle risk factors that contribute to many diseases are already known, and future research will undoubtedly uncover many more.

Some personalized medicine tests have now become well established and have good clinical utility. Some of the personalized medicine tests that are currently available, however, are limited in their usefulness, because pressure to commercially exploit new discoveries has caused companies to bring what they could to the market

quickly, to establish their presence in the market, rather than to wait until the testing evolved to a form with better clinical utility. For example, with respect to pharmacogenomic tests, it has been easier to determine the genetic variants that affect the pharmacokinetics of drugs than variants that affect their pharmacodynamics, because one merely needs to measure the levels of the drug in the blood to determine the effects of the individual's genetic status, rather than some aspect of the drug response that requires more invasive measurement techniques. Consequently, most of the pharmacogenetic tests that are well-developed enough to be marketed focus on pharmacokinetically relevant gene variants. In addition, even tests that include pharmacokinetically and pharmacodynamically relevant gene variants rarely combine this genetic information with family history, clinical data or diet, environment and lifestyle data.

We have come to recognize that most common diseases are multifactorial diseases, and that genetic factors only constitute a portion of the factors that influence the individual's risk for any given disorder. Because nongenetic factors influence the individual's risk for many diseases and response to many drugs, there will always be a limit to the predictive value of tests that use only genomic information. Genetic and genomic tests should not be seen as replacements for conventional predictors such as family history, age and clinical data, but rather as additional weapons in the arsenal. The most clinically useful tests will combine genetic/genomic information with family, personal and clinical data to maximize predictive power.

The Necessary Infrastructure Is Evolving

In the last few years, the institutional, educational and legal infrastructure that is necessary to support personalized medicine has begun to evolve. A number of medical schools have developed genetic and genomic medicine training programs. In addition, the federal government and several states have initiated programs to help educate practicing physicians. The Genetic Nursing Credentialing Commission offers a program whereby practical nurses and registered nurses can be certified as specialists in genetics. A number of medical centers have developed Centers devoted to the practice of personalized medicine, and several hospitals have adopted policies which use personalized medicine tests to guide treatment decisions from the beginning of the patient's care.

Major changes in the health care system are not possible without the involvement of the federal government, and the new federal administration has made it clear that it appreciates the ability of personalized medicine to improve efficiency and reduce the cost of health care and new drug development. The Genomics and Personalized Medicine Act was introduced to Congress in March, 2007, by then-senator Barack Obama. Now that Mr. Obama is President, he has included in the American Recovery and Reinvestment Act of 2009 a plan to spend $19 billion to upgrade the nation's medical information technology and create electronic health records, in part to enable the more effective and efficient use of genetic testing data to reduce the cost of health care. This dovetails with the Department of Health and

Human Services' Personal Health Care Initiative (PHCI), which began in March 2007. The PHCI is intended to encourage better communication between basic researchers and clinical researchers, improve the information technology in the health care industry, and protect individuals from misuse of their genetic information. Lastly, the director of the National Institutes of Health (NIH), Dr. Francis Collins, who was instrumental in overseeing the completion of the Human Genome Project, now plans to emphasize five "themes," including health care reform and translating genomic research into medicine, as his platform.

We are increasingly recognizing the need for better communication between basic researchers and clinicians, and programs are slowly developing to integrate the different subdomains of several fields. For example, the National Cancer Institute's (NCI) Biomedical Informatics Grid initiative, which began in 2004, is building a network of communication between research laboratories, clinical laboratories, academic centers and private corporations that will maintain an integrated cycle of discovery, application and feedback that will make the process of going from discovery to clinical application more effective and efficient. In addition, the Centers for Disease Control and Prevention (CDC) is developing a process for evaluating the analytic validity, clinical validity, clinical utility and ethical, social and legal implications of genetic tests.

Major changes in the health care system are also not possible without the cooperation of the health insurance industry. This, too, is moving forward as it becomes obvious that genetic medicine can reduce health care costs throughout the individual's lifespan. Better understanding of the genetic basis for disease risk can help some people tailor their diet, environment and lifestyle to reduce their preventable risk of diseases for which their genetic susceptibility is greatest, avoiding the cost of treatment. Further, by informing decisions about the choice and dose of medications, genetic tests can reduce the costs associated with ineffective treatments and adverse side effects.

The American Association of Health Plans, as well as individual insurers such as Aetna, United Health and Kaiser Permanente, have already recognized that personalized medicine can sharply reduce the cost of health care for many individuals. Some insurers are now paying for pre-symptomatic genetic tests that can predict risks and guide treatment decisions before the patient begins to exhibit symptoms of the disorder. In addition, the CDC is working with the insurance industry to help resolve issues related to approval of and payment for tests.

Finally, we predict that public interest will contribute to the development of personalized medicine to a considerably greater degree than public interest usually contributes to the development of medical fields. Aided by the popularity of forensic television shows, public acceptance of genetic testing is growing. Having one's genome screened has even become fashionable in some well-to-do circles, and as the cost of genome screening or sequencing drops, many others will become interested as well. Some people are sharing the results of their genomic scans on social network websites, or signing up for involvement in online research studies at commercial genomic companies. As the number of people who decide to have genetic/genomic analyses and share their information increases, the accumulating data will give rise to new discoveries and developments.

One of the biggest obstacles to the public's willingness to undergo genetic testing and share data has recently been addressed. Many people have feared that their genetic information could be used to deny them a job or health insurance policy because they possess genetic variants that increase their risk for a particular chronic disease. The federal Genetic Information Nondiscrimination Act of 2008 has instituted safeguards against the unethical use of genetic information, and set punishments for unauthorized agents who try to access genetic information. Now that this obstacle has been overcome, the public's acceptance of genetic testing, and willingness to undergo testing and share data, may increase. In addition, as the public gets better educated about multifactorial diseases and preventable nongenetic factors, people may also become more likely to take a more active role in their own medical choices, and to promote their own health and wellness.

Chapter 1
Genetic Variability Provides the Biochemical Basis for Our Individuality, Including Differences in Our Susceptibility to Many Common Diseases

Abstract The human DNA sequence is highly polymorphic; for a typical gene, different people have different specific sequences (or alleles) for that gene. As a result, for the typical protein, different people have different levels of activity in that protein. If the activity of a protein is significantly greater or less than the level of activity that is seen in the typical person, that individual will have a greater or lesser susceptibility to the diseases the protein's function influences than the typical person does. Whether a high-activity or low-activity gene allele represents a risk-increasing allele or a risk-decreasing allele depends on the specific function the protein performs. While some risk-increasing alleles are relatively common, and some increase the individual's risk for the associated disease dramatically, most risk-increasing alleles are relatively rare, and most only increase the individual's risk for the associated disease by a small amount. This chapter contains a thorough review of the foundational material that the reader must understand in order to understand the aforementioned principles. This includes a review of the process whereby a gene makes a protein, the role of promoter and other regulatory sequences in influencing gene activity, and the role that intronic sequence variants, interfering RNAs and epigenetic factors play in regulating gene activity.

1.1 Defining and Differentiating Between Genetics and Genomics[1]

Genetics deals with the process whereby our genes make our proteins, the effects that changes in our chromosome structure or DNA sequence have on our health, and the inheritance of genetic variants through the generations. "Genomics" is a much broader term. The term "genome" is used in several contexts: it can refer to

[1]These are not comprehensive definitions of these terms. These definitions are merely intended to illustrate several important differences in the way the terms "genetics" and "genomics" are used.

K.M. Sweet and R.C. Michaelis, *The Busy Physician's Guide To Genetics, Genomics and Personalized Medicine*, DOI 10.1007/978-94-007-1147-1_1,

an individual's entire DNA sequence, the collection of genes a particular species possesses, or the specific versions of gene sequences an individual possesses. A genomic researcher studies not only the process whereby a gene makes a protein, but also how that gene and that protein fit within the context of the rest of your genes, proteins and other biomolecules.

Genomics expands the field of inquiry to include studies of the mechanisms whereby gene activity is regulated, such as the chemical modifications of cytosine (C) nucleotides or histone proteins (discussed below). Genomics also emphasizes that all proteins work in concert with other proteins and other biomolecules. In order to understand the way in which our genes affect our risk for a particular disorder, or our likelihood for a successful response to a particular treatment, we need to study all the genes and proteins that participate in the metabolic pathway(s) that underlie the pathophysiology of that disease, influence the patient's response to the treatment, and maintain the tissues and pathways that are affected by the disease or acted upon by the treatment.

At this point in time, most of the clinical offerings for personalized medicine focus on genetic tests. As continued research leads to increased understanding of our molecular machinery (genetic, epigenetic, RNA, protein and other biomarkers), the field will expand to include true genomic tests that have high clinical utility. In addition, new algorithms will be developed that reflect new discoveries about the way these genes, proteins and other molecules interact with dietary, environment and lifestyle factors to allow for more extensive patient profiling at point of contact.

1.2 The Structure of DNA, the Variability of the DNA Sequence and the Independent Inheritance of Gene Alleles by Siblings

1.2.1 The Structure of DNA

The building blocks of DNA and RNA are called *nucleotides* (Fig. 1.1). A DNA nucleotide is made from three components: the sugar *deoxyribose*, a *phosphate group*, and a *nitrogen-containing base*. There are four different types of nucleotides in DNA. They all have the same deoxyribose and phosphate group, but they differ in the type of nitrogenous base they contain: some contain the base *adenine*, some contain *cytosine*, some *guanine* and some *thymine*.

Although the deoxyribose and phosphate group are important in their own right, for our purposes, the base is the important part of the nucleotide. The sequence of bases determines which stretches of your DNA serve as genes, and what specific proteins they make. In fact, because the base is the important portion of the nucleotide for our purposes, the terms "base" and "nucleotide" are often used interchangeably in the literature. For convenience, we can refer to the bases, and the nucleotides that contain them, by the letters "A," "C," "G" and "T."

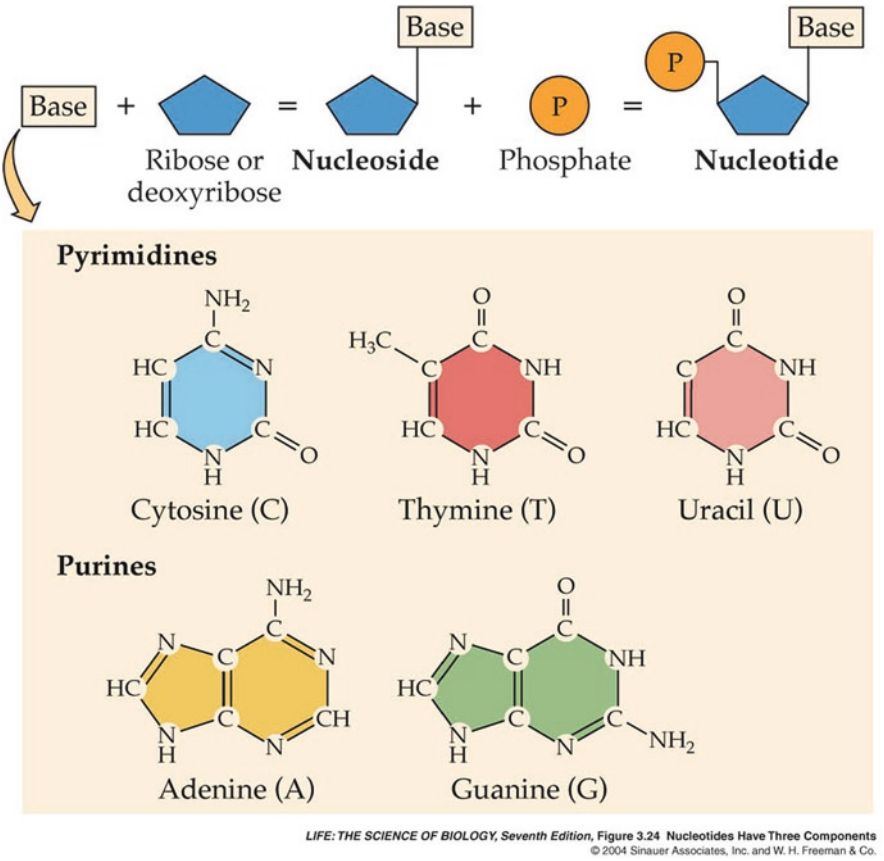

Fig. 1.1 A nucleotide consists of a deoxyribose, a phosphate group and a nitrogen-containing base (Reprinted from Life: The Science of Biology, 7th ed. Purves, Sadava, Orians and Heller. Copyright Sinauer Associates, 2004)

When a DNA strand is synthesized, each nucleotide that is added to the strand is attached by the formation of a bond between the phosphate group that is attached to the 5′ carbon[2] of the new nucleotide and the oxygen that is attached to the 3′ carbon of the nucleotide to which it is being attached. Because of this, a strand of DNA is said to have a *5′ to 3′ orientation*. The *5′ end* refers to the end at which the first nucleotide that was incorporated into the strand lies; there is a phosphate group on the 5′ carbon at this end of the molecule. The *3′ end* refers to the end at which the last nucleotide that was incorporated into the DNA strand lies; this end of the molecule has a hydroxyl group on the 3′ carbon.

[2] The carbons in the base are given the numbers 1, 2, 3, etc. In order to distinguish between the carbons in the base and the sugar, the carbons in the sugar are designated as 1′, 2′, 3′, etc.

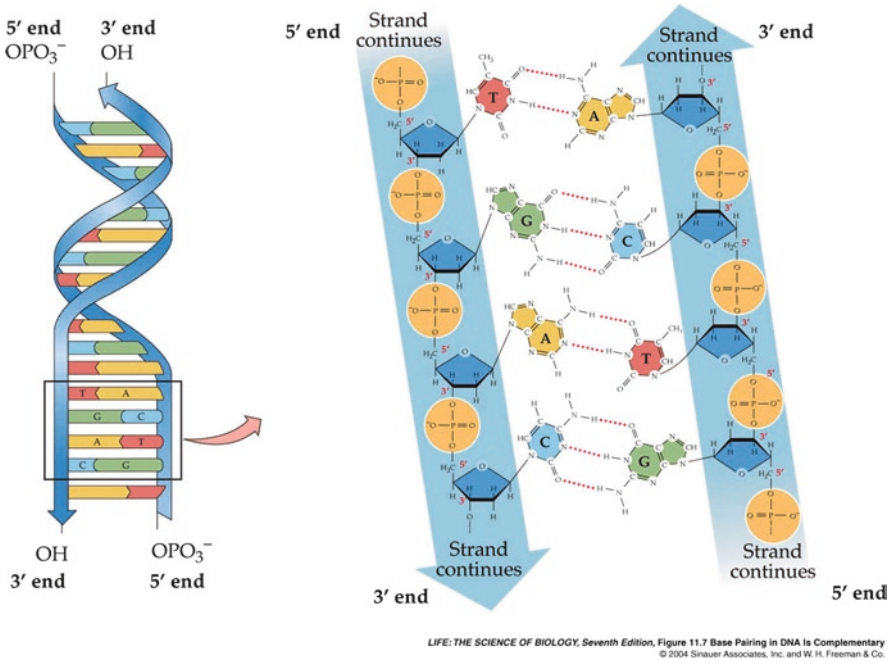

Fig. 1.2 The two strands of a DNA molecule lie in an antiparallel configuration (Reprinted from Life: The Science of Biology, 7th ed. Purves, Sadava, Orians and Heller. Copyright Sinauer Associates, 2004)

DNA consists of two strands of nucleotides that lie in an *antiparallel* orientation to each other (Fig. 1.2). An antiparallel orientation is the same orientation one sees in a two-lane road, in which there are two parallel lanes, with an opposite direction of traffic flow in each. In a double-stranded DNA molecule, each of the two DNA strands has the opposite 5′ to 3′ orientation.

Notice that hydrogen bonding between the bases of the two different DNA strands holds the double helix together. Notice also that As bond with Ts and Cs bond with Gs; they are said to bond as *complementary basepairs*. Two nucleic acid sequences are said to be complementary if their sequences contain complementary bases, allowing them to bind together to make a double-stranded molecule. For example, the two following sequences are complementary, and illustrate the antiparallel arrangement of the two nucleic acid strands.

5′-ACTGATCTGAGGCACTC-3′
3′-TGACTAGACTCCGTGAG-5′

Because of the bonding forces that arise when nucleotides are strung together, each strand of DNA adopts a helical shape; the descriptive term “double helix” is often used to describe DNA’s configuration. Because the DNA molecule is

double-stranded, the physical size of a gene or the physical distance between two genes is often stated in *basepairs* (abbreviated "bp"). For example, you might say gene A and gene B lie two million bp apart, or the coding sequence of gene X is 150,000 bp long.

1.2.2 The Arrangement of Genes on Chromosomes

There are approximately 3.3 billion nucleotides in a single strand of human DNA, and therefore approximately 3.3 billion bp of DNA in the double-stranded DNA molecule. These 3.3 billion bp of DNA are distributed among 23 chromosomes, each of which is a separate molecule of DNA.

Figure 1.3 illustrates how the DNA molecule is condensed into chromosomes. The DNA is wound around a cluster of *histone proteins* to make a *nucleosome*. The string of nucleosomes is then supercoiled to such a degree that 46 chromosomes, which if laid out in linear orientation would make a molecule approximately 6.6 ft (2.04 m)long, can fit inside a microscopic cell nucleus. The combination of DNA and proteins is called *chromatin*, which is held in this supercoiled configuration by several specialized proteins.

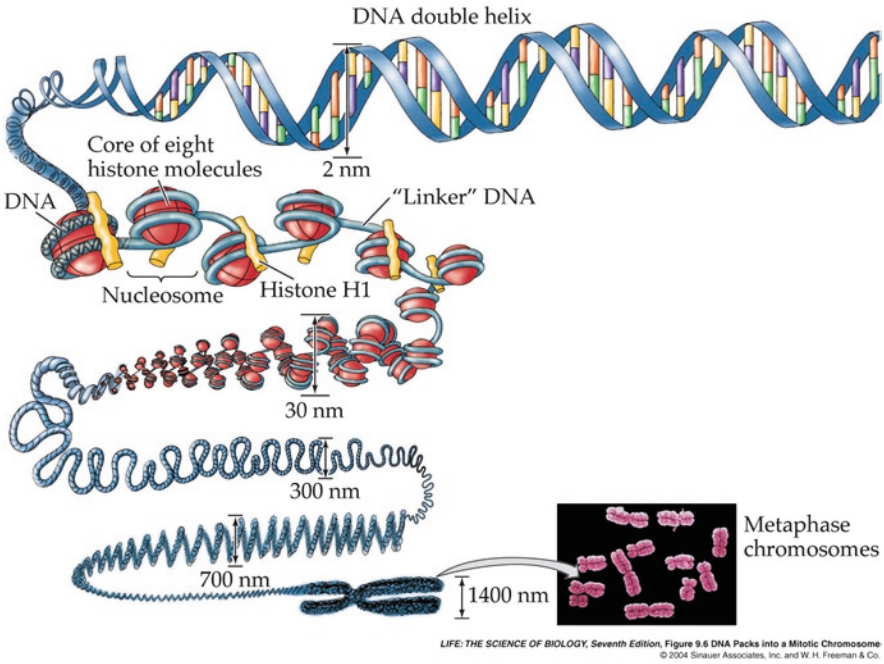

Fig. 1.3 DNA is wrapped around histone proteins and supercoiled into chromosomes (Reprinted from Life: The Science of Biology, 7th ed. Purves, Sadava, Orians and Heller. Copyright Sinauer Associates, 2004)

One complete set of 23 human chromosomes is often referred to as a *haploid genome*. Because most human cells contain two full sets of chromosomes, they are said to be *diploid*. The numbered chromosomes, 1–22, constitute the *autosomes*, and the X and Y chromosomes are referred to as the *sex chromosomes*. Males and females both have 22 pairs of autosomes. In addition, females have a pair of X chromosomes as their sex chromosomes, while males have one X and one Y chromosome. Chromosomes are usually depicted (Fig. 1.4) as having a short *p arm* on top, a longer *q arm* on the bottom, and a constriction called the *centromere* between them. At the end of each p arm and q arm is a region called the *telomere*.

The supercoiling of the chromatin not only enables you to store a large amount of DNA in a microscopic cell nucleus, it is also part of the mechanism whereby the activity of your genes is controlled. We will return to the histone proteins later, when we discuss some of the ways in which chemical modifications of histone proteins influence the activities of genes.

A *gene* is a stretch of DNA sequence that makes an *RNA*.[3] A gene may be anywhere from a few hundred to a few million bp long. A chromosome, on the other hand, is between 33 million and 250 million bp long, and can contain several hundred to several thousand genes. There are approximately 21,000–23,000 genes in the human genome.

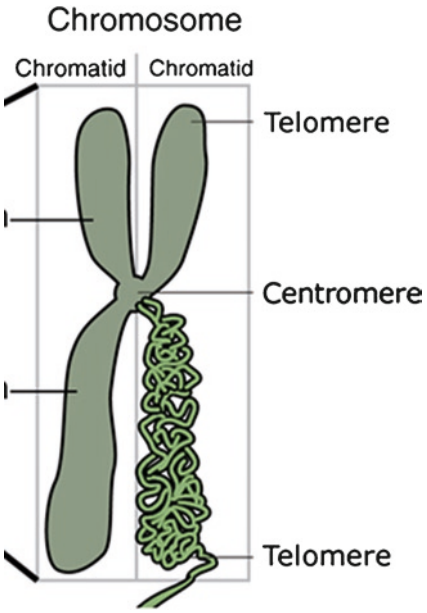

Fig. 1.4 Diagram of a human chromosome, illustrating the p arm, q arm, centromere and telomeres (From mstiboldo.blogspot.com)

[3] Most RNAs then go on to make proteins, but some RNAs function as enzymes (ribozymes), play critical roles in protein synthesis (rRNA and tRNA) or regulate the production of other genes' proteins (microRNAs).

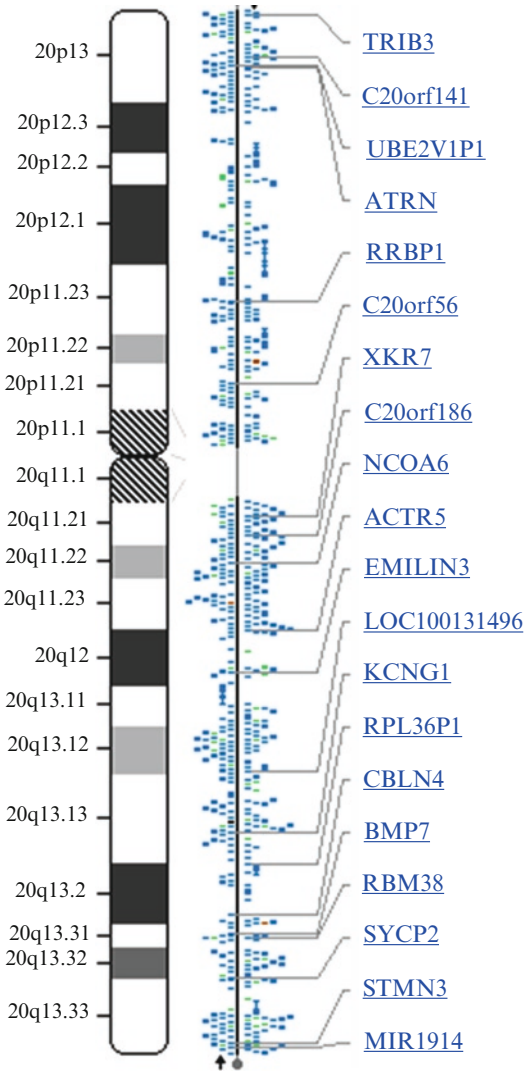

Fig. 1.5 Diagram illustrating the distribution of genes along chromosome 20 (From the National Center for Biotechnology Information)

Figure 1.5 illustrates some of the genes on chromosome 20. The genes that are on chromosome 20 are represented by rectangles to the left and right of the line in the middle of the figure. As you can see, there are many more genes on chromosome 20 than can be listed on the right side of the figure. As you can also see, some regions of chromosome 20 have a denser concentration of genes than other regions do. As a rule of thumb, when an individual is missing a region of a chromosome, or has an extra copy of a region of a chromosome, the larger the deletion or duplication,

the more severe its consequences. Because genes are unevenly distributed through each chromosome, however, it is possible for a smaller deletion or duplication involving a gene-rich region to be more clinically apparent than a larger deletion involving a more gene-poor region.

This has become an important consideration recently, as it has been reported that most healthy people have many (perhaps as many as 200–250) small deletions and duplications in their DNA molecule. These are collectively referred to as *copy number variations* (*CNVs*), because they cause the individual to have either fewer or more than the normal number of copies of the gene(s) that lie in that region of the chromosome. Because these CNVs appear to be common and widely distributed through the DNA as they are, they may influence the susceptibility to many diseases, and the response to many drugs.

1.2.3 The Polymorphic Human DNA Sequence: Gene Alleles, Protein Isoforms and Genotypes

Because of the way in which our chromosomes are paired, we have two copies of each of the genes that reside on the autosomes. Females also have two copies of each gene that resides on the X chromosome, while males have one copy of each gene that resides on the X chromosome, and one copy of each gene that resides on the Y chromosome.

As we discuss below, the sequence of bases in a typical human gene varies between different individuals, just as a typical physical trait such as height and eye color does. If you sequence the DNA of a group of people, there are many places, or *loci*,[4] at which one can find slightly different sequences in different individuals. In fact, there are many loci in the human genome at which the sequence is so variable that most people possess different versions of the gene's sequence at each of the two copies of that locus that they possess. Any stretch of DNA sequence for which you can observe two or more different versions of the sequence in a population is considered to be *polymorphic* (from the Latin poly = many and morph = form), and is referred to as a *polymorphism.* There are many different types of polymorphisms in the human DNA molecule; we will discuss them in more detail later in this chapter.

Each of the different specific versions of a locus' or gene's sequence that has been observed (in any individual) is referred to as an *allele* of that locus/gene.[5] Because the sequence of bases in the gene's coding sequence determines the sequence of amino acids in the gene's protein, the different alleles of a gene often

[4] The term "locus" is frequently used to refer to a stretch of sequence in the DNA molecule. The term is used very broadly; there is no specific length or sequence requirement for a stretch of DNA to be a locus. A locus can be a single nucleotide, or a string of nucleotides of any length.

[5] There is no length or sequence requirement for a stretch of DNA to be considered an allele, either. Any version of a locus's or gene's sequence, whether it differs from other versions by a single nucleotide or many, is referred to as an allele of that locus/gene.

make different versions of the protein whose amino acid sequences are slightly different from each other. Just as each specific version of a gene's sequence is called an allele of that gene, each specific version of a protein's amino acid sequence that has been seen (in any individual) represents a different *isoform* of that protein. For any polymorphism, the most common allele of the gene, or isoform of the protein, is referred to as the *wild-type* allele/isoform. Other alleles or isoforms are often referred to as *variant alleles* or *variant isoforms*.

Each individual possesses two copies, or alleles, of each gene (except for males having one allele for their X and Y chromosome genes). The two alleles an individual possesses at a locus constitute the individual's *genotype* for that locus. An individual for whom both alleles of a locus/gene have identical base sequences is said to have a *homozygous* genotype for that locus/gene, while someone who possesses different sequences in the two alleles of a locus/gene is said to have a *heterozygous* genotype for that locus/gene. Males are considered *hemizygous* with respect to loci/genes on the X and Y chromosomes, for which they have only one allele. Instead of a genotype, the collection of alleles the male has for the loci/genes on his X or Y chromosome (one allele for each locus/gene) is referred to as a *haplotype*. In addition, recent research suggests that many people have small deletions in their DNA, rendering them *hemizygous* (possessing one copy) for one or more genes. Figure 1.6 illustrates the concepts of chromosomes, gene alleles and genotypes.

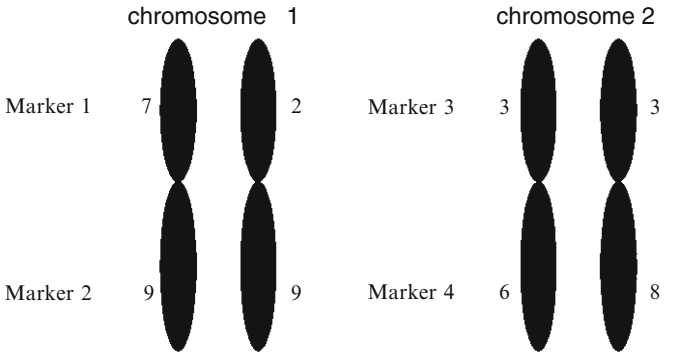

Fig. 1.6 The alleles an individual has for a genetic marker is referred to as the individual's genotype for that marker

The large, whole-genome sequencing projects that have arisen in the last few years are beginning to provide important insights into the variability of the human DNA sequence. The 1,000 Genomes Project aims to obtain the entire DNA sequence from 1,000 volunteers, as well as track their health status over their entire lifetime. The 1,000 Genomes Project Consortium has recently reported the results they observed after obtaining whole-genome sequence data from 179 subjects from four different ethnic populations, as well as exon-targeted sequencing data from 697 subjects from seven different ethnic populations. They reported the presence of approximately 15 million single nucleotide polymorphisms, 1 million short insertions and deletions, and 20,000 structural variants in the human genome, the majority of which had not yet been reported. They also reported that the typical person carries 10,000–11,000 sequence variants that would be expected to alter the amino acid sequence of one of his/her proteins, as well as 10,000–12,000 sequence variants that are not expected to alter the amino acid content of a protein. Because there are many different specific amino acid sequences that will produce a functional isoform of the protein, the variants that change the amino acid content of a protein will not necessarily alter the individual's risk for any particular disease. The Consortium also reported, however, that the typical person also carries approximately 250–300 sequence variations that are predicted to reduce or abolish the activity of a known gene or protein, including 50–100 sequence variants that are already know to cause specific inherited disorders.

1.2.4 Each Sibling Inherits a Unique Combination of Gene Alleles from the Parent

Figure 1.7 illustrates the phenomenon known as *recombination*, which occurs during *spermatogenesis* and *oogenesis.* During the process of spermatogenesis and oogenesis, *spermatogonia* and *oogonia*, which have 46 chromosomes, undergo a process known as *meiosis* to create sperm and eggs, respectively, which have 23 chromosomes. During the process of making the sperm and eggs, the chromosomes are doubled, whereupon each chromosome possesses two *sister chromatids* that are joined at the centromeres. At one point during meiosis, the two members of each chromosome pair line up alongside each other in a *tetrad*, so named because there are a total of four chromatids in the arrangement.

As is illustrated in Fig. 1.7, two of the four chromatids in the tetrad engage in recombination, in which the two chromatids exchange material. This recombines the alleles from some of the genes on that chromosome arm, so that the recombinant chromatids now contain different combinations of gene alleles than were present on the chromosomes the individual inherited from his/her parents. The two other chromatids in the tetrad usually do not undergo recombination, however. This means that, after recombination, there are two chromatids in the tetrad that contain the same combination of alleles that are present in the chromosomes the individual inherited from his/her parents, and two chromatids in the tetrad that contain different

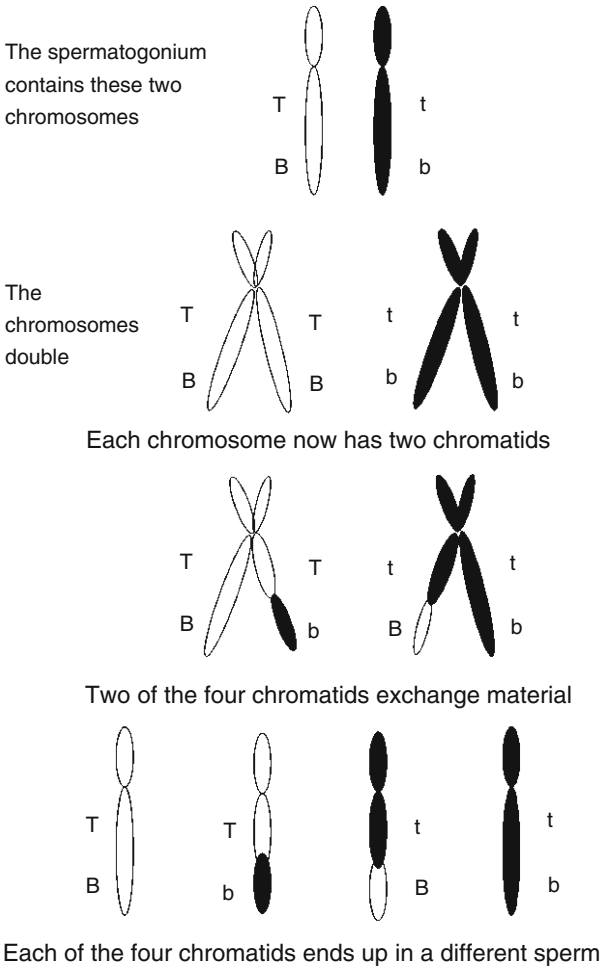

Fig. 1.7 Because of recombination, every gamete an individual makes contains a different set of gene alleles

combinations of alleles from those that are present on the chromosomes the individual inherited from his/her parent.

As the cells continue through the rest of meiosis, the sister chromatids eventually separate, and each of the four chromatids from the tetrad eventually ends up as a chromosome in one of the sperm or eggs that are produced by that meiosis.[6] This means that, for any given chromosome, some of the sperm or eggs the individual

[6]This discussion of spermatogenesis and oogenesis omits some important details, as well as some important differences between the two processes. It is merely intended to illustrate the inheritance of polymorphic gene alleles.

produces contain the same combination of gene alleles as the individual inherited from one of his/her parents (referred to as a *nonrecombinant chromosome*), while other sperm or egg cells contain a chromosome that contains a combination of gene alleles that the individual inherited from his/her mother and gene alleles he/she inherited from his/her father (referred to as a *recombinant chromosome*).

One thing that further increases the diversity of gene allele combinations that are found in the different gametes the individual produces is the fact that the specific point at which the recombining chromatids break, and therefore the amount of material they exchange, varies from meiosis to meiosis (and therefore from one sperm to another or one egg to another). Because the recombination breakpoint on any given chromosome is so variable from one meiosis to another, almost every recombinant chromosome that is created contains its own unique combination of gene alleles that were inherited from the individual's mother and gene alleles that were inherited from the individual's father. Because of the inheritance of recombinant and nonrecombinant chromosomes, and the variability in the recombination breakpoint, each sperm or egg the parent creates contains a different combination of recombinant and nonrecombinant chromosomes, and therefore carries a different set of gene alleles to each child.

1.3 A Review of the Process Whereby a Gene Makes Its Protein

1.3.1 Coding Sequences and Regulatory Sequences

A gene's sequence contains two critical elements: the *coding sequence* and the *regulatory sequences*. The coding sequence is made up of the bases that actually provide instructions as to what protein is to be made. The regulatory sequences are stretches of the gene's sequence where specialized proteins bind to the DNA in order to regulate the level of activity of the gene. The gene's *promoter region* is usually located shortly before the coding sequence, but some regulatory sequences may lie a considerable distance away from the coding sequence.

Because there is a relationship between the gene's coding sequence and a protein's amino acid sequence (discussed in the next few sections), if you know the amino acid sequence of a protein, it is relatively easy to identify the coding sequence that is responsible for producing that protein. It is more difficult, however, to identify the regulatory sequences that control the gene's level of activity. The difficulty in identifying regulatory sequences is an important impediment to the effort to identify all the sources of variability in the activity of any given gene or protein. Variations in these regulatory sequences often alter the level of activity of the gene, and some authorities hypothesize that polymorphisms in regulatory sequences make a stronger contribution to the variability in gene/protein activity than polymorphisms in coding sequences do.

1.3.2 Transcription: Deoxyribonucleic Acid (DNA) Makes Ribonucleic Acid (RNA)[7]

The process whereby the gene's DNA is used to make RNA is called *transcription*. In order to begin the process whereby the gene makes its protein, specialized proteins called *transcription factors* bind to the promoter region of the gene, and recruit the enzyme *RNA polymerase* and other specialized proteins that are needed to transcribe the gene's sequence into RNA. Once the transcription factor proteins have bound to the gene's promoter region, the RNA polymerase reads the gene's DNA base sequence, and chains RNA nucleotides (*ribonucleotides*) together to synthesize an RNA called the *primary transcript*, or the *pre-messenger RNA* (pre-mRNA). Each RNA nucleotide contains the sugar ribose, rather than DNA's deoxyribose, and one of the bases adenine, cytosine, guanine or *uracil* **(U)**. As with DNA, the bases are the most important portion of the nucleotides for our purposes, and we refer to RNA nucleotides as "A," "C," "G" and "U."

The RNA polymerase follows the complementary basepairing rule (A-U and C-G); if it reads a C in the gene's sequence, it incorporates a G into the corresponding locus in the newly synthesized pre-mRNA. If it reads an A in the gene's sequence, it incorporates a U nucleotide into the pre-mRNA. The RNA polymerase reads one of the strands of the DNA (the *template strand*), and makes a pre-mRNA strand whose base sequence is complementary to the DNA strand's sequence. Because the sequence of the RNA is complementary to that of the template strand, and the sequence of the other DNA strand (the *coding strand*) is also complementary to the sequence of the template strand, the sequence of bases in the pre-mRNA is identical to that of the DNA's coding strand, except the pre-mRNA contains U's where the DNA sequence contains T's. This way, if you know the sequence of a gene's coding strand, you can predict the sequence of bases in the gene's pre-mRNA, and ultimately the sequence of amino acids in the gene's protein.

1.3.3 Posttranscriptional RNA Processing

The pre-mRNA is processed into the *messenger RNA (mRNA)*, which is then used to direct the cellular machinery to string the appropriate amino acids together to make the protein. There are several things that happen during posttranscriptional RNA processing, but we will focus solely on the fact that major portions of the pre-mRNA are spliced out, and are therefore not included in the mRNA sequence.

As Fig. 1.8 illustrates, a gene's coding sequence can be divided into *exons* and *introns*. As you can see from the figure, the pre-mRNA contains nucleotides that

[7] This is not an exhaustive description of the transcription or translation process. We merely emphasize those points that have some direct relationship to personalized medicine.

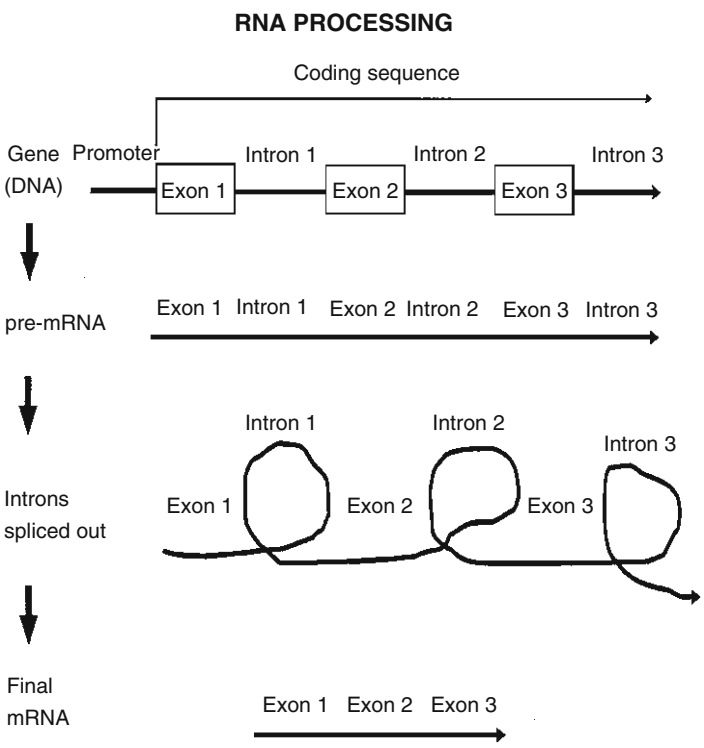

Fig. 1.8 The introns are spliced out during post-transcriptional RNA processing

correspond to the nucleotides from both the gene's exons and its introns. During RNA processing, however, the pre-mRNA forms what is often called a loop and lariat structure. The loops are cut out and the remaining pieces spliced together, so that the portions of the pre-mRNA that were encoded by the introns are lost, but the portions of the pre-mRNA that were encoded by the exons remain in the mRNA.

The cellular machinery recognizes the exon/intron boundaries, because there are certain sequences that signal where the exons end and the introns begin. For example, almost all human introns begin with GT (GU if you are reading the pre-mRNA sequence) and end with AG. In addition, the AG is usually preceded by a stretch of 20 nucleotides, most of which contain C or T bases (C or U in the pre-mRNA). These nucleotides, and some others in the intron as well, inform the cellular machinery which nucleotides are to be spliced out and which are to be kept in the mRNA. If anything goes wrong with the splicing of the pre-mRNA, the mRNA will either contain intronic nucleotides it isn't supposed to contain, or be missing exonic nucleotides it is supposed to contain. These *splice site mutations* usually cause the gene to make a nonfunctional protein, and can therefore contribute to the individual's risk for disease.

1.3.4 Translation of mRNA into a Polypeptide

Transcription and posttranscriptional RNA processing occur in the cell's *nucleus*. Once the gene's mRNA has been processed, it exits the nucleus. An organelle called the *ribosome* reads the base sequence in the mRNA, and uses the sequence of bases in the mRNA as instructions to chain together the proper sequence of amino acids to make the desired protein. The ribosome reads the base sequence of the mRNA three bases at a time; each three-base unit is called a *codon*. Each codon instructs the ribosome to add one specific amino acid to the growing chain. The chain of amino acids that is assembled by the ribosome is called a *polypeptide*. The polypeptide is subjected to further processing (discussed below) in order to become a functional protein.

Transfer RNAs carry amino acids to the ribosome-mRNA complex, so the ribosome can chain the appropriate amino acids together to make the polypeptide (Fig. 1.9). Each tRNA has a 3-nucleotide *anticodon* that binds to the mRNA, and another domain that binds an amino acid. Because the binding of the tRNA with the mRNA follows the complementary basepairing rule, each of the different mRNA

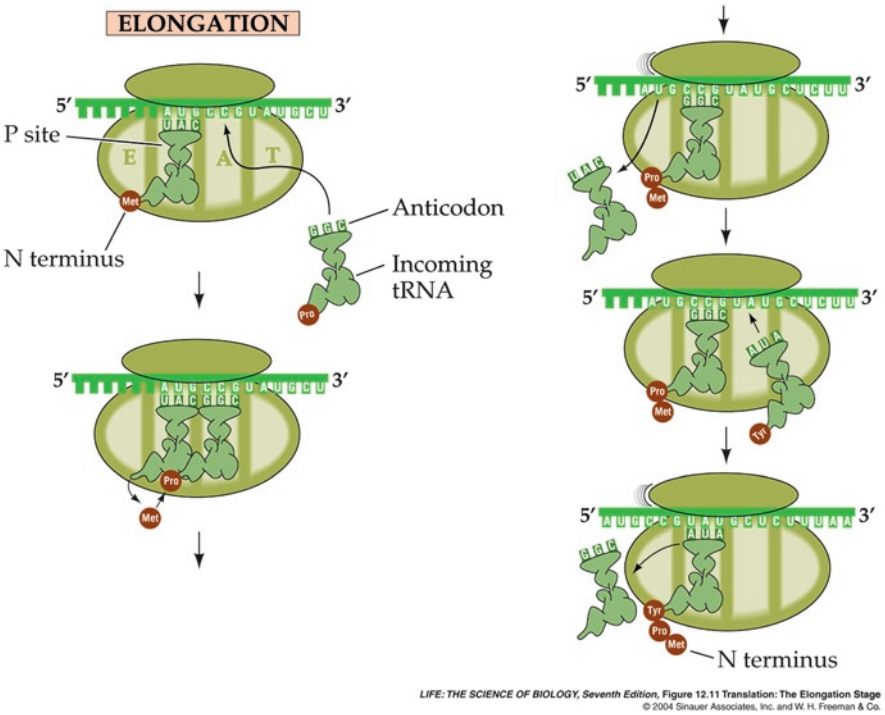

Fig. 1.9 The ribosome reads the mRNA sequence and chains amino acids together (Reprinted from Life: The Science of Biology, 7th ed. Purves, Sadava, Orians and Heller. Copyright Sinauer Associates, 2004)

codons calls for a tRNA with a specific anticodon sequence. Every tRNA with a particular anticodon sequence always carries the same amino acid. Because of this arrangement, if you know the sequence of the mRNA, you can predict the amino acid sequence of the protein.

1.3.5 The Genetic Code and the Structures of Our Amino Acids

The relationship between the mRNA codon sequences and the amino acids they instruct the ribosome to incorporate into the polypeptide is called the *genetic code*. Figure 1.10 illustrates the genetic code. Note that three of the codons (UAA, UAG or UGA) are *translation termination codons*, or *STOP codons*; they signal the ribosome to stop adding amino acids to the polypeptide and release the polypeptide so it can move on to the next stage of processing. AUG is the *translation initiation codon*, or *START codon*; the ribosome begins incorporating amino acids into the chain by reading the START codon (AUG) and incorporating a methionine as the first amino acid in the protein's sequence. There can be several AUG codons in a gene's coding sequence. A special sequence just before and after the START codon identifies that AUG codon as the START codon; all the other AUG codons merely instruct the ribosome to incorporate methionine into the growing amino acid chain.

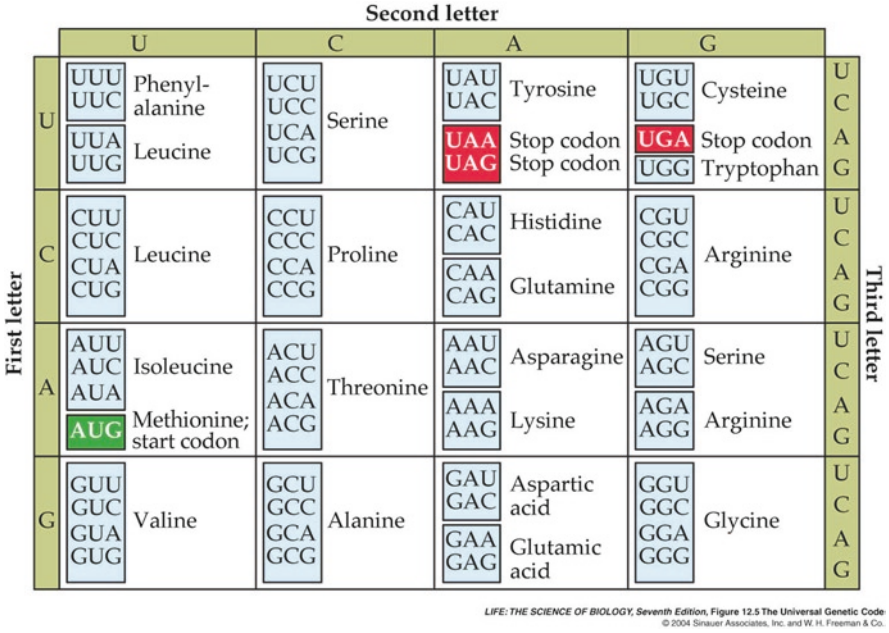

Fig. 1.10 The genetic code illustrates the relationship between the mRNA codon sequence and the amino acid that gets incorporated into the polypeptide (Reprinted from Life: The Science of Biology, 7th ed. Purves, Sadava, Orians and Heller. Copyright Sinauer Associates, 2004)

1.3.6 Posttranslational Processing of the Polypeptide

While some authorities call the ribosome's product a protein, it is more appropriate to call it a *polypeptide* (the bonds between amino acids are called *peptide bonds*), because it must go through several more processing steps before it becomes a functional protein. After the polypeptide is synthesized, it must be folded into its characteristic three-dimensional shape and have certain chemical side groups, such as phosphate groups, amine groups and oligosaccharide chains, attached to specific amino acids.

Posttranslational processing occurs in the *rough endoplasmic reticulum* and the *Golgi apparatus*. In addition, proteins must be transported to their proper locations inside the cell, or secreted outside the cell. Some proteins are produced in an inactive form, and are cleaved by enzymes to produce the functional form of the protein. Others must form complexes with other proteins. Anything that disrupts any portion of this process can result in an impairment of the protein's function.

1.4 A Typical Gene's Sequence and the Level of Activity in the Associated Protein Are as Variable as Any Other Human Trait

The effect that a variation in a gene's sequence can have on the associated protein's function run the entire gamut; some variants will have no effect on the gene's protein, some will cause a small to moderate increase or decrease in the protein's activity, and some (referred to as *null alleles*) will completely abolish the protein's activity. Recall that, like so many other human traits, if you measure the level of activity of a typical human protein in a large population, you would obtain a distribution of values that approximates the normal distribution, or "bell" curve. For the typical human gene, there are many different specific versions of the gene's sequence (gene alleles) present in the population. Most people possess alleles that produce an isoform of the protein that has an approximately average level of activity, but there are always people who possess gene alleles that produce isoforms of the protein that have significantly greater or less activity than the typical-activity isoforms do.

To avoid confusion, we must clarify our terminology.[8] Many authorities limit the term "mutation" to indicate alleles with a frequency less than 1% in the population, and use "polymorphism" to indicate alleles that have a frequency greater than 1%. For our purposes, however, it is more useful and meaningful to classify genetic variants according to their effect on the level of activity in the protein and the individual's susceptibility to disease, rather than their frequency.

[8] We are not proposing this as the standard system of terminology for the field. This is merely a useful system of terminology for the purposes of this book.

In this book we will classify gene sequence variants according to the level of activity in the protein isoforms they produce, and the effect that a change in protein activity has on one's susceptibility to disease. The vast majority of disorders that impact human health are *multifactorial disorders* (also known as *complex disorders*). As the term implies, one's susceptibility to a multifactorial disorder (and also one's probability of an adverse drug reaction, or ADR) is influenced by multiple factors, including both genetic and nongenetic factors. Factors that influence the level of activity of a protein influence the individual's risk for any diseases that involve pathways in which that protein participates.

A gene allele that produces an isoform of the protein that has the typical level of activity will be referred to as a *typical-activity allele*. Because one of these alleles conveys the typical level of risk for the associated diseases, these alleles will also be referred to as *typical-risk alleles*. In addition, we will refer to sequence variants that produce high-activity or low-activity protein isoforms as *high-activity alleles* or *low-activity alleles*, respectively.

Because these high-activity and low-activity gene alleles will alter one's susceptibility to specific diseases, we will also refer to them as *risk-increasing alleles* or *risk-decreasing alleles*, depending on how the change in protein activity influences the individual's susceptibility to a specific disease. Some proteins protect us from disease-causing agents; low-activity alleles in the genes that produce these proteins will be risk-increasing alleles, and high-activity alleles will be risk-decreasing alleles. In contrast, some proteins produce disease-causing agents; high-activity alleles in the genes that produce these proteins will be risk-increasing alleles, and low-activity alleles will be risk-decreasing alleles.

You should note that some authorities refer to gene alleles that increase the individual's risk for a particular disease or adverse drug response (ADR) as "high-risk alleles," and to gene variants that reduce the individual's risk for a particular disorder or ADR as "low-risk alleles." These terms may be misleading, however, because a patient may equate possession of a high-risk allele or a low-risk allele with having a high overall risk or a low overall risk for the associated disorder. This is not necessarily the case. For most multifactorial disorders, each risk-increasing or risk-decreasing allele makes a relatively small contribution to the individual's overall susceptibility to the associated disorder. Possessing one risk-increasing or risk-decreasing allele does not necessarily mean the individual has a high or low overall risk of developing the disorder.

In addition, many authorities use the term *mutation* to indicate any gene allele that contributes to a disease, whether it is a risk-increasing allele that increases the individual's risk for a multifactorial disease, or a mutation that is deleterious enough to be the sole cause of a single-gene disorder (see Sect. 1.3.2). Although we prefer "risk-increasing allele" for those gene alleles that contribute to multifactorial disorders, you will see the term "mutation" used in a number of contexts when you read literature in this field, and we will use it in a few places in this book.

There are several circumstances in which a sequence variation will not have any effect on the activity of any of the individual's proteins. For example, relatively little of our DNA (approx. 2–4%) actually constitutes protein-coding sequence, so

many DNA sequence variants lie outside of the sequences of known genes. In addition, a variant might lie in an intron, and be spliced out during post-transcriptional RNA processing. If that occurs, the variant will not affect the content of the mRNA, so it will not affect the amino acid sequence of the protein. Note, however, that some intronic nucleotides inform the cellular machinery exactly which nucleotides are to be spliced out of the pre-mRNA, and which are to remain in the mRNA. Changes in these nucleotides will disrupt the splicing of the pre-mRNA, alter the content of the mRNA, and disrupt the production of the protein.

Sequence variants that lie within a gene's coding region do not necessarily change the activity level of the gene's protein. Note that the genetic code is partially redundant (Fig. 1.10); there are a number of situations in which substituting one base for another does not change the amino acid that is incorporated into the protein. These *synonymous substitutions* are often assumed to have no effect on the activity of the gene or its protein. It should be noted, however, that some of these synonymous substitutions can alter the pattern of RNA splicing or level of activity of the gene, thereby altering the production of the gene's protein.

For example, the *ABCB1* gene, which encodes a drug transporter that transports several drugs, including phenytoin, across the blood-brain barrier, has a synonymous C>T polymorphism in exon 26. One's status for this polymorphism influences the level of expression of the ABCB1 protein. A high level of this protein correlates with a reduced response to a number of drugs, including phenytoin.

Even when a sequence variant causes a change in the amino acid sequence of the gene's protein, this does not necessarily mean it will change the level of activity in the protein. Figure 1.11 illustrates the structure of some of the amino acids that are found in human proteins. As Fig. 1.11 illustrates, each amino acid has a unique structure, and different amino acids have different sizes, charges and polarities. In order for a protein to perform its function properly, it must

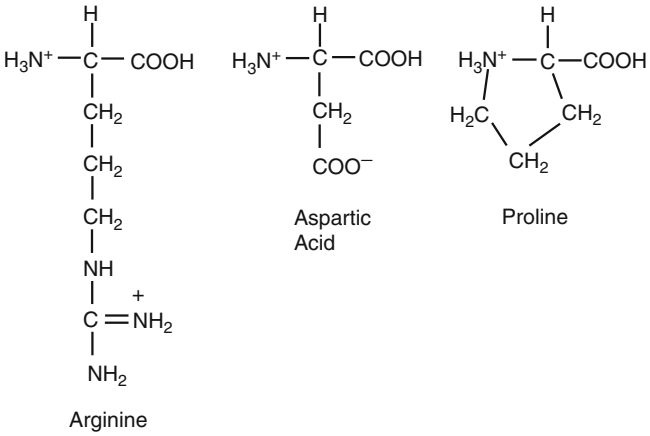

Fig. 1.11 The chemical structures of three of the amino acids that are commonly found in human proteins

adopt its characteristic three-dimensional shape, and many proteins also need to be able to change their shape to some degree as they perform their functions. The three-dimensional shape and degree of flexibility of a protein is determined by steric considerations (the spatial considerations that arise due to the sizes of the atoms in the molecule), as well as by several types of bonding forces that exist between the different amino acids in the protein. The size, charge and polarity of the amino acids determine the types of bonds they can form with other amino acids, as well as the three-dimensional shape and degree of flexibility the protein has.

If one amino acid is substituted for by another amino acid, the effect the amino acid substitution has on the protein's activity depends on how similar the two amino acids are with respect to their size, charge and polarity. For example, consider a polymorphism in which the wild-type allele of the gene's coding sequence has an A, and the variant allele has a G, in that position (symbolized as A>G). Imagine that the wild-type allele has the sequence AAA in codon 7. The seventh amino acid in the wild-type isoforms of this protein, therefore, will be lysine (Fig. 1.10). If the A>G polymorphism occurs in the second nucleotide of the codon, the sequence for codon 7 will now be AGA, and will instruct the ribosome to incorporate arginine into the polypeptide instead of lysine. Substituting lysine in for arginine replaces a positively charged amino acid with another positively charged amino acid of similar size. This substitution will probably not drastically change the activity of the protein, unless the amino acid resides in a domain of the protein that has very stringent requirements for three-dimensional shape (ex. the catalytic site of an enzyme).

In contrast, imagine that the A>G substitution occurs in the first nucleotide of codon 7. This changes the codon's sequence to GAA, and instructs the ribosome to incorporate the negatively charged amino acid glutamate as the seventh amino acid in the polypeptide. Because this substitution replaces a positively charged amino acid with a negatively charged one, this substitution will be tolerated in many fewer locations, and is more likely to cause a greater change in the activity of the protein than the replacement of lysine by arginine is.

One of the core principles underlying personalized medicine is that, because the sequence of a typical human gene is so variable, the activity level of a typical protein will vary from one individual to another, just as physical traits such as height and skin tone do. Minor differences in a gene's sequence between two individuals result in minor changes in the amino acid content of their proteins, which in turn produce small, moderate or large changes in the level of activity in that gene's protein, which in turn alter the individual's level of susceptibility to certain diseases or response to certain drugs. The fact that we each possess our own unique combination of high-activity, typical-activity and low-activity isoforms of our proteins provides the biochemical basis for inter-individual differences in the susceptibility to many common diseases and response to medical treatments, including the probability of experiencing an ADR.

1.5 Risk-Increasing Alleles Have Variable Frequencies and Variable Levels of Penetrance

The specific combination of risk-increasing and risk-decreasing gene alleles the individual possesses is referred to as the individual's *genetic load* for that disorder/ADR. The dietary, environmental and lifestyle factors the individual is exposed to that influence his/her risk for a specific disease are referred to as the individual's *nongenetic load* for that disorder/ADR. The genetic load and the nongenetic load add up to constitute the individual's *aggregate load* for the disorder.

The degree to which any single risk-increasing allele influences the individual's overall risk for the disorder is referred to as the *penetrance* of that allele (see Sect. 3.3). The penetrance of a risk-increasing allele is defined as the percentage of people who possess the risk-increasing allele and also develop the associated disorder. An individual who possesses the risk-increasing allele, but does not develop the disorder, is said to represent a case of *nonpenetrance* of that allele.

There are many reasons for nonpenetrance of a risk-increasing allele. As discussed above, the risk-increasing allele may only change the activity of the gene/protein to a small degree. In addition, the second copy of that gene may be able to compensate for its partner's reduced activity. Further, there are situations in which a decrease in the activity of one protein can be compensated for by the activity of other proteins. The most important example of this involves the genes that encode the CYP450 family of enzymes, which metabolize many commonly prescribed drugs. There are a number of prescription drugs that are capable of serving as substrates for several of the CYP450 enzymes, particularly CYP2D6, CYP2C19 and CYP2C9. The effect of a sequence variant that drastically reduces the level of activity in one CYP450 protein may be diluted by the other CYP450 proteins' abilities to metabolize that protein's target drugs.

There are some risk-increasing alleles that have such high penetrance that possession of that risk-increasing allele singlehandedly endows the individual with a high overall risk for the disease. At the most extreme end of the spectrum, the gene mutations that cause single-gene disorders can be considered risk-increasing alleles that have 100% penetrance. In addition, there are a few highly penetrant alleles that increase one's risk for specific multifactorial disorders that have attracted a great deal of attention among both medical professionals and the lay public (that cause hereditary breast ovarian cancer: see section...). For example, several of the risk-increasing alleles in the *BRCA1* and *BRCA2* genes have up to 85% penetrance. A woman who possesses one of these risk-increasing alleles has an 85% risk of developing breast cancer in her lifetime. Because of their high penetrance, genetic tests that determine the individual's status for these gene variants are clinically useful tests, because they enable one to accurately estimate the risk for breast cancer in the different members of a family.

These highly penetrant alleles are relatively rare, however. The vast majority of gene variants that influence an individual's risk for the multifactorial

disorders are more common than these highly penetrant alleles, but only make small contributions to the individual's risk for the disorder. For this reason, in order to truly appreciate the influence genes have on a particular metabolic pathway, one must know the individual's genotype for dozens, hundreds or perhaps even thousands of polymorphisms from genes whose proteins mediate different aspects of the pathway. This is one of the primary benefits of microarray technologies (see Sect. 4.19.8); the arrays allow you to examine hundreds of thousands of individual DNA or RNA sequence variants at once, enabling the researcher to study all the known genetic variants or gene activities in an entire metabolic pathway.

There may be more than one risk-increasing or risk-reducing allele for any particular gene. In addition, although this is rare, the same gene allele can be a risk-increasing allele for one disorder, and a risk-reducing allele for another disorder. The best-known example of this is the common sickle cell anemia mutation, which involves a single amino acid substitution in the beta-globin protein, which is a subunit of hemoglobin. Possessing one copy of this beta-globin gene variant increases the individual's risk for sickle-cell anemia, but also provides the individual some protection from a lethal strain of malaria. Another example can be seen in the interleukin-1-receptor agonist (IL-1RA) gene, which encodes a protein that inhibits the inflammatory response. There is an allele of the IL-1RA gene (referred to as IL1RN*2, discussed below) that reduces the expression of the IL-1RA protein, and increases the individual's risk for ulcerative colitis, Crohn's disease, systemic lupus erythematosus and vulvar vestibulitis, and possibly for osteoporosis, coronary artery disease, recurrent spontaneous abortion, preterm birth and severity of preeclampsia. This allele is also associated, however, with an increased resistance to several types of infectious agents, including vaginal mycoplasmas, human cytomegalovirus, Epstein-Barr virus and human immunodeficiency virus. In addition, it may protect against ovarian cancer.

1.6 Polymorphisms in Promoter Regions and Other Non-coding Sequences Influence the Activity of Our Proteins

1.6.1 Polymorphisms in Promoter Regions Alter the Gene's Level of Activity

As described above, in order to begin the process whereby the gene makes its protein, specialized proteins called transcription factors bind to the promoter region of the gene, and recruit the RNA polymerase and other specialized proteins that are needed to transcribe the gene's sequence into RNA. A change in the sequence of a promoter region can change the efficiency with which the transcription factors initiate transcription, and thereby change the activity of the gene.

For example, some people have a 44 bp stretch of DNA in the promoter region of the serotonin transporter (*5HTT*) gene,[9] while other people lack those 44 bp (referred to as the *5HTTLPR* polymorphism). This is an example of an insertion/deletion polymorphism, as discussed below. The insertion allele (I allele) of this *5HTTLPR* polymorphism, which contains the 44 bp, conveys a higher level of activity on the gene than the deletion allele (D allele) does. The I allele of the gene makes more mRNA than the D allele does, and animal studies report that synaptosomes from subjects who are homozygous for the I allele (i.e. have the II genotype for the *5HTTLPR* polymorphism) show more 5HTT activity than synaptosomes from individuals who are heterozygous (i.e. who have the ID genotype) at this locus. The *5HTTLPR* polymorphism has been reported to influence the individual's risk for several psychiatric diseases and response to a number of psychiatric drugs (see Sect. 8.3.8), although several of these initial findings must be confirmed before they can used to develop clinically useful genetic tests.

The promoter region of a gene can contain several different individual promoter sequences, each capable of responding to different transcription factor proteins. This enables different tissues to turn the gene on and off at different times or in response to different events. This promoter specificity can also result in mutations in the same gene producing markedly different clinical presentations. For example, consider that different mutations in the dystrophin gene, which is located on the X chromosome, can result in either Duchenne muscular dystrophy (DMD), Becker muscular dystrophy (BMD) or X-linked dilated cardiomyopathy (XLDC).

DMD and BMD both result from mutations that change the amino acid content of the dystrophin protein. The primary clinical presentation involves the dysfunction and degeneration of the skeletal muscles, with involvement of the cardiac muscle as well. DMD and BMD are the same disorder, differing only in severity and clinical course; mutations that only mildly impair dystrophin's ability to function cause the milder BMD, while mutations that severely impair dystrophin's function cause DMD.

Because the DMD and BMD mutations change the amino acid sequence of the dystrophin protein, the protein is deficient in every tissue in which it is expressed, and both skeletal and cardiac muscle are involved. Mutations in a certain portion of the promoter region, however, cause the selective loss of dystrophin expression in cardiac muscle, while preserving normal or nearly normal levels of dystrophin in the skeletal muscles. Because the mutation is in a sequence that controls the activity of the gene, but not the amino acid content of the gene's protein, the gene is capable of making a functional protein when it is active. It is most likely that the mutations that cause XLDC do so by altering a promoter sequence that is a target of the transcription factors that are present in cardiac muscle cells, but preserves intact the promoter sequences that respond to the transcription factors that are present in skeletal muscle cells. As a result, the cardiac muscle cells do not produce sufficient dystrophin to support normal function, but the skeletal muscle cells are unaffected.

[9] Serotonin's chemical name is 5-hydroxytryptamine, or 5HT. Therefore the 5HT transporter gene is referred to as *5HTT*.

1.6.2 Chromosome Rearrangements Can Cause Promoters to Drive Transcription of the Wrong Sequences

Sometimes when chromosomes break and rearrange themselves, as in inversions and translocations,[10] the rearrangement results in a promoter sequence being moved to a new locus on the chromosome, and being followed by a different sequence than it is followed by in its normal position. The cellular machinery may recognize that promoter sequence, and make a protein using the nucleotides that follow that promoter sequence as a source of instructions. If that protein turns out to be functional, it can disrupt metabolism in dangerous ways.

For example, in 95% of individuals with chronic myelogenous leukemia, a translocation involving chromosomes 9 and 22 [formal nomenclature = t(9;22)(q34;q11), often referred to as the Philadelphia chromosome] results in a *fusion gene*, in which the promoter region of the *BCR* gene is followed by a portion of the *BCR* gene's coding sequence, then by a portion of the *ABL* gene's coding sequence. The BCR-ABL[11] *fusion protein* has biological activity; it speeds up the rate of cell division and impairs the process of DNA repair, causing excessive proliferation of neutrophils, eosinophils and basophils.

Sometimes the presence of a fusion gene may provide some insights into the pathogenesis of the disease, and by extension the prospects for treating the disease. For example, in some individuals with idiopathic hypereosinophilic syndrome, a deletion in chromosome 4q12 causes the fusion of the Fip1-like 1 (FIP1L1) gene and the platelet-derived growth factor receptor alpha (*PDGFRA*) gene. The fusion protein FIP1L1-PDGFRA contains the PDGFRA protein's tyrosine kinase domain, and its transcription is under the control of the FIP1L1 promoter. The FIP1L1 promoter appears to be a constitutive, or 'housekeeping gene' promoter; the gene under its control is active in most cell types at most times during the lifespan. The constitutively activated FIP1L1-PDGFRA protein transforms hematopoietic cells, causing the excessive production of eosinophils. Like other tyrosine kinases, the fusion protein is inhibited by the drug imatinib. Individuals with idiopathic hypereosinophilic syndrome who have the FIP1L1-PDGFRA fusion protein are more likely to benefit from imatinib treatment than individuals who lack the fusion protein are.

---

[10] The rearrangements we speak of here are not the exchange of material between recombining chromatids in meiosis. These rearrangements involve breakage of the entire chromosome, and either reinstatement of the excerpted piece in the opposite orientation, or exchange of material between different chromosomes, not between chromatids from the same chromosome.

[11] When perusing published works, you will often find the gene's name written in italics, and the name of the protein written in regular font.

1.6.3 Intronic Gene Variants That Influence RNA Splicing or Gene Activity May Be Unrecognized Risk Factors

Because the introns are spliced out of the pre-mRNA before the mRNA provides the ribosome its instructions for chaining the appropriate amino acids together, intronic polymorphisms that do not involve the nucleotides that are known to be critical for proper RNA splicing are often assumed to have no effect on the activity of the gene or its protein. We do not know all the means by which intronic nucleotides control splicing, however, and there are undoubtedly numerous intronic nucleotides that influence RNA splicing in ways we do not yet understand. Variations in these nucleotides may be reported as benign variants, or variants of uncertain significance. One must always keep in mind, however, that they may change the pattern of RNA splicing, and drastically disrupt the production of the gene's protein.

Some intronic sequence variants can alter the level of expression of the protein. For example, intron 2 of the interleukin-1-receptor antagonist (*IL-1RA*) gene contains a 344 bp stretch of DNA in which an 86 bp sequence is repeated four times. In some people, however, there are only two repetitions of the 86 bp sequence motif; this allele is referred to as the IL-1RN*2 allele. Although the effect of this variant may depend on the specific cell type, most research suggests that having the IL-1RN*2 allele results in decreased production of the IL-1RA protein. The IL-1RN*2 allele is a risk-increasing allele for a number of disorders that primarily involve epithelial and endothelial tissues.

1.6.4 Intronic Polymorphisms Can Influence the Ratio of Protein Isoforms or the Balance of Allelic Expression

The process whereby the pre-mRNA is spliced to make the mRNA is actually an important means by which the activity of genes is regulated. Most genomic researchers estimate that there are approximately 21,000–23,000 genes in the human genome, but approximately 90,000–100,000 distinct proteins in the human proteome (our collection of proteins). This means that many, if not all, our genes make multiple isoforms of their respective protein, not only when caused to do so by a functional polymorphism in the gene's sequence, but also as part of the normal process whereby our genes make our proteins. Many genes engage in *alternative splicing* when they process their pre-mRNA into mRNA. These genes make different isoforms of the protein by splicing some exons out of some isoforms along with the introns, keeping a different combination of exons' nucleotides in each of the mRNAs that produce each of the protein's different isoforms.

Sequence variants can sometimes change the ratio of the different isoforms the gene produces, which may have functional consequences. For example, there is an alternative exon (exon 5 N) of the gene that encodes the alpha subunit of the type 1 voltage-gated sodium ion channel (*SCN1A*) that is present in the SCN1A isoform

that is expressed primarily in fetal tissues. A G > A polymorphism in intron 5 of the *SCN1A* gene affects the proportion of *SCN1A* mRNAs that include versus exclude exon 5 N. In patients with epilepsy, seizures can increase the proportion of the *SCN1A* mRNAs that contain exon 5 N. One study has reported that this effect is most pronounced in individuals who have the GG genotype for this intronic *SCN1A* polymorphism.

Another related issue involves *allelic imbalance*. It is often assumed that the two alleles one has of a gene are expressed at approximately equal levels, and contribute approximately equally to the individual's supply of that protein. Many genes demonstrate an imbalance of allelic expression, however, wherein one allele expresses itself at a significantly higher level than the other allele does. If the two alleles encode isoforms of the protein with different levels of activity, this can not only influence the level of activity in that protein, but it can also confound the interpretation of an analysis that reports the individual's allele status for that gene. One study has reported finding an allelic expression imbalance in approximately 55% of the genes studied, including several that have been reported to influence the response to several prescription drugs.

1.6.5 Interfering RNAs Exert an Important Influence over Gene Activity

One interesting mechanism whereby protein production is regulated involves *interfering RNAs*. *MicroRNAs* (*miRNAs*) in particular are a class of interfering RNAs that are receiving considerable attention from medical researchers. Some have suggested there may be more than 400 distinct miRNAs in a typical human cell. Many miRNAs are produced by genes that lie in stretches of sequence that have historically been considered untranscribed stretches of sequence between genes, but some miRNA genes lie in the introns, and occasionally exons, of known genes.

The miRNA genes produce RNAs that have several regions in which the base sequence is complementary to other regions of that same RNA. This causes the RNA to form hairpin loops and other double-stranded configurations. The enzyme dicer cleaves these RNAs to yield fragments of double-stranded RNA ~ 20–25 bp in length. The double strand is unwound, whereupon one of the two strands (the passenger strand) is degraded. The other strand (the guide strand) joins the RNA-induced silencing complex (RISC), and binds the RISC to the complementary sequence in its target mRNA. This causes cleavage of the mRNA by the RISC enzyme argonaute, thereby preventing translation of the mRNA by the ribosome.

There are two ways in which interfering RNAs can contribute to personalized medicine. One application involves testing polymorphic miRNA genes for variants that influence the risk for or progression of certain diseases. For example, miRNAs may play an important role in carcinogenesis by inhibiting tumor suppressor genes, thus initiating the cancer process. Variants in the sequence of miRNA genes may therefore influence the risk for or progression of certain cancers.

The second type of application involves the development of chemically synthesized *small interfering RNAs* (*siRNAs*) as possible treatments for a number of diverse disorders. These siRNAs can be synthesized in the laboratory and introduced into a cell by transfection, whereupon they will be processed by dicer, joined with the RISC and able to reduce the production of their target genes' proteins. Because the sequences of most human RNAs is well known, and siRNAs can be synthesized that contain any possible sequence of bases, almost any gene can be a target for siRNA therapy. siRNAs are being actively investigated as possible treatments for metastatic melanoma, macular degeneration and some solid tumors, and recent basic research findings suggest that they may be useful for combating viral infections and neurodegenerative diseases.

1.7 Epigenetic Factors Also Control Gene Activity

If a gene is to play its proper role in maintaining our health, the level of activity of the gene must also be regulated appropriately. A gene's level of activity is controlled in part by chemical modification of the DNA and/or the histone proteins around which the DNA is wound. *Epigenetic*s is a term that has been coined to refer to these regulatory mechanisms; epigenetics is formally defined as stable changes in a gene's level of activity that stem from mechanisms other than changes in the base sequence of the gene.

The primary epigenetic modification that is observed in the DNA involves the methylation of C nucleotides in the gene's promoter region, especially C nucleotides that are followed by G nucleotides. The histone proteins around which the DNA is wound also undergo several different chemical modifications, including methylation, acetylation, phosphorylation and ubiquitination. Relatively little is known about the role most of these histone modifications play in the process, but the interplay between DNA methylation and histone protein acetylation has been the major focus of research on this aspect of gene regulation.

As a general rule, transcriptionally active DNA is often unmethylated, and the histone proteins in that region are acetylated. In regions or gene alleles in which transcription is repressed, the DNA is often methylated and the histones are deacetylated. This is the mechanism by which many *imprinted genes* are silenced. An imprinted gene is a gene for which only one allele is supposed to be active. For some imprinted genes, the copy that was inherited from the mother is active (MAT-ON/PAT-OFF), while for other imprinted genes, the copy that was inherited from the father is active (PAT-ON/MAT-OFF). In either case, for many imprinted genes some of the C's in one allele's promoter region are methylated, and the histones in that region of the chromatin are deacetylated.

The Prader-Willi syndrome (PWS) and Angelman syndrome (AS) illustrate examples of syndromes that are caused by abnormal regulation of imprinted genes. PWS is characterized by severe hypotonia, poor suck and feeding problems in early infancy followed later in infancy by excessive eating. If unchecked, this excessive

eating can gradually lead to significant obesity. Individuals with PWS have mild-to-moderate mental retardation with multiple learning disabilities. In contrast, AS is characterized by microcephaly, severe motor and intellectual retardation, ataxia, frequent jerky limb movements and flapping of the arms and hands, seizures, absence of speech, and unusual facies.

There are several imprinted genes clustered in a small region of chromosome 15. PWS results when the normally active allele of a PAT-ON/MAT-OFF gene is silenced, while AS results when the normally active allele of a MAT-ON/PAT-OFF gene is silenced. Aberrant silencing of the (PAT-ON/MAT-OFF) gene that encodes the small nuclear ribonucleoprotein-associated protein N (SNRPN) is thought to be primarily responsible for PWS. Aberrant silencing of the (MAT-ON/PAT-OFF) gene that encodes the ubiquitin protein ligase E3A (UBE3A) is thought to be primarily responsible for AS.

These epigenetic mechanisms control gene activity on both the individual gene scale and the regional scale. Recall from above that the DNA molecule, plus the histone and other proteins that hold it in its supercoiled configuration, are collectively referred to as chromatin. Human chromatin exists in two primary configurations. *Euchromatin* is the more relaxed, less compacted form. In euchromatin, the DNA is less methylated, the histone proteins are acetylated, and most of the genes in the region express their proteins. In *heterochromatin*, on the other hand, the DNA is heavily methylated, the histones are deacetylated and the chromatin is highly compacted. Further, most of the genes that reside in the heterochromatic regions are silenced. The centromeres of all human chromosomes consist of large blocks of heterochromatin. Not only are all the genes in these regions silenced, but if a chromosome rearrangement moves a normally active gene too close to the centromere, being in close proximity to the heterochromatic centromere can silence the normally active gene.

Epigenetic factors can influence an individual's risk for a disease or response to a drug by affecting the level of activity in proteins that influence the pathophysiology of the disease, the drug's metabolism or the drug's effects on its targets. For example, alkylating agents battle tumors by damaging the DNA in the rapidly dividing malignant cells. Unfortunately, through its normal function, the DNA-repair enzyme O6-methylguanine-DNA methyltransferase (MGMT) opposes attempts by alkylating agents to kill tumor cells. The level of activity in the *MGMT* gene is determined by the methylation status of the gene's promoter; methylation of the promoter silences that particular allele of the gene. Methylation of the *MGMT* gene promoter has been associated with a more favorable response of gliomas to the alkylating agent carmustine. Methylation of the *MGMT* gene also was associated with more extensive regression of the tumor and better overall survival as well as disease-free survival.

There is one very interesting difference between these epigenetic mechanisms of gene regulation versus sequence-based mechanisms of gene regulation that makes epigenetic factors a promising field of research. While the base sequence of a gene is expected not to change over the individual's lifetime, these epigenetic chemical modifications are all reversible. In fact, it has been demonstrated that several drugs and foods can affect the level of DNA methylation of one or more genes, and thereby

change their level of activity. This means that epigenetic factors can contribute to the development of a multifactorial disorder if the level of DNA methylation (and the level of activity of one or more critical genes) changes during the individual's lifetime. Epigenetic changes in gene activity have been demonstrated to occur in the malignant cells in several types of cancers, and to affect genes involved in every aspect of the growth, differentiation and metastasis of cancer cells.

For example, consuming high levels of green tea has been reported to decrease methylation of the gene that encodes the caudal type homeo box transcription factor 2 (CDX2). The CDX2 protein is critical for differentiation of the intestinal epithelial cells, and the loss of CDX2 activity may increase one's risk for colon cancer. Green tea may reduce the risk of colon cancer by decreasing the level of methylation of *CDX2*, thereby maintaining a high level of *CDX2* activity.

Because these epigenetic factors can be modified, they represent potentially fruitful avenues for drug development research. Several drugs have been designed that are capable of inhibiting the activity of the DNA methyltransferases, histone acetyltransferase and histone deacetylase enzymes that catalyze the reactions that bring these epigenetic modifications about. Drugs that modify the level of DNA methylation are being studied as potential therapeutic agents for a number of disorders, including cancer, some of the autism spectrum disorders, Angelman syndrome and Fragile X syndrome.

1.8 Common Types of Variants in the Human DNA Sequence

There are several different types of variants in the human DNA sequence. Among the most common are:

A *single nucleotide polymorphism* (*SNP*) involves a single nucleotide; where one person has an A, another person may have a C, G or T. Approximately 15 million SNPs have been characterized in the human genome as of this writing. It is easy to see from the genetic code (Fig. 1.10) how some SNPs alter the amino acid content of the gene's protein, but others do not. If the SNP causes one amino acid to be substituted for another one of similar size, polarity and electrical charge, the SNP may constitute a benign polymorphism. If the new amino acid differs from the one it replaced, however, the SNP may alter the activity of the protein, and thereby alter the individual's risk for a disorder or ADR.

An *insertion/deletion (ins/del or indel) polymorphism* involves a situation in which some people have a certain stretch of nucleotides in their DNA, while other people do not. Genomic researchers estimate that there are several hundred thousand different ins/del polymorphisms in the human genome. The 44 bp serotonin transporter gene polymorphism discussed earlier in this chapter provides an example of an ins/del polymorphism that has functional significance. There is also a 287 bp ins/del polymorphism in the angiotensin converting enzyme (*ACE*) gene that has functional significance; people who are homozygous for the deletion allele have higher plasma ACE levels than people who are homozygous for the insertion allele.

The human DNA sequence also includes a surprising number of *repetitive sequences*. Depending on the specific locus, the repeated sequence may range in size from a mononucleotide repeat (ex. CCCCCCCC) or dinucleotide repeat (ex. CACACACA) to thousands of repetitions of a 1,000 bp motif. Short tandem repeats (STRs), especially tetranucleotide repeats (ex. GAAT), are the cornerstone of forensic DNA testing. Trinucleotide repeats, on the other hand, have a special significance in the field of medicine. Unlike other types of short tandem repeats such as dinucleotide or tetranucleotide repeats, trinucleotide repeats are unstable, and often expand in size during meiosis. A parent with 50 repetitions of a trinucleotide repeat may produce gametes with 250 repetitions of the trinucleotide motif at that locus, which may expand to thousands of repetitions in the next generation. Expanding trinucleotide repeats are responsible for several disorders, including Fragile X syndrome and several degenerative neuromuscular disorders. Several of these disorders exhibit *anticipation*, in which the expansion of the repeat through the generations results in the members of each succeeding generation being affected at earlier ages and more severely than the members of the generation before.

As we explore the human genome further, it is becoming apparent that there are many repeat length polymorphisms that influence the activity of the gene in which they reside. The mechanisms whereby these variants influence protein function often provide insights into the means by which gene activity is regulated. Many of these repeated sequences lie within an intron, or within the promoter region of their respective gene. Instead of altering the amino acid sequence of the gene's protein, many of these polymorphisms influence the level of activity of the gene or its protein.

For example, some genes have "TATA box" promoter sequences, which are stretches of sequence at the 5′ end of a gene that are rich in T and A bases, including sequences such as TATA or TATAAA. The uridine diphosphoglucuronosyl transferase 1A1 (*UGT1A1*) gene is one such gene. Most people have the sequence $A(TA)_6TAA$ at a particular place in the *UGT1A1* promoter (i.e. they have six repetitions of the TA repeat), but some people have an extra TA in the TA repeat, changing the sequence to $A(TA)_7TAA$. It is already known that the G71R and Y486D mutations in the coding sequence of *UGT1A1* can cause Gilbert's syndrome, which is characterized by a mild unconjugated hyperbilirubinemia with no structural liver disease or overt hemolysis. Gilbert's syndrome can also be seen in individuals who are homozygous for the 7-repeat TA allele of the *UGT1A1* promoter polymorphism.

The 7-repeat allele of the UGT 1A1 promoter polymorphism also increases the individual's risk for an adverse reaction to irinotecan (CPT-11) administration. In one study, possession of the 7-repeat allele increased the frequency of toxic reactions to the drug; 71% of the individuals with the 7/7 genotype experienced grade 3 or 4 toxicity after irinotecan administration.

Although most of the repeat length polymorphisms that are known to contribute to multifactorial disorders involve small repeated sequences, several known functional polymorphisms involve larger repeated sequences. For example, the *IL-1RA* polymorphism mentioned above involves an 86 bp repeat. The allele with two repetitions of the 86 bp sequence (the *IL1RN*2* allele) is associated with a poor prognosis in individuals with several chronic inflammatory diseases, including systemic

lupus erythematosus, ulcerative colitis and alopecia areata. It has also been reported to be associated with diabetic nephropathy.

One of the more surprising recent discoveries is the fact that there is a lot of *copy number variation* (*CNV*) in human DNA as well. While it is accepted dogma to assume that each individual has two copies of each gene (except for the fact that males have only one copy of all X and Y chromosome loci/genes), a surprising number of people have duplications and deletions that result in them having more or fewer than the assumed two copies of certain genes. Many of these duplicated/deleted sequences are hundreds of thousands of kilobases in length; many contain an extra copy or copies of at least one gene. If these extra copies are functional, as many of them are, the individual will produce more of that particular protein than the typical person does. Deletions, on the other hand, result in the individual having only one copy (hemizygous) or no copies of the gene. The phenotypic consequences of these CNVs are determined by several factors, including whether there is another protein capable of serving the necessary function if the protein in question is reduced or absent, or how easily the cell can adapt to an excess of the protein in question.

1.9 Common Multifactorial Diseases Are Genetic Disorders, Despite Their Non-Mendelian Patterns of Inheritance

Until recently, the medical community's understanding of the means by which our genes influence our health has been dominated by single gene disorders and their classic Mendelian patterns of inheritance. These single-gene disorders often represent dramatic alterations in the individual's health: they often affect the individual from birth or shortly thereafter, and they often have severely deleterious effects on the individual's development and function throughout his/her entire lifetime.

As we learn more about the role our genes play in maintaining our health, however, the definition of "genetic diseases" is expanding. We have come to recognize that the more common, multifactorial disorders that affect so many individuals (ex. cancer, cardiovascular disease, diabetes, Alzheimer disease, respiratory disorders) are also genetic diseases, because genetic factors play an important role in determining the individual's susceptibility to these disorders. These multifactorial disorders are far more common than the single-gene disorders, and together they represent a huge burden on people's lives. In addition, while these multifactorial disorders often arise later in life than the single-gene disorders do, they often reduce the quality of a significant portion of the individual's life.

We need to recognize that these common multifactorial diseases are genetic diseases, despite the fact that they do not produce the Mendelian patterns of inheritance we have come to equate with "genetic diseases." It is considerably harder to characterize the effects that particular genes have on the individual's susceptibility to multifactorial disorders than it is for the single-gene disorders. Unlike the single-gene disorders, each individual risk-increasing allele only makes a small contribution to the individual's overall risk for a complex disorder. In addition, it is harder

to characterize the phenotype that is associated with one of these multifactorial disorders, because the expression of these disorders is highly variable. The age of onset of many of these multifactorial diseases is highly variable. In addition, the expression of symptoms is highly variable in multifactorial disorders, and their progression is often more gradual than that of single-gene disorders.

Despite the extra challenges, the medical genetics community's attention is clearly shifting from the rare single-gene disorders to the much more common, complex multifactorial disorders. We have become aware that far more people's health is affected by the relatively common gene polymorphisms that contribute to common multifactorial disorders than by the relatively rare gene mutations that cause single-gene disorders. We have also become aware that, despite the extra challenges presented by multifactorial disorders, this is a field in which our expanding knowledge of genetics and genomics can be applied most effectively to reduce human suffering.

Another way in which multifactorial disorders present an extra challenge to the medical community is the difficulty physicians and genetic counselors will have communicating this complex information to their patients. With multifactorial disorders, determinations of risk and susceptibility are more complicated, and physicians, genetic counselors and other healthcare workers alike will need more guidance on how to explain this information to their patients within the framework of our lack of complete understanding of the disease process. These challenges, however, may be offset by the growing understanding and promise that multifactorial diseases should also be more amenable to early intervention and treatment, and to personalized approaches to medicine. In addition to using personalized medicine tests to improve the way doctors treat patients, the practitioner may also have the opportunity to improve the course of the patient's disease by making the patient aware of the need to modify potential contributing nongenetic factors, such as diet, exercise, medication and other behavior changes.

1.10 Personalized Medicine Testing May Allow You to Better Tailor the Treatment to the Individual, and May Allow the Individual to Make Healthier Choices

At present, most health care is disease-oriented. It is often true that every individual who has a particular disease is prescribed the same first-line treatment. If there is a choice of several medications for the disorder, the physician often has no means to predict which one will be safest and most effective for that individual, and therefore must use a trial-and-error process to determine which drug, or which dose of the drug, is safest and most effective for that individual. As we learn more about the way in which our genes and proteins influence our risk for disorders and responses to treatments, however, we are developing the ability to perform genetic tests that can identify the individuals who are most likely to benefit from, or to suffer adverse

side effects from, a particular treatment. Genetic tests can help the doctor choose which of several medications to prescribe, and guide dosing strategies for that individual. Genetic tests can also help the nutritionist/dietician design a diet that will provide maximum benefit for that individual.

Perhaps most importantly, genomic medicine can identify the diseases for which the individual is at greatest risk, and thereby help the individual maintain his/her own health as effectively as possible. The lay public's health literature contains a cornucopia of advice on diet, environment and lifestyle that is designed to improve people's health. Many of these articles contain valid information, and many people may very well reduce their risk for common diseases if they followed the right advice. Most people do not need to follow most of the advice that is out there, however. Most people only need to follow the advice that reduces their risk for the diseases for which they have an increased genetic risk.

Unfortunately, without knowing the individual's genotype for the critical genes, one cannot determine the diseases for which the individual is at increased risk for, and which advice is most important for that individual. No one can follow a regimen of diet and lifestyle that provides them optimal protection from every disease that plagues human health. As genetic/genomic research continues to generate new information regarding the way in which our proteins influence our health, new genetic tests will be developed that will identify the diseases for which the individual has an increased genetic risk. This information will enable many people to tailor their diet, environment and lifestyle to reduce their risk for those disorders to which they are most susceptible. Future research will determine the degree to which people comprehend, respond to and incorporate this information to adopt healthy behaviors and lifestyles.

1.11 Summary

- At the present time, most of the emphasis is on genetic tests, but in time the field will expand to include assessments of RNAs, proteins and the interactions between proteins.
- The human DNA sequence is highly polymorphic. For the typical human gene, there are many different alleles of the sequence in the population, and some of the alleles cause the gene to make an isoform of the protein that has significantly higher or lower level of activity than other isoforms of that protein have.
- Depending on the specific function of the protein, the alleles that produce higher or lower levels of activity in the protein can be risk-increasing or risk-decreasing alleles.
- Different risk-increasing alleles have different levels of penetrance. There are a few diseases for which a few highly penetrant risk-increasing gene alleles have been identified, but for most diseases, there are many risk-increasing alleles with low penetrance that contribute to the disease.

- Polymorphisms in promoter regions and other regulatory sequences represent important influences over gene activity, and may provide numerous clinically useful tests in the future.
- Polymorphisms in intronic sequences and microRNA genes also represent potential subjects for clinically useful tests.
- Epigenetic factors, such as the methylation of DNA, are important regulators of gene activity, and may provide clinically useful tests in the future.
- Common variants in the human DNA sequence include not only changes in gene sequences but also changes in the number of copies of a gene the individual possesses.
- Personalized medicine tests may not only enable you to tailor treatments to the individual patient, they may also enable you to identify the diseases the individual has the highest genetic risk for, and thereby provide the most effective advice regarding the patient's diet, environment and lifestyle.

Further Readings

Auffray C, Chen Z, Hood L (2009) Systems medicine: the future of medical genomics and healthcare. Genome Medicine 1(1): 2.

Bartel, DP (2009) MicroRNAs: target recognition and regulatory functions. Cell 136(2): 215–33.

Cuccato G, Della Gatta G, di Bernardo D (2009) Systems and Synthetic biology: tackling genetic networks and complex diseases. Heredity 102(6): 527–32.

Izquierdo M (2005) Short interfering RNAs as a tool for cancer gene therapy. Cancer Gene Ther 12 (3): 217–27.

Mattick JS and Gagen MJ (2001) The Evolution of Controlled Multitasked Gene Networks: The Role of Introns and Other Noncoding RNAs in the Development of Complex Organisms. Molecular Biology and Evolution 18(9): 1611–1630.

Schneider JA, Pungliya MS, Choi JY, Jiang R, Sun XJ, Salisbury BA, Stephens JC (2003) DNA variability of human genes. Mechanisms of Ageing and Development 124(1): 17–25.

Simmons D (2008) Epigenetic Influences and Disease. www.nature.com/scitable/topicpage/epigenetic-influences-and-disease-895.

The 1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. Nature 467: 1061–1073.

Zhang F, Gu W, Hurles ME, and Lupski JR (2009) Copy Number Variation in Human Health, Disease, and Evolution. Annual Review of Genomics and Human Genetics Vol. 10: 451–481.

Chapter 2
Making the Most of Family History Information, Single Gene Disorders and Mendelian Patterns of Inheritance, and When to Refer to a Genetic Specialist

Abstract Family history is an essential tool for disease risk assessment, because it incorporates not only shared genetic susceptibilities but also shared environmental, behavioral and cultural factors. Family history data should be taken as part of routine patient care by all medical practitioners. This chapter discusses single-gene disorders, i.e. disorders for which there is a single cause, such as a gene mutation. Single-gene disorders cause the patterns of inheritance that obey the laws of segregation and independent assortment that Gregor Mendel first described (autosomal dominant, autosomal recessive, X-linked dominant, X-linked recessive and Y-linked). Because the laws that govern transmission of a single genetic factor are known, a physician or genetic counselor can give accurate estimates of the probability that other family members might possess the mutation. In this chapter we discuss the foundational principles that the reader must know to understand these principles, as well as guidelines on when to refer to a genetic specialist.

2.1 Maximizing the Use of Family Medical History in Disease Risk Assessment

Family history is an essential tool for disease risk assessment, because it incorporates not only shared genetic susceptibilities but also shared environmental, behavioral and cultural factors. Family history remains the single least expensive and most effective means of disease risk assessment. Family history data should be taken as part of routine patient care by all medical practitioners. By noticing patterns of disease within a family, practitioners can often identify family members who have an increased risk for the disorder before they exhibit symptoms, and help high-risk family members tailor their screening, management and/or behavior. It also serves as a vital record of health information about a person and his/her family members. Given the dynamic nature of any family history, the family medical history should be reviewed and updated annually.

The family medical history should consist at a minimum of first-degree relatives (parents, children, and siblings) and second-degree relatives (grandparents, aunts,

K.M. Sweet and R.C. Michaelis, *The Busy Physician's Guide To Genetics, Genomics and Personalized Medicine*, DOI 10.1007/978-94-007-1147-1_2,

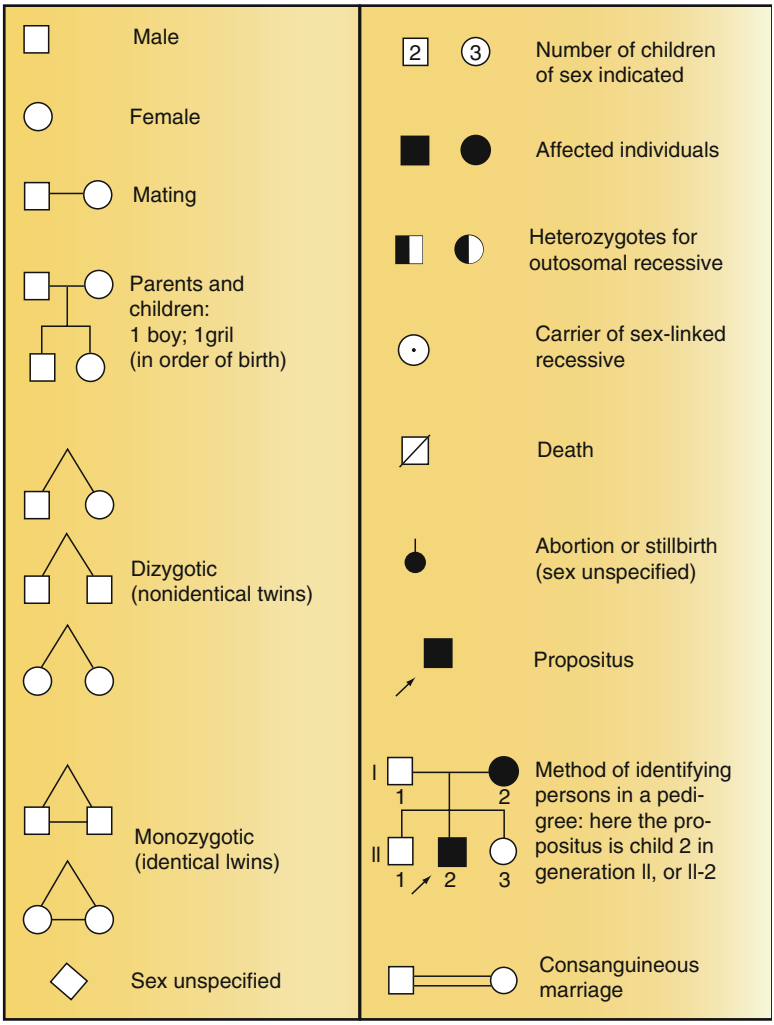

Fig. 2.1 Symbols used in drawing human pedigrees (From the National Center for Biotechnology Information)

uncles, cousins, nieces and nephews). Figure 2.1 illustrates the symbols that are used to illustrate human pedigrees. Start with the person of interest and his/her immediate relatives (parents, siblings, children, nieces, nephews), then expand upward (aunts, uncles, past generations) and downward (if appropriate, grandchildren, great-nieces and great-nephews). Ask not only about major medical problems, but also about morphological abnormalities, cognitive abilities and sensory functions. Be sure to determine the age at which each affected individual was diagnosed. For all deceased family members, be sure to include the age at which the individual died, as well as the cause of death. Determine each individual's ethnicity,

as many genetic disorders vary in frequency among different ethnic groups. Ask about religious heritage as well, as some religious groups may be at increased risk for particular genetic diseases (e.g. rare autosomal recessive disorders in the Amish). For family members with known medical problems, also ask about environmental risk factors such as whether they smoke, what their diet and exercise habits are (were), and if they are (were) overweight.

One of the reasons physicians may not use family histories routinely is that it takes a considerable amount of time to gather the information necessary to construct a pedigree. Patients and family members can construct their own pedigrees, however, using free online programs such as My Family Health Portrait (https://familyhistory.hhs.gov/fhh-web/home.action), which is sponsored by the Surgeon General's office. The patient or family member can construct the pedigree at home, where he/she can consult with other family members to insure the maximal level of accuracy of the information, then bring it or send it as an electronic file to the physician's office.

2.2 Single Gene Disorders

Applications of human genetics in clinical medicine have historically been based on the analysis of *single-gene disorders*. As the name implies, in a single-gene disorder a single gene mutation has a deleterious enough effect on the individual's metabolism to cause the individual to have a disorder, without requiring any other causative genetic or nongenetic factors. Single-gene disorders are also referred to as *Mendelian disorders*, because the patterns of inheritance that are seen in families with these disorders obey the laws of inheritance that Gregor Mendel first described. Collectively there are more than 6,000 single gene disorders, and they affect approximately 4% of the population. Although the typical single-gene disorder is somewhat rare, and some physicians may only encounter a few cases of single-gene disorders in their entire careers, their combined impact on human health is significant.

Because single gene disorders each have a single cause (a gene mutation), it is relatively easy to predict the probability of recurrence for future pregnancies, or the likelihood that an unaffected individual might be a carrier of the causative mutation. In many cases the causative mutation can be identified, and unaffected family members (including fetuses) can be tested to directly determine whether or not they have the causative mutation. In some cases, in vitro fertilization and preimplantation testing can be used to assist a couple who are at risk for having a child with a single-gene disorder.

If the physician thinks a patient has or is at risk for a single-gene disorder, he/she can take a multigenerational pedigree and ask about affected family members on both sides of the family. This may provide a clue to the inheritance pattern and the risk for the patient. There are also online resources available, such as Online Mendelian Inheritance in Man (OMIM at http://www.ncbi.nlm.nih.gov/omim), which allow you to search for genetic disorders that include specific clinical

features. Another useful site is the NIH's GeneTests website (http://www.genetests.org), which provides expert-authored disease reviews as well as an international directory of genetic clinics and testing laboratories. The U.S. Library of Medicine also has information on genetic conditions at the Genetics Home Reference website (http://ghr.nlm.nih.gov/info=inheritance). These websites often enable the physician to narrow the diagnosis down to a specific category of diseases, and may even permit him/her to specify the disorder. Because there can be considerable overlap with respect to the clinical features seen in different disorders within a specific disease category, however, referral to a genetic specialist may be appropriate.

2.3 Understanding Mendelian Patterns of Inheritance (Single Gene Disorders)

Each of the classic Mendelian patterns of inheritance has a characteristic set of features that can be identified by taking a pedigree. The differences in the patterns of inheritance can best be understood if one considers (1) how many copies of the gene an individual possesses and (2) how many copies of the gene must be defective in order for an individual to exhibit the disorder. The terms that are used to describe these patterns of inheritance are:

Autosomal = the gene is on one of the numbered chromosomes, or autosomes. This means everyone has two copies of the gene.

X-linked = the gene is on the X chromosome. Females have two copies of the gene, but males only have one.

Y-Linked = the gene is on the Y chromosome. Males have one copy of the gene, and females don't have any.

Dominant = the individual will have the disorder if he/she only has one defective copy of the gene (Note – the mutation and the disorder can both be referred to as dominant)

Recessive = the individual will not be affected with the disorder unless both copies of the gene are defective (Note – the mutation and the disorder can both be referred to as recessive)

Homozygous = both copies of the gene have the same sequence (normal or mutant[1])

Heterozygous = the two copies of the gene have different sequences

Hemizygous = the individual has only one copy of the gene, as is true for X and Y chromosome genes in males

[1] The term "normal" is often used to indicate the wild-type, or most common, allele of the gene. It is possible, though, to have several different alleles of a gene produce isoforms of the protein that can support healthy development. All such alleles can be seen as "normal," while any versions of the gene's sequence that make a version of the protein that does not support healthy development can be considered "mutant."

Unaffected carrier = someone who has a recessive mutation in one copy of a gene, but a normal sequence in the other copy, and therefore does not have the disorder

Genotype = the alleles the individual has for one or more genes.[2]

Phenotype = the status of the individual's development, including physical characteristics, health status, biochemical measurements and level of susceptibility to diseases.

De novo mutation = A *de novo* mutation is one that arose during the process in which the sperm or the egg that produced the child was made, and was therefore not present in the parent from whom the sperm or egg came. If an individual has a de novo mutation, there will be no prior family history, but the individual's offspring will now be at risk.

When you draw a pedigree, each individual is assigned a number that reflects his/her position in the pedigree. The generations are numbered with Roman numerals, with the oldest generation (usually depicted at the top of the figure) as generation I. Within each generation, individuals are assigned Arabic numbers to indicate their position in the pedigree, with the leftmost person being assigned as number 1. For example, the third person from the left in the second generation of a pedigree is assigned the number II-3.

2.3.1 Autosomal Dominant Inheritance

An autosomal dominant pedigree is depicted in Fig. 2.2. Because the gene in question resides on an autosome, and because an individual only needs to inherit one

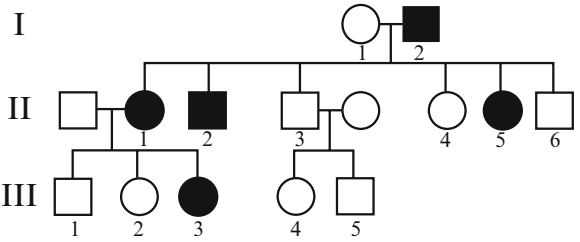

Fig. 2.2 A pedigree illustrating autosomal dominant inheritance (From biologie.uni-hamburg.de)

[2] For our purposes, if you use "N" for a functional, or normal, version of the gene's sequence, and "M" for a defective, or mutant, version of the gene's sequence, there are three possible genotypes for an autosomal gene: NN, NM and MM. For an X-linked gene, using "Xn" to denote an X chromosome with a functional copy of the gene and "Xm" to denote an X chromosome with a mutant copy of the gene, the possible genotypes for a female are XnXn, XnXm or XmXm, while the possible genotypes for a male are XnY or XmY. For Y-linked genes, a male's possible genotypes are XYn or XYm.

mutant copy of the gene to exhibit the disorder, the pedigree exhibits the following features:

- Except for the first person to be affected in the family, who has a *de novo* mutation, affected individuals have at least one affected parent. (The affected individual had to inherit the mutation from one or both parents, and whoever had the mutation would be affected.) One typically sees a vertical pattern of inheritance, with affected individuals in multiple generations in either the maternal or paternal lineage.
- Males and females are affected equally often, and with equal severity. (Males and females both have two copies of each autosome, so the rules apply equally, unlike with X and Y gene mutations discussed below.)
- The trait can be passed either through the male or female germline. (Every child gets one set of autosomes from the father and one from the mother, unlike X and Y genes discussed below.)
- An individual who is homozygous for the mutation is usually more severely affected than an individual who is heterozygous for the mutation. (If one mutant copy of the gene is deleterious enough to cause the disorder, two mutant copies of the gene are often lethal.)
- The descendants of two unaffected parents will not be affected. If the parents are unaffected, they do not carry the familial mutation, and so cannot pass it on. The exception to this is that some affected individuals will possess *de novo* mutations, which arose anew in them. Their parents will be unaffected.
- Each child that results from the mating between a heterozygous affected individual and an unaffected individual has a 50% probability of being affected. (The heterozygous affected parent has one functional and one mutant copy of the gene; there is a 50% probability he/she will pass the defective copy to any given child; any child who inherits the mutation will be affected).

2.3.2 Autosomal Recessive Inheritance

An autosomal recessive pedigree is depicted in Fig. 2.3. Because the gene in question resides on an autosome, and because an individual must inherit two mutant copies of the gene in order to exhibit the disorder, the pedigree exhibits the following features:

- The parents of an affected individual may be either affected or unaffected. (They could be unaffected carriers and still possess a mutation they can pass down. In fact, because these causative genetic mutations are rare in general, most of the people who possess the mutation only have one mutant copy of the gene, and are therefore unaffected carriers.)
- An autosomal recessive disorder can skip one or more generations in the pedigree. (There can be several generations that contain unaffected carriers, but no affected individuals. The disorder will reappear in the pedigree when two

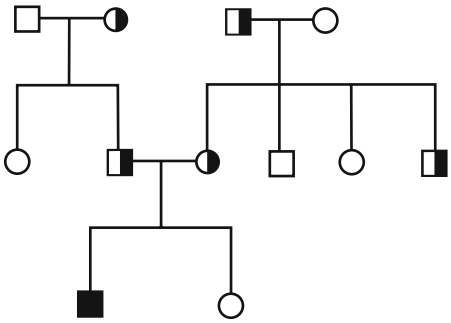

Fig. 2.3 A pedigree illustrating autosomal recessive inheritance (From athenadiagnostics.com)

unaffected carriers mate and one of their children inherits the mutation from each of them.)

- In some families, affected children are the result of consanguineous matings (matings between relatives). (We all possess a small number of recessive gene mutations, but two relatives are more likely to possess recessive mutations in the same gene than two unrelated people are. They are therefore more likely to have a child who has inherited two recessive mutations in the same gene, and has the recessive disorder.)
- Males and females are affected equally often, and with equal severity. (Males and females both have two copies of each autosome, so the rules apply equally, unlike with X and Y gene mutations discussed below.)
- The trait can be passed through both the male and female germline. (Every child gets one set of autosomes from the father and one from the mother, unlike X and Y genes discussed below.)
- When both parents are affected, all their offspring are affected. (If the parents are affected, they must have the MM genotype. Therefore, they only have mutant copies of the gene to pass down. Each of their offspring therefore must have the MM genotype.)
- When both parents are unaffected carriers of the mutation, each of their children has a 25% probability of being affected. (There is a 25% probability the child will have the NN genotype, a 50% probability the child will have the NM genotype, and a 25% probability the child will have the MM genotype.)

2.3.3 X-Linked Recessive

An X-linked recessive pedigree is depicted in Fig. 2.4. Because the gene in question resides on the X chromosome, sons and daughters have different probabilities of inheriting the mutation from an affected father. In addition, a daughter must inherit

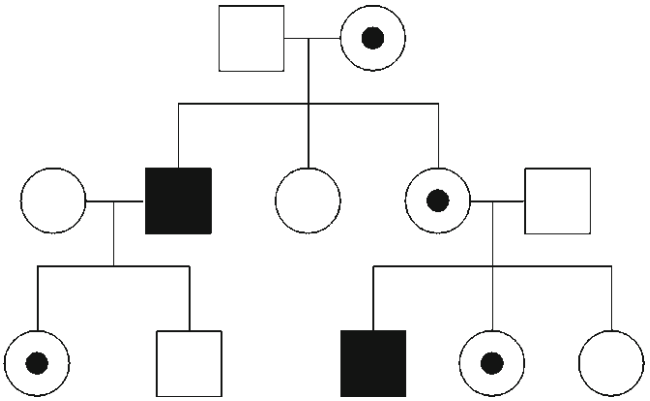

Fig. 2.4 A pedigree illustrating X-linked recessive inheritance (From ocw.tufts.edu)

two mutant copies of the gene in order to be affected, but a son will be affected if his only copy of the X-linked gene is defective. The pedigree exhibits the following features:

- Many more males are affected than females. (A female can be affected under certain circumstances, but in most cases, the second copy of the gene can support normal function. The male does not have a second copy of the gene, so if he gets a mutation, he will have the disorder.)
- Affected males get their mutation through their mother; there is no affected-male-to-affected-male transmission. (The son gets his Y chromosome from his father, and his X chromosome from his mother.)
- A mating between an affected male and a mutation-free female produces no affected offspring, but all daughters are carriers. (An affected male will pass a mutant gene down to every daughter, because he only has one, mutant, copy of the gene to pass down. He passes his Y chromosome to all his sons, however.)
- If an unaffected carrier female mates with an unaffected male, each of their sons has a 50% chance of being affected, and each of their daughters has a 50% chance of being an unaffected carrier, and a 50% chance of being mutation-free. (Each child has a 50% chance of inheriting the mother's mutant gene. If a son inherits it, he will be affected. If the daughter inherits it, she will be an unaffected carrier, because she will inherit a functional copy of the gene from her unaffected father.)

2.3.4 X-Linked Dominant Inheritance

X-linked dominant inheritance is relatively rare. An X-linked dominant pedigree is depicted in Fig. 2.5. Because the gene in question resides on the X chromosome, sons and daughters have different probabilities of inheriting the mutation from an

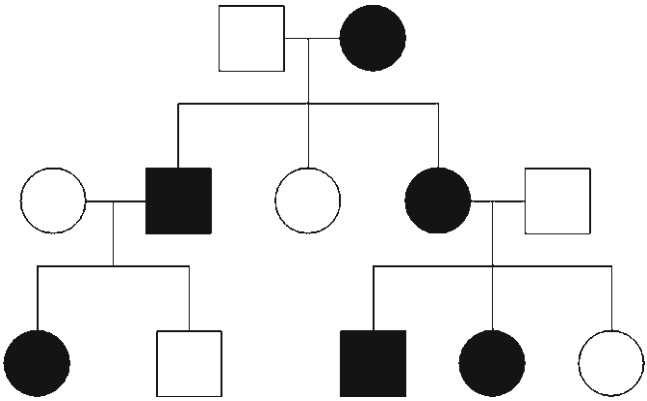

Fig. 2.5 A pedigree illustrating X-linked dominant inheritance (From uic.edu)

affected father. Because the mutation is dominant, however, sons and daughters will both be affected if they inherit one mutant copy of the gene. The pedigree exhibits the following features:

- An X-linked dominant pedigree can resemble an autosomal dominant pedigree in that there will be many affected family members, but there is no affected male-to-affected-male transmission. (The son gets his father's Y chromosome, not his X chromosome.)
- Both males and females may be affected, but males are often more severely affected than females (Expression of the trait is more variable among females than among males, because females possess another copy of the gene, which can modify the phenotype.)
- If a heterozygous, affected female mates with an unaffected male, each child has a 50% probability of being affected. (Because the mother has one functional copy and one mutant copy of the gene, each child has a 50% probability of inheriting the mother's mutant gene. If the child inherits the mutation, he/she will be affected.)

2.3.5 Y-Linked Inheritance

Y-linked disorders are very rare; there are only a few genes on the Y chromosome that, when disrupted, will produce a change in the individual's phenotype. Because only the males carry a Y chromosome, a Y-linked pedigree exhibits the following features:

- Only males are affected. (None of the females has a Y chromosome.)
- Every affected male passes the mutation to all his sons. (Each son gets his father's only Y chromosome.)

2.4 Assessing the Risk of Recurrence in Mendelian Pedigrees

In order to determine the probability that any individual has a mutation that is present in a family member, one must apply the rule of probability theory that states:

> The probability of a set of independent events is equal to the product of the probabilities of the individual events.

In other words, to find the probability that a series of events will happen, you determine the probability of each individual event happening, and multiply those probabilities together. As a simple example, consider why the probability of tossing a coin three times and getting three heads is 1/8. The probability of getting a head on any given toss is 1/2, and the outcome of each coin toss is not influenced by the outcomes of the other coin tosses. Multiplying $1/2 \times 1/2 \times 1/2 = 1/8$.

In order to calculate the risk of a disorder occurring in any given individual, you must factor in the probability that the individual in question will inherit the mutation that exists in one of his/her parents, along with the probabilities of any other events that must occur for the individual to be affected. For example, the individual must inherit a second mutation from the other parent if the disorder/mutation is recessive, or the child must inherit a Y chromosome from the father to be affected if it has inherited an X-linked recessive mutation from the mother. On the other hand, if the mutation is dominant, nothing needs to happen, other than inheriting the mutation, for the individual to be affected.

When considering a family of siblings, keep in mind that the laws of probability apply to each child independently. Many people make the mistake of thinking that, if the probability of an affected child is 1/4, if the first child is affected, the next three are safe. This is not true; each child has a 1/4 probability of being affected, and there is no reason why a couple with four children cannot have several affected children.

In order to determine the probability that any given individual possesses a mutation that appears in a family member, you must determine which events must occur in order to allow that individual to inherit the mutation, then multiply those probabilities together. This is made easiest if you write the genotypes of the critical individuals (ex. NM or XnY) on the pedigree; this makes it easier to see the probability that the individual will pass down his/her mutation. For example, in the pedigree depicted in Fig. 2.6, individual I-2 is an unaffected heterozygous carrier of an autosomal recessive mutation. Because she has the NM genotype, there is a 1/2 probability that she will pass the mutation down to any given child. Therefore, her son II-1 has a 1/2 probability of having the mutation. Because I-2's husband I-1 is mutation-free, he passes the N allele to II-1. If II-1 inherits his mother's mutation, he will therefore have the NM genotype as well. Similarly, because II-1's wife II-2 is mutation-free, their son III-1 can only inherit the mutation from II-1. III-1 therefore has a 1/2 probability of inheriting the mutation if II-1 possesses it.

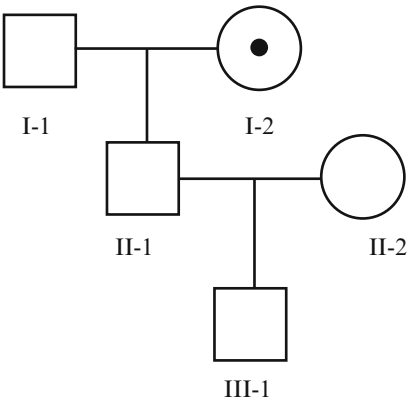

Fig. 2.6 A pedigree illustrating autosomal recessive inheritance

Imagine that all you know is that I-2 has the NM genotype, and you have been asked to determine the probability that III-1 has inherited I-2's mutation. In order for this to have happened, I-2 must have passed the mutation down to II-1 (probability = 1/2, because I-2 has the NM genotype), and II-1 must have passed the mutation down to III-1 (probability 1/2, because II-1 would have the NM genotype if he inherited I-2's mutation). Therefore, without knowing whether II-1 inherited his mother's mutation, you know that III-1 has a 1/4 probability ($1/2 \times 1/2$) of having inherited I-2's mutation. If II-1 has been tested and found to have the NM genotype, then the probability that III-1 has inherited the mutation is 1/2.

2.5 Carrier Frequencies for the More Common Recessive Single-Gene Disorders

If you want to calculate the probability that an individual will have a child who is affected by a recessive disorder, you often know the probability of the child inheriting the mutation from the parent who has a family history of the disorder, but you also need to know the probability that the other parent carries the mutation. If the other parent does not have a family history of the disease, you must rely on the estimate of the carrier frequency in the general population. The carrier frequency gives you the probability that the other parent has the NM genotype. After that, the parent must pass the mutation down to the child (probability = 1/2) for the child to be affected.

Carrier frequencies for recessive mutations can be surprisingly high; Table 2.1 lists the carrier frequencies for some of the more common recessive mutations. Note that there can be significant differences in carrier frequencies between different ethnic groups.

Table 2.1 Carrier frequencies for some recessive mutations

Disorder	Gene symbol and name	Carrier frequency and population[a]
Cystic fibrosis	*CFTR*; cystic fibrosis transmembrane regulator	1/25 C; 1/150 As; 1/70 Af
Hemochromatosis	*HFE*; hereditary hemochromatosis protein	1/10 C; 1/43 Af; 1/36 H
Alpha-1 antitrypsin deficiency	*SERPINA1*; serine protease inhibitor type A1	1/50 U
Spinal muscular atrophy	*SMN1*; survival of motor neuron 1	1/50 U

[a] *C* Caucasians, *As* Asian-Americans, *Af* African-Americans, *H* Hispanic-Americans, *U* Unspecified American population

2.6 Referring to a Genetic Specialist

Although medical practitioners can provide basic risk assessment using personal and family medical history, more detailed analysis and focused risk assessment is at times necessary. In these cases, it is appropriate to refer the patient to a genetic specialist for consultation. The National Society of Genetic Counselors offers a searchable directory of genetic counselors (http://www.nsgc.org). One can search by city, name, area of practice and zip code.

Genetic consultation is a health service that provides information and support to people who have, or may be at risk for, common and rare diseases. It is an important part of the decision-making process for genetic testing, and helps people better understand their risk for disease in the context of their personal and family medical history. It is important to recognize that while a family medical history provides information about the risk of specific health concerns, this alone does not provide a complete assessment of the individual's risk. Nongenetic influences such as environmental and lifestyle factors also impact risk for multifactorial disorders. Some individuals with no family history of a disorder may still be at risk of developing that disorder if they are exposed to a high level of one or more of the dietary, environmental or lifestyle factors that can cause the disease.

Genetic consultation is typically provided by a genetic counselor with an advanced degree, at times in tandem with a physician who specializes in clinical genetics. Genetic counselors and clinical geneticists have completed certified training and board examinations. Like other medical specialists' services, the consultation and risk assessment service is covered by most health insurance companies. Some insurance plans require a physician referral.

Genetic consultations usually take place in a hospital, genetic center or other type of medical center or office. These meetings are often in-person visits with an individual or family, and usually take 1–2 h. Some centers are starting to develop telemedicine programs for interactive genetic counseling as well. The genetic professional will review the medical, family, lifestyle and exposure history, and assess familial risk for common and rare disease through analysis of a four-generation pedigree. The consultation may also include a targeted physical examination.

In addition to providing a diagnosis, the genetic professional will also educate the patient or family member about the basic genetic and medical issues involved, the expected pattern of inheritance for single gene disorders, the potential risk for other family members and the options for testing and treatment.

Historically, genetic testing has been offered in this context of in-person risk assessment and counseling. This allows the genetic professional to discuss the benefits, limitations and risks associated with testing with the individual and his/her family members. It allows opportunity to illuminate the necessary facts, clarify alternatives and anticipate consequences associated with the test results. It allows acquisition of informed consent, blood (or other samples) for testing, and preauthorization from the patient's insurance company. Once the results of the test are available, the genetic professional meets with the patient or parent to discuss the results of testing, and as necessary, aid in the adjustment process. Counseling and other support is provided to help the patient and family members cope with psychosocial issues that may arise. Targeted interventions, such as recommending that the patient undergo screening tests regularly, or change certain aspects of his/her diet, environment or lifestyle to help lower his/her risk for the disease, may be provided. The option to participate in genetic research is offered when appropriate.

Understanding what to look for in the family history and how this affects risk can be challenging. If any of the following apply, we recommend referral for genetic consultation where more comprehensive risk assessment can be provided:

- A personal or family history of a single gene disorder such as cystic fibrosis, biochemical abnormality, birth defect, mental retardation or chromosomal disorder.
- A person having two or more close relatives on the same side of the family with the same or related condition (e.g. breast cancer, heart disease, Parkinson disease).
- A person having a disease at an earlier than expected age (e.g. colon cancer before age 50; cardiovascular disease in a male prior to age 55 years, in a female prior to age 65 years).
- A person with more than one primary disease, or exceptional presentation of common conditions (e.g. breast and ovarian cancer in the same person).
- Abnormal or unusual test results for which you have no explanation.
- Two or more unexplained pregnancy losses (miscarriages).
- A woman who is pregnant or plans to become pregnant at or after age 35.
- A person who is very anxious about his/her disease risk.

2.7 New Genomic Applications for Complex Disease Will Change Approaches to Genetic Counseling and Personalized Medicine

The recent introduction of genome wide association (GWA) studies (see Sect. 4.7), whole-genome sequencing and especially the rise in commercial applications of these technologies is affecting the landscape of genetic counseling, just as it is

changing approaches to personalized medicine. The ability to test an individual to determine risk for multiple common complex diseases will soon become commonplace and increasingly cost-effective. A number of genomic tests for multifactorial diseases have also become directly available to consumers, who can simply send in a saliva sample and receive their results through an Internet web portal. Basic education on genetics and what the results of the tests might mean is often provided. Some of the commercial companies offer access to genetic counselors, but others do not.

Whether or not one endorses increased public access to genomic information, it is here to stay. One must remain cautious, however, as appropriate clinical application will continue to be a delicate endeavor. What is required are new approaches to genetic consultation for multiple disease risk as well as genomic medicine delivery models based on participant and provider education and feedback. Much remains to be learned about both the genetic and nongenetic factors that influence one's risk, especially for common multifactorial diseases, and how this risk could modify health and behavior. Basic questions of how to properly interpret and apply this information are critical, because the fact that an individual possesses risk-increasing gene alleles does not always influence the physician's management strategy. Many gene variants have no direct clinical significance. And of those that do have clinical significance, most have relatively low penetrance and minimally influence disease risk. How this information not only impacts a patient's health but that of their family members must also be taken into consideration.

The growing wealth of genomic information will increase our ability to customize approaches to medical treatment. In the not too distant future, increased understanding of our molecular machinery (genetic, epigenetic, RNA, protein and other biomarkers) and how these biological factors interact with dietary, environment and lifestyle factors will allow for more extensive patient profiling at point of contact. We will have the means to more fully understand a patient's disease or predisposition toward disease, and to develop disease management approaches that are likely to work best for that individual patient.

What we have learned so far about hereditary forms of breast and colon cancer, as well as other single gene disorders, has helped to pave this road. In subsequent chapters, you will learn about several areas in which personalized medicine tests are providing real, albeit often limited, benefits. We now need to refine and extend this knowledge for all multifactorial diseases, as therein lies the heart of the new era of predictive, preventive, personalized and participatory (P4™) medicine. The focus of this new approach to medicine is to define and understand the origins of disease at the genomic level, to allow for more accurate prediction of a person's risk for disease, and to enable health care providers to better promote the individual's health and well-being from birth. Translating this knowledge into patient applications and therapeutics will allow for more effective preventative measures, and personally targeted approaches based on an individual's unique genetic makeup. The last "P" of P4™ Medicine, patient participation, refers to the fact that knowing one's level of genetic risk for specific diseases enables patients to take a more active role in their own health care, because it may enable some people to adjust their diet,

environment and lifestyle habits to reduce their exposure to the nongenetic factors that contribute to those specific diseases, thereby reducing their overall probability of developing the disease.

Selected References

American Society for Human Genetics (2004) Your Family History. http://www.ashg.org/genetics/ashg/educ/007.shtml

Bennett RL (1999) The Practical Guide to the Genetic Family History. New York: Wiley-Liss, Inc.

Centers for Disease Control and Prevention, Office of Genomics and Disease Prevention (2010) Using Family History to Promote Health. http://www.cdc.gov/genomics/public/famhistMain.htm

U.S. Department of Health and Human Services (2004) U.S. Surgeon General's Family History Initiative. http://www.hhs.gov/familyhistory/

Chapter 3
Types of Genetic Tests and Issues Associated with the Interpretation of Their Results

Abstract With continued development and availability of genetic test offerings, there will be greater need for physicians to help their patients interpret their results and to know when it is appropriate to refer to a genetic specialist. In addition, physicians should be able to understand the process whereby basic research discoveries are translated into clinically useful genetic tests. This chapter describes the genetic tests that are most commonly used in research and clinical settings, as well as by genetic testing companies, including some that advertise their services directly to consumers. We discuss the specific type of genetic variation that each test assesses, along with several principles that must be taken into account in order to correctly interpret the clinical significance of a genetic test result. The chapter also describes the process that is used to evaluate the clinical utility of newly developed genetic tests, and provides the reader with websites that maintain up-to-date information regarding the genetic tests that are available in all fields of medicine.

3.1 Accessing Current Information on Available Personalized Medicine Tests

There are literally thousands of genetic tests now available, and the list of laboratories and commercial companies that offer personalized medical testing is growing too rapidly for any book to provide a listing of available tests that will be current when the reader acquires the book. The Internet sites listed below maintain current listings of the genetic tests that are available, and will provide you with the most up-to-date information possible.

At the present time, most of the personalized medicine tests that have true clinical utility detect the presence of risk-increasing alleles that have high penetrance, such as the mutations that cause Mendelian single gene disorders (discussed in Chap. 2). If the risk-increasing allele's penetrance is high enough, testing that one polymorphism can provide clinically useful information regarding the individual's risk for a specific disease or a specific response to a drug.

K.M. Sweet and R.C. Michaelis, *The Busy Physician's Guide To Genetics, Genomics and Personalized Medicine*, DOI 10.1007/978-94-007-1147-1_3,

It also appears, however, that the vast majority of the genetic variants that influence risk for common complex diseases such as cancer, diabetes and asthma have relatively low penetrance. Because each variant only makes a small contribution to the individual's risk, in order to produce the most clinically useful test possible for a specific disease, it will be necessary to combine information from dozens, and perhaps hundreds, of these low-penetrance variants, as well as determine exactly how the information about the different variants will be factored into the predictive algorithm. Despite the low penetrance of these individual variants, together they account for a significant portion of our medically relevant genetic variability. As the field progresses, disease-based tests will be developed that include hundreds, perhaps thousands, of tests for variants that have low penetrance for that particular disease.

Genomic researchers are identifying more of these low-penetrance variants every day, and it will not be long before researchers determine the best way to use these low-penetrance variants in personalized medicine tests. At the present time some genetic testing companies, especially those that offer genome-wide screens directly to consumers, offer tests that detect risk-increasing alleles that have low penetrance. It may be easy for a lay person to overestimate the medical significance of such a variant, and medical specialists must be prepared to help the patient understand that some of the genetic tests that are available do not provide medically actionable information.

There are several types of personalized medicine tests that can provide information that will help you better estimate risks for disease and make choices regarding the way you treat your patients. This is not a comprehensive list; the Internet websites listed below provide the most comprehensive listing of tests that are currently available. This list merely serves to illustrate the types of services that are useful to the practitioner.

- Presymptomatic testing for prediction of risk for adult-onset disorders such as Marfan syndrome, neurofibromatosis, hereditary hemochromatosis, or certain forms of adult-onset cancer such as hereditary breast-ovarian cancer syndrome
- Carrier screening to determine if a healthy individual carries an autosomal recessive mutation that is known to cause a specific disease, such as cystic fibrosis (the typical healthy individual may carry between 50 and 100 such mutations)
- Preimplantation genetic diagnosis to screen embryos for genetic disease
- Prenatal diagnostic testing to detect deleterious mutations in fetuses
- Newborn screening

The following Internet resources will enable you to access the current information regarding the clinically useful personalized medicine tests that are available at any time.

The NIH's GeneTests website (http://www.ncbi.nlm.nih.gov/sites/GeneTests/) provides the most comprehensive and current listing of laboratories that offer genomic testing. The website includes:

Gene Reviews – Expert-authored articles, each devoted to a specific disorder or related family of disorders. Gene Reviews focuses primarily on single-gene

disorders for which molecular genetic testing is available and disorders for which one of the causative genes has been identified.

Laboratory Directory – An international directory listing not only the CLIA-certified American laboratories that perform clinical testing, but also clinical testing laboratories from other countries, research laboratories that perform other types of genetic tests pertinent to a particular gene or disease, and support resources for the patient and his/her family as well. Research laboratories can often provide highly useful information, and some of the research laboratories' findings can be confirmed by a CLIA-certified or otherwise accredited clinical laboratory.

You can search the laboratory directory by disease, gene, protein, laboratory or laboratory director, and refine your search by location, if desired. The website provides detailed information about the tests, including the techniques used (ex. sequencing versus focused mutation detection versus comparative genome hybridization) and the extent of coverage (ex. entire coding sequence plus splice junctions or screening for specific mutations).

Clinic Directory – An international directory of genetics and prenatal diagnosis clinics.

Educational Materials – Including an illustrated glossary, information about the types of genetic services that are available, and a set of PowerPoint® slide presentations.

In addition, the Association for Molecular Pathology (AMP) maintains a directory of information on the diagnosis and/or monitoring of solid tumors, hematologic malignancies and infectious diseases as a service to its members and the public (http://www.amptestdirectory.org/). AMP members voluntarily list their laboratories and research or clinical molecular tests for inclusion in this Directory. The AMP Test Directory is not a membership directory, nor is it a comprehensive list of members whose laboratories provide testing. This information is available to identify laboratories that might serve a given need.

3.2 Standard Format for Genetic Test Results

There are many ways in which the human genome is variable, and different tests assess different aspects of the DNA sequence or the regulation of gene activity. Regardless of the type of test that is performed, however, the results are always reported as a comparison between the patient's sequence and a specific reference sequence. Recall that there are many different specific versions of a gene's sequence that are capable of producing an isoform of the protein that has the typical (i.e. "normal") level of activity. There is therefore no single sequence that serves as the "normal" sequence for any gene; any sequence that is predicted to produce an isoform of the protein that has the typical level of activity can be considered a "normal" sequence.

Positive for a deleterious mutation. This notation indicates that a known risk-increasing gene allele has been identified. This would include mutations that are

believed to significantly change the level of activity in the gene, mutations that alter the amino acid sequence of the protein drastically or result in a truncated protein, and deletions, duplications, insertions or other chromosome rearrangements that significantly alter the activity of one or more genes.

When a deleterious mutation has been found, the laboratory will usually include a statement recommending that family members who are in a position to have inherited the mutation be tested. This testing can help estimate the unaffected family members' risks of developing the disease more accurately. If the penetrance of the risk-increasing allele that is detected is limited, this may also enable some family members to adjust their diet, environment or lifestyle to reduce their exposure to the critical nongenetic factors, thereby helping them postpone, mitigate or possibly even avoid developing the disease. If the risk-increasing allele has too high a penetrance, however, there will be little the individual can do to reduce his/her overall risk for developing the disease.

Genetic variant of uncertain significance. These are sequence variants whose clinical significance has not yet been determined. Some coding sequence variants cause the substitution of one amino acid for another that is only slightly different in size, charge and/or polarity. It often requires extensive laboratory research to determine whether this amino acid substitution changes the production or activity level of the gene's protein. In addition, the promoter regions of many genes contain polymorphic sites whose influence on gene activity has not yet been determined. There are also many variations in intronic sequences for which the functional consequences are unknown. Sometimes further studies, such as testing other family members with and without the disorder, or testing a population of healthy control subjects, can provide insights regarding the clinical significance of the sequence variant.

No potentially causative mutations found (*a* "*negative*" *test*). If the patient's gene sequence completely matches a sequence that is believed to produce an isoform of the protein that has the typical level of activity (i.e. one of the "normal" sequences), the report will say that no potentially causative mutations have been found.

One should always interpret a report of no causative mutations carefully, because there are several reasons why a sequence variant that is assumed to have no effect on the gene's protein might actually affect protein production or activity level. For example, there are two assumptions that many analysts adhere to when they interpret the results of genetic tests: that a "synonymous," or "silent" base substitution that is not predicted to change the amino acid sequence of the protein has no effect on the protein's function; and that most single nucleotide substitutions in introns will not affect the protein, because the introns get spliced out during posttranscriptional RNA processing (see Sect. 1.3.3). Because the intronic nucleotides are spliced out, some changes in intronic nucleotide sequences have no effect on protein production. Some intronic nucleotides help the cell's splicing machinery identify the splice sites, however; changes in these nucleotides may alter RNA splicing and drastically reduce or abolish production of the protein.

There are known examples of synonymous exonic variants that alter the splicing pattern of the RNA or alter the rate of transcription of the gene or translation of the mRNA. In addition, while there are intronic nucleotides that are known to be essential for proper RNA splicing, it is also known that other, as yet unidentified, intronic nucleotides also control the splicing process in some genes. If a variant has been found, it is important to note what that variant is, in case future research reveals that the variant has clinical significance. Some laboratories will send out followup reports if this happens, but some will not.

3.3 Risk-Increasing Gene Alleles Often Have Limited Penetrance

It is important to know the degree of *penetrance* of any given risk-increasing allele in order to interpret the results of a genetic test properly. As we discussed in chapter (see Sect. 1.5), the penetrance of a risk-increasing allele is defined as the percentage of people who possess the risk-increasing allele and also develop the associated disorder. An individual who possesses the risk-increasing allele but does not exhibit the associated disorder is considered a case of *nonpenetrance* of that risk-increasing allele.

There are many reasons why someone who possessed the risk-increasing allele of the gene in question would not exhibit the disorder. For example, as described in Sect. 3.18.4, some of the polymorphisms that are tested do not assess the individual's status at the critical functional polymorphism, but assess the status at a locus that lies close to the critical functional polymorphism. The individual under study may represent an example of the rare recombinations that can occur between the locus that was tested and the critical functional polymorphism. In this case, the individual inherited a rare chromosome that has the "risk-increasing" allele of the locus that was tested, but did not also contain the true risk-increasing allele of the critical functional polymorphism.

Even if the test directly assesses the individual's status at a critical functional polymorphism, it is no surprise to see someone who possesses a risk-increasing allele for that polymorphism, but still does not have the disorder. As the term "multifactorial disorder" denotes,[1] there are usually a number of other genetic and/or nongenetic factors that must be present for the individual to develop the disorder. The individual may not possess any of the other critical genetic factors, or may not have encountered enough of the relevant nongenetic factors that contribute to the disease. Note that this may be a function of the individual's age; many nongenetic factors will accumulate over the individual's lifespan (ex. iron accumulation leading

[1] These diseases are often referred to as complex diseases or multifactorial diseases. We have chosen the term "multifactorial," to emphasize the fact that there are multiple contributory factors that must combine for the individual to develop the disorder.

to hemochromatosis). A properly worded risk estimate should specify the time frame within which one is estimating the individual's risk (ex. by age 50, or during the entire lifetime).

The physician must be careful to ascertain that the patient understands that possessing what some reports call a "high-risk" allele does not automatically put the individual at high risk for the disorder or adverse drug reaction (ADR). The most highly penetrant risk-increasing alleles do convey a high risk for the disease upon the individual; these alleles can be called "high-risk" alleles. The vast majority of risk-increasing alleles have lower penetrance, however; these alleles only increase the individual's risk by a small amount. If the individual possesses a risk-increasing allele that has low penetrance, if there are any dietary, environmental or lifestyle-related factors the patients can regulate to minimize his/her overall risk, the patient should be advised to take a more active role in his/her health care by reducing those factors to whatever degree is reasonably achievable.

3.4 The ACCE and EGAPP Projects Evaluate Emerging Genetic Tests

The Centers for Disease Control and Prevention (CDC) have developed a process for evaluating emerging genetic tests. The CDC-sponsored ACCE Project performs an evaluation that hinges around four aspects of the test: *A*nalytic validity; *C*linical validity; *C*linical utility; and the *E*thical, legal and social issues related to genetic testing.

The *analytic validity* of the test refers to whether the test properly measures the phenomenon it is intended to measure. This is often the least controversial aspect of the evaluation. The technologies that are used for DNA testing have wide application, and have not only been subjected to rigorous quality control procedures, but often have withstood at least a moderately stringent test of time and usage. The laboratories that perform the medically useful tests are all certified under the Clinical Laboratory Improvement Amendments of 1988 (CLIA), which requires stringent quality control measures. In addition, the kits that are developed and sold to other laboratories are classified as medical devices, and as such are subject to regulation by the US Food and Drug Administration (FDA). There is usually little doubt regarding the analytic validity of the results these laboratories generate.

There is much less information available regarding the probability for error in the newer whole-genome scans that are becoming increasingly available, including those that are offered directly to consumers. Many of the tests that are included in these whole-genome screens have been developed in the laboratory that uses them (laboratory-developed tests, LDTs). While the test is validated in the laboratory that develops it, LDTs are not subject to FDA regulation. The laboratory that developed the test cannot sell it to other laboratories, but it can use the test as part of the service it offers to consumers.

The probability for error in any single genetic test is very small. When you perform a huge number of individual tests, however, as is done in a whole-genome

screen, the probability that you will make an error on one of the tests grows significantly. For example, if a test is 99% accurate, the probability that you will not make an error on that test is 99% (0.99). If, however, you add a second test, your probability of not making an error on either test is now 98% ($0.99 \times 0.99 = 0.98$). After 20 tests, your probability of not making any errors is only 82%. In fact, if every test is 99% accurate it only takes 68 tests for your probability of not making any errors to be less than 50%. Even if each single test is 99.9% accurate, it only takes 693 tests before your probability of not making a mistake in one of your tests is less than 50%.

Clinical validity refers to how reliably the result of the test predicts a clinical outcome, such as the individual's risk for developing a disorder or his/her response to a drug. *Clinical utility* refers to whether the result of the test changes the plan for treating the patient. Tests that detect risk-increasing alleles that have high penetrance are clinically useful, because knowing the individual possesses one of these alleles will often prompt the physician to prescribe an especially rigorous schedule of screening tests, and perhaps other measures that have been reported to detect or prevent the disease. If the risk-increasing allele has low penetrance, however, possessing one risk-increasing allele may increase the individual's overall risk for the disorder to such a small degree that learning that the patient possesses the risk-increasing allele will not change the plan for the patient's treatment.

The final aspect of the evaluation assesses the ethical, legal and social issues that attend genetic testing. Given the diverse nature of people's opinions and the ethical concerns that arise over personal health-related information being generated and stored in a database, it is no surprise that there are many who worry that this information can either have adverse psychological impacts on people, or be misused by insurance agencies and employers to discriminate against people with high risks for chronic diseases.

If personalized medicine tests are to be widely used in medicine, tests must be developed that have enough clinical validity that insurance providers will see them as cost-effective, and cover genetic testing as a routine part of an individual's health care. The CDC and representatives of the insurance industry are currently involved in a project entitled The Evaluation of Genomic Applications in Practice and Prevention (EGAPP). This project is designed to establish standards for bringing new tests from the research stage to clinical application, so that insurance carriers can be assured that covering the cost of genetic testing will lead to reduced health care costs in the long run.

3.5 Assessing the Usefulness of a Genetic Test

There are two ways in which the concept of usefulness can be applied to a genomic test. A test is useful at the level of the individual patient if it helps clarify the individual's level of risk for the disorder. Tests that help estimate the individual's level of risk often shed light on the etiology of the disorder, and may provide insights

into how to treat that individual patient. In order to recommend a genomic test for routine clinical use, however, population-level issues such as sensitivity and specificity are of paramount importance. The test must detect the vast majority of those who are or will be affected by the disorder, with an acceptably small rate of false positives and false negatives.

A genomic test's clinical validity can be subdivided into *positive predictive validity* (*PPV*) and *negative predictive validity* (*NPV*). A test's PPV is defined as the percentage of people with a positive test result (i.e. who possess the risk-increasing allele) who develop the disorder within a specified time frame. Note that because the genetic and nongenetic factors that contribute to the disease can accumulate over time, the PPV should refer to the probability of developing the disease over a specified time frame (the lifetime risk is often given). The PPV of a genetic test is directly related to the test's level of sensitivity (sensitivity = the proportion of people who have the disorder or who will develop the disorder who had a positive test). There are several reasons why there may be people who develop the disorder despite having had a negative result (or lack of risk-increasing alleles) on the test. For one, the test may have failed to detect a risk-increasing allele that was present. Alternatively, the individual may not possess that particular risk-increasing allele, but may have developed the disorder because he/she possesses risk-increasing alleles in other critical genes, and/or has been exposed to one or more causative nongenetic factors. Finally, the PPV of any genetic test is reduced by anything that reduces the penetrance of the risk-increasing allele. As described above, there are a number of reasons for nonpenetrance of a risk-increasing gene allele.

A test's NPV is defined as the percentage of people who have a negative test result (i.e. they do not possess the risk-increasing allele) and will not develop the disorder during a specific time period after the test. As with the PPV, the NPV should refer to the probability of developing the disease over a specified time frame. The NPV is directly related to the specificity of the test (specificity = the percentage of people who do not have the disorder and were classified as low-risk by the test). People who possess the risk-increasing allele the test is intended to detect but do not develop the disorder represent examples of the test's lack of specificity. Anything that reduces the specificity of a test results in people who are not going to develop the disorder being put through the expense, inconvenience, emotional challenge and potential danger of the treatment because of their positive test result.

Tests for gene mutations that cause single-gene disorders have excellent clinical validity, because in those cases there are no other causative factors to be considered, and the penetrance of these mutations is essentially 100%. Unfortunately, however, because the genetic variants that contribute to multifactorial disorders often have limited penetrance, the genetic tests that assess the individual's status for these functional polymorphisms have considerably less predictive power than tests for single-gene disorders do. When reporting the results of genetic tests to patients, one must be certain the patient does not automatically equate having one or even a few risk-increasing alleles that have low penetrance with having a high overall risk for the disorder.

3.6 Even an Informative Genetic Test May Have Limited Clinical Utility

Clinical utility refers to whether the results of a genetic test will change the plan to treat or manage the disorder. Unfortunately, the fact that a genetic variant contributes to one's risk of developing a multifactorial disorder does not necessarily mean that the test for this genetic variant will have substantial clinical utility. There are many occasions when a new genetic discovery provides important insights into the pathogenesis of that particular disease, but determining the patient's status for the associated genetic polymorphism will not change the way the physician treats the patient.

Because of the number of genetic and nongenetic factors that influence one's risk for the typical multifactorial disorder, most of the risk-increasing gene alleles that contribute to the common multifactorial disorders confer only a small increase in the individual's risk for developing the disorder. At the present time, our ability to specify the individual's risk is still limited to categorical assessments such as "high risk" versus "moderate risk" versus "general population risk." In this context, the small increase in overall risk that usually results from possessing a risk-increasing allele is not likely to elevate an individual into the next categorical level of risk. Only the most highly penetrant risk-increasing alleles are capable of singlehandedly elevating the individual's categorical risk, and playing the preeminent role in shaping the decisions regarding the patient's treatment.

The most important nongenetic factors are often more powerful elevators of the individual's risk than the currently known genetic factors are. For example, it is believed that cannabis use during the teenage years significantly increases one's risk for schizophrenia, while it is considerably less certain what genetic variants contribute to the disease, and to what degree.

3.7 Single Nucleotide Polymorphisms (SNPs) Are the Most Commonly Tested Polymorphisms

As mentioned in Sect. 3.2, a SNP is merely a single nucleotide position at which one person will have an A, for example, while another person will have a C, G or T in the sequence. As of this writing there have been approximately 15 million SNPs reported to exist in the human genome, and SNPs are the genetic polymorphisms that are most commonly used for genetic testing. As we discussed in Chap. 1 (see Sect. 1.3), if the nucleotide substitution causes an amino acid substitution, and the amino acid that gets added into the protein is different enough in electrical charge, polarity or size from the amino acid it is replacing, this may cause the protein to work at an activity level that is significantly higher or lower than the activity level at which most of the isoforms of that protein work. If that protein's function is such that it influences the individual's susceptibility to a multifactorial

disorder, the individual with the unusual-activity isoform of the protein will be either more or less susceptible to the disorder than the typical person is.

A SNP test is sometimes named for the DNA nucleotide that is involved, but if the different alleles of the SNP are known to cause different amino acids to be incorporated into the protein, the SNP may be named according to the amino acid substitution that is involved. For example, one of the genetic factors that influence an individual's susceptibility to hereditary hemochromatosis is a SNP in the human hemochromatosis (*HFE*) gene. Some people have a G at nucleotide number 845 in the *HFE* gene's coding sequence, while others have an A in that position. At the nucleotide level, this SNP can be symbolized as G845A, or n845G>A, using the letter "n" to indicate that we are describing the SNP at the nucleotide, rather than amino acid, level.

The G allele of the *HFE* gene causes the amino acid cysteine to be incorporated as the 282nd amino acid in the protein (recall that three nucleotides are read from the coding sequence for each amino acid that gets incorporated into the protein, so nucleotide 845 is part of codon 282). The A allele, on the other hand, causes the amino acid tyrosine to be incorporated in that position instead. At the amino acid level, this SNP can be abbreviated as C282Y or cys282tyr, using the single-letter or three-letter abbreviations for the amino acids cysteine and tyrosine, respectively. At the amino acid level, a sequence change that causes one amino acid to be substituted for another is referred to as a *missense variant*, or sometimes *missense mutation*.[2]

The cys282tyr missense variant is a good candidate for a functional polymorphism. Recall that a protein's ability to perform its function depends on the protein being able to adopt the proper three dimensional shape, and to move as it needs to as it performs its function. Recall also that each amino acid contains an organic chemical group that gives it its unique size, electrical charge and polarity. Cysteine's organic group contains a terminal sulfhydryl (SH) group. The HFE protein's shape is maintained in part by a disulfide bond that is formed between the cysteine in position 282 and another cysteine in the HFE protein. Tyrosine's organic group, on the other hand, consists of a methylene (CH_2) group plus a hydroxylated 6-carbon benzene ring (phenol group). Because tyrosine does not have a sulfhydryl group, it cannot form the disulfide that helps maintain the HFE protein's shape; this reduces the level of activity in the HFE protein, and increases the individual's risk for hemochromatosis.

3.8 There Are Many Small Deletions and Insertions in Different People's DNA

There are a number of places in the human genome at which some people will possess a certain stretch of nucleotides, while others will not. These polymorphisms are referred to as *insertion/deletion polymorphisms* (often abbreviated as "indel"

[2] Note that some authors use the term "mutation" to describe any change in the DNA sequence, regardless of whether it has any influence on the activity of any protein(s).

or "insdel"). For example, the panel of polymorphisms that are tested in order to estimate an individual's susceptibility to breast cancer includes a dinucleotide ins/del polymorphism (185insdelAG) and a mononucleotide insdel polymorphism (5382insdelC) in the *BRCA1* gene.

Recall from Sect. 1.3.4 that the ribosome reads the mRNA in 3-nucleotide codons, and each codon directs the ribosome to add one amino acid into the growing polypeptide chain. Deletions and insertions in a gene's coding sequence that involve an even multiple of three nucleotides will often result in the deletion or insertion of one or more amino acids from the protein, but will leave the amino acid sequence of the protein normal both before and after the deletion/insertion. Depending on how many amino acids are deleted or added, and exactly where in the protein the deletion/addition occurs, this may or may not change the level of activity in the protein.

Deletions and insertions that lie in the coding sequence of a gene and involve a number of nucleotides that is not evenly divisible by three (such as the 185delAG and 5382insC alleles in the *BRCA1* gene) shift the ribosome's reading frame, and cause the ribosome to read a completely different set of codons. The ribosome will add a certain number of novel amino acids into the polypeptide chain, and sooner or later encounter a STOP codon that causes it to cease adding amino acids and release the polypeptide. In most cases, this will produce a nonfunctional protein.

In addition to coding-sequence indels, there are a number of genes that have functional indel polymorphisms in the sequences that regulate their level of activity. As discussed in Sect. 1.3.1, transcription factor proteins bind to a gene's promoter sequences and turn the activity of the gene up and down as your body's needs change. Variations in the critical regulatory sequences often change the efficiency with which the promoter sequences bind the transcription factor proteins, and change the gene's activity level.

3.9 Repeated Sequence Length Polymorphisms and Microsatellite Analysis

3.9.1 The Repeated Sequence Motif Can Vary in Length

The human DNA contains a large variety of repeated sequences, ranging from mononucleotide repeats (ex. CCCCCCCC) to repetitions of a 1,000 bp motif. There are a few repeated sequence length polymorphisms in coding regions of genes, but most of the repeated sequence length polymorphisms that have been identified as risk-influencing genetic factors lie in the promoter region of their gene.

For example, the FDA has recommended determining a cancer patient's status for polymorphisms in the gene that encodes the enzyme uridine diphosphoglucuronosyl transferase 1A1 (*UGT1A1*) before deciding whether to prescribe irinotecan. *UGT1A1* has a "TATA box" promoter sequence, in which most people have

the sequence $A(TA)_6TAA$ (i.e. they have six repetitions of the TA repeat), but some people have the sequence $A(TA)_7TAA$. Possessing the 7-repeat allele of this polymorphism increases the risk for ADRs after irinotecan administration.

Although most of the repeat length polymorphisms that are known to contribute to multifactorial disorders involve short repeated sequences like the *UGT1A1* dinucleotide repeat, several known functional polymorphisms involve larger repeated sequences. For example, there is an 86 bp repeat polymorphism in intron 2 of the interleukin-1-receptor antagonist (*IL-1RA*) gene, which encodes a protein that inhibits the inflammatory response. There are five known alleles, ranging from 2 to 6 repeats of the 86 bp sequence. Most people have four repeats of the 86 bp motif at this locus, but some people have only two (this allele is referred to as *the IL1RN*2* allele). The *IL1RN*2* allele reduces the level of expression of the IL-1RA protein. The *IL1RN*2* allele is associated with a poor prognosis in individuals with several chronic inflammatory diseases, including systemic lupus erythematosus, ulcerative colitis and alopecia areata. It has also been reported to be associated with diabetic nephropathy, a chronic hypochlorhydric response to Helicobacter pylori infection and an increased risk for gastric cancer.

Trinucleotide repeats have the interesting property of being unstable during meiosis. They can expand during meiosis, whereupon the members of each succeeding generation will have increasingly longer strings of the repeated sequence. In many cases, this leads to the clinical phenomenon of anticipation, in which members of each succeeding generation are affected earlier in life and more severely than members of previous generations. Trinucleotide repeat testing is performed for several disorders, including Fragile X syndrome, Huntington disease, myotonic dystrophy, Friedrich's ataxia, dentatorubropallidoluysian atrophy, spinobulbar muscular atrophy (Kennedy disease) and several forms of spinocerebellar ataxia. For each disorder, there is a threshold number of repeats of the trinucleotide motif at which the risk for meiotic instability, and therefore rapid expansion from generation to generation, increases.

3.9.2 Microsatellite Instability (MSI) Is Observed in Certain Types of Cancer

Short tandem repeats such as dinucleotide, trinucleotide and tetranucleotide repeats are often referred to as *microsatellites*. The presence of the microsatellite in the DNA sequence can cause the DNA polymerase to slip during DNA replication, and add more or fewer repetitions of the repeated sequence to the newly synthesized DNA strand than it is supposed to. If left uncorrected, this would causes the length of our microsatellite repeats to increase or decrease as we passed them from one generation to the next. Our DNA repair proteins usually correct these DNA replication errors, however. When our DNA repair mechanisms are working properly, microsatellites are stable during meiosis, and rarely change in

length from one generation to the next (note that trinucleotides are the exception to this, see Sect. 3.9.1).

When our DNA repair mechanisms are not working properly, however, microsatellites will be unstable, not only during meiosis, but during the DNA replication that is involved in mitosis as well. Some have suggested that disruption of the ability to repair genetic abnormalities is an essential step in the progression from the pre-cancerous state to cancer. Mutations in the genes whose proteins are involved in DNA repair can therefore constitute important somatic mutations that contribute to the development of certain types of cancers. The presence of MSI in malignant cells indicates that the ability to repair damaged DNA has been compromised in the tissue being investigated.

The presence of MSI is detected simply by determining the size of the repeated sequence, as described in Sect. 3.19.4. When MSI is observed in a patient's malignant cells, it is often wise to perform a mutation analysis to search for mutations in genes whose proteins help repair damaged DNA, and referral for genetic counseling is appropriate.

3.9.3 Loss of Heterozygosity (LOH) of Microsatellites Is Seen in Several Disorders

Microsatellites also serve to indicate other important mechanisms for disease. Microsatellites are often highly polymorphic; in fact, many individuals are heterozygous for many microsatellite polymorphisms. SNP array analyses have revealed that patients with disorders as diverse as cancer, mental retardation and blindness have sometimes long (perhaps several Mb) stretches of contiguous microsatellites for which the affected cells have homozygous genotypes, but the individual's normal cells from another tissue have heterozygous genotypes. This LOH may result from a simple deletion, reduplication after deletion, recombination during mitosis or other mechanisms.

LOH in one or more microsatellites may indicate a deletion that has removed one copy of a gene that lies close to the microsatellite(s). When it comes to influencing the individual's phenotype, some gene alleles are dominant over others; the recessive allele would have relatively low penetrance, because the dominant allele would control the individual's phenotype. If an individual had one allele that was dominant over the other for a particular gene, and the deletion removed the dominant gene allele, this would allow the normally recessive or low-penetrance allele that remained to influence the individual's phenotype more strongly than it usually does. When one finds LOH in one or more microsatellites in a patient who has cancer, it may indicate that the patient has a deletion in his/her DNA, and that the deleted region contains a gene that in turn contains a polymorphism that influences risk for that disease. If the risk-increasing gene allele can be detected, this can provide a means to screen other family members to better estimate their risk for the disorder.

3.10 Chromosome Rearrangements Can Contribute to Some Complex Disorders

It is well known that chromosome rearrangements such as translocations, inversions, duplications and deletions can have sufficiently deleterious effects on the individual's metabolism to causes disorders on their own. Recent studies have demonstrated, however, that chromosome rearrangements are much more common than we once thought, and that they are capable of contributing to the development or progression of certain complex multifactorial disorders as well.

Chromosome rearrangements are especially important in cancer, as the progression to malignancy is often accompanied by several chromosome rearrangements in transformed cells. For example, it is common to determine whether a patient with chronic myelogenous leukemia, acute lymphoblastic leukemia or acute myelogenous leukemia possesses a specific translocation between chromosomes 9 and 22 [t(9;22)(q34;q11)], commonly called the *Philadelphia chromosome*, before deciding whether to prescribe that patient tyrosine kinase inhibitors such as Dasatinib.

3.11 Copy Number Variation Is Surprisingly Frequent

It is generally assumed that we have two copies of each gene (except males' X and Y chromosome genes), but recent research has shown that many people have deletions and duplications of portions of their DNA that result in them having greater than or less than two copies of certain genes (generically referred to as *copy number variations*, or CNVs). One example that has wide clinical importance is the fact that a small percentage of people have duplications of a portion of chromosome 22 that result in them having 3–13 potentially functional copies of the cytochrome p450 2D6 (CYP2D6) gene (see Sect. 4.6). As described in Sect. 3.19.6, fluorescence *in situ* hybridization assays and microarray analyses are capable of detecting CNVs.

3.12 It Is Sometimes Necessary to Determine the Level of Activity in Specific Genes

When we think of gene mutations, we usually think of *loss-of-function mutations*, in which the activity of the protein is impaired. Some gene variants constitute *gain-of-function mutations,* however. As described above, there are a surprising number of individuals who possess duplications that result in extra functional copies of one or more genes. In addition, there are many situations in which we cannot predict the effect of a sequence variation on gene activity, and must

directly measure the gene's level of activity in order to determine the significance of the variant.

Determining the level of activity in a specific gene(s) can help classify a cancer at the molecular level and guide the choice of treatments. This may be determined indirectly by detecting extra copies of the gene in the patient's DNA, or directly by determining the level of the mRNA or protein in question in the appropriate cells (see Sects. 3.19.6 and 3.19.9).

3.13 Mitochondrial DNA Variants Are Also Relevant

Although the mitochondrion is best known as the cell's energy generator, it also has its own DNA molecule. While most personalized medicine tests assess the individual's status for polymorphisms in nuclear genes, polymorphisms in the individual's mitochondrial DNA may also influence his/her risk for disease or drug response. For example, between 1% and 15% of patients with cystic fibrosis who are given aminoglycoside antibiotics to treat *Pseudomonas aeruginosa* respiratory infections suffer some degree of hearing loss. Possessing the G allele of the mitochondrial A1555G polymorphism increases the patient's risk of suffering hearing loss after taking aminoglycosides.

3.14 Many Epigenetic Factors That Influence Gene Activity Are Amenable to Testing

As discussed in Sect. 1.7, epigenetic factors are an important influence over many genes' levels of activity. Epigenetic factors involve chemical modifications that do not alter the gene's base sequence, but influence the activity level of the gene instead. The two best-understood epigenetic modifications involve methylation of C nucleotides in CG-rich promoter regions and acetylation of specific lysine residues in the histone proteins.

Special polymerase chain reaction (PCR, Sect. 3.19.1) protocols already exist that can detect the degree of methylation that exists at any particular C locus. In addition, microarray-based assays (see Sect. 3.19.2) have been developed that assess methylation status at loci all across the genome. At some point in the near future the level of methylation at key sites in the DNA will be one of the epigenetic factors that gets incorporated into many predictive algorithms. In addition, researchers are discovering ways to influence the methylation patterns of genes, and thereby silence overactive genes or restore activity in a gene that has been silenced by aberrant methylation. In these cases, methylation-specific PCR assays will be needed to monitor the progress of therapy.

3.15 Some Tests Assess Characteristics of the Pathogen

In addition to determining the patient's genotypes, personalized medicine can also include tailoring drug therapy to the specific strain of virus or bacterium that is infecting the individual. For example, some strains of the HIV virus infect T cells by using the T cells' surface protein CCR5 to gain entrance. A blood test can be performed to determine whether the patient has been infected with CCR5-tropic HIV-1, and is therefore a candidate for "entry inhibitor" drugs such as maraviroc. DNA sequencing and/or PCR/restriction endonuclease assays (Sect. 3.19.3) can be used to determine the type of pathogen the patient has been infected with.

3.16 Cancer Analyses Often Must Include Somatic Mutations as well as Germline Mutations

Genetic analyses in patients with cancer often must include determining what genetic abnormalities have occurred in the malignant cells themselves. This often necessitates a search for *somatic mutations* in the malignant cells themselves. Somatic mutations are mutations that arise after fertilization, as the cells are replicating, dividing and differentiating into their individual cell types. Because somatic mutations arise after fertilization, they only exist in cells that have descended from the cell in which the mutation originally arose. A somatic mutation can drastically alter the metabolism of the cells in which it resides. Because cancer begins with the disruption of metabolism in a single cell, somatic mutations make significant contributions to the development and progression of many cancers. As a cancer progresses, the malignant cells often accumulate gene mutations and chromosome rearrangements that can be used not only to predict the individual's prognosis, but also monitor the individual's response to drugs.

3.17 Predictive Algorithms Must Include both Genetic and Nongenetic Factors

In order to derive the most accurate possible estimate of the individual's susceptibility to a disease, or the best prediction of the individual's response to a treatment, one must combine genetic information, clinical data, family history and information related to the individual's diet, environment and lifestyle. The genetic data must include polymorphisms from genes whose proteins influence as many aspects of the disease process or drug's actions as possible, including polymorphisms from genes whose proteins mediate the normal function of the pathways that are affected by the disease or the drug. Genetic information should be used to complement traditional indicators, not replace them, as predictors of disease susceptibility and drug response.

3.18 Genome-Wide Association (GWA) Studies Provide Insights into the Mechanisms for Disease, But Their Results Are Often Not Clinically Useful

3.18.1 Direct-to-Consumer GWA Testing Services May Provide Results That Have Limited Clinical Utility

To develop the most clinically useful tests, one must incorporate all the necessary genetic information into the predictive algorithms. There are a great many genetic variants that constitute risk-increasing alleles with low penetrance; together, they account for a significant portion of the clinically relevant variation in human DNA. In order to use genetic information as effectively as possible, researchers must incorporate all the genetic variants that contribute to a specific disease into the predictive algorithm, including the ones that have low penetrance. There is currently considerable effort being expended to identify these variants and incorporate them into clinically useful tests.

Association studies are used by researchers to identify risk-increasing gene alleles. Appropriately designed GWA studies can identify risk-increasing alleles that have low penetrance, and this is a key tool for researchers. This is only the beginning of the process whereby a clinically useful genetic test is produced. Unfortunately, however, when it is reported that an association has been found between a genetic variant and a disease, many lay people may underestimate the time it takes to translate that finding into a clinically useful test.

The genome-wide SNP screens that are advertised directly to consumers include many tests that are designed to detect risk-increasing alleles that have low penetrance. Critics of personalized medicine testing warn that people who have received the results from a genome-wide SNP screen, but who do not understand the concept of penetrance and the multifactorial nature of complex diseases, may become unnecessarily anxious upon finding out they possess a "risk allele." This is indeed a concern, and it highlights the importance of genetic counseling in the process, even when the process does not require an appointment with a physician.

As the cost of genetic testing declines, physicians and other health care practitioners will see an increasing number of patients who have obtained information about their genetic status, but do not know how to interpret that information within the framework of their health care plan. In this section, we discuss the GWA studies that provide the foundation for these genome-wide SNP screens, and provide background information that will help you explain the limitations of the results of these tests to patients.

3.18.2 GWA Studies Identify Risk-Increasing Alleles, But Have Their Limitations

When discussing single-gene disorders, it is expected that in virtually all cases the causative mutation and the disorder will coexist; everyone who has the disease will

have the causative mutation, and everyone who has one or two copies of the causative mutation (depending on whether the mutation is dominant or recessive) will have the disease. Because of the complex etiology of multifactorial disorders, however, the relationship between a risk-increasing allele and its associated disorder is never that consistent. Identifying a risk-increasing gene allele involves determining the degree of *association* between specific alleles of the gene in question and the disorder or ADR in question.

Most association studies (often called *case-control studies*) include a group of people who have the disorder (the affected cases) or show a particular response to a treatment, and a group of people who do not have the disorder (control subjects) or show a different response to the treatment. The researchers determine which alleles the subjects possess of the genes of interest, and determine whether any specific gene allele appears more frequently in one group than it does in the other. If the gene/protein in question actually does influence one's susceptibility to that disease/ADR, and the test is providing accurate information regarding the status of a functional polymorphism, there should be one or more risk-increasing alleles that appear more frequently in the case group than in the controls, and/or one or more risk-reducing alleles that appear more frequently in the controls than in the cases.

Researchers have found GWA studies highly useful tools for detecting risk-increasing gene alleles. GWA studies have identified over 300 polymorphisms that influence the individual's risk for more than 80 complex diseases (summarized at www.genome.gov/26525384). By identifying the genes whose proteins influence susceptibility to specific diseases, GWA studies have enhanced our understanding of the pathophysiology of many complex diseases, including several types of cancer, cardiovascular disease, inflammatory bowel disease, diabetes, respiratory diseases and several neurodegenerative disorders.

GWA studies have their limitations, however. For example, the association between a risk-increasing gene allele and the disorder will be weakened by any factor that reduces the penetrance of the risk-increasing allele (see Sect. 3.3). Because the penetrance of so many risk-increasing alleles is so low, few published studies include the large population of subjects that is necessary in order to demonstrate a statistically reliable association between a risk-increasing allele that has low penetrance and the disorder. In addition, a rare allele with high penetrance may not be identified in an association study, because the risk-increasing allele may not be observed frequently enough in the case population for the statistical analysis to support the claim that the allele was found significantly more frequently in the case population than it was in the controls. Factors such as these undoubtedly contribute to many of the discrepancies that can be found among the findings of different research studies in this field.

One of the most important limitations of the association studies that are currently being conducted is that they only look at a small portion of the relevant gene's sequence. These tests often assess the individual's status for one or a few specific polymorphic sites in the gene's sequence. If an association is found between the disease and a specific allele of a specific polymorphism, that test may be incorporated into a personalized medicine test.

As more and more genomes are sequenced, however, it is becoming apparent that the human DNA sequence is so variable that there are many rare variants, and many that have not yet been reported in the literature, that influence different individuals' risks for these diseases. In addition, as discussed in Sect. 3.18.4, this is a particularly important concern for tests that assess the individual's status at a site that is not believed to be a functional polymorphism, but instead is believed to lie close to a functional polymorphism.

Because GWA studies focus on specific sites in the sequence, they are not capable of detecting any risk-increasing alleles other than the ones they were specifically designed to detect. This means that personalized medicine tests that focus on single sites in the DNA will detect only a small subset of the risk-increasing alleles that are present in the human population. The fact that no deleterious mutations were found in a test that assesses several risk-influencing polymorphisms from the same gene is by no means equivalent to saying that a disruption of that gene's function is not present and increasing that person's risk for the disease. The individual may have another sequence variant in that gene that is increasing his/her risk for that disease.

As the cost of DNA sequencing declines, sequencing will come to replace these site-specific tests as the method of choice for whole-genome screens. Sequencing detects all variants that exist in the DNA, even those that have never been reported before. As a result, it is capable of identifying not only novel mutations in known genes, but also novel genes whose disruption can contribute to the disease. One example of the use of whole-genome sequencing to discover disease-associated sequence variations is the 1,000 Genomes Project (www.1000genomes.org). The 1,000 Genomes Project aims to sequence the entire genome of 1,000 volunteers and track their health information over their lifetime. By providing complete sequence information as well as extensive health information, on a subject population of this size, the 1,000 Genomes Project will overcome many of the limitations of the GWA studies that have been conducted in the past.

3.18.3 Measures of the Association Between the Risk-Influencing Allele and the Disorder/ADR

The association between the disorder/ADR and either a single allele or a genotype is often expressed in terms of the *odds ratio* (*OR*) associated with that allele or genotype. The odds ratio states how much more likely the allele or genotype is to be found in an affected person versus an unaffected person. For a biallelic A/G SNP, saying that the A allele has an OR of 1.3 means that the A allele is 1.3 times more likely to be carried by an affected individual than the G allele is. When there are more than two possible genotypes or alleles at a locus, a study may report the OR of one genotype or allele versus another specific genotype or allele, or the OR of one genotype or allele against all other genotypes or alleles. Because of the multifactorial nature of these common diseases and the low penetrance of most

risk-increasing alleles, the OR for a typical risk-increasing allele or genotype is less than 2.0. Therefore, any time the OR for an allele or genotype exceeds 1.0, it is considered evidence that there is a real association between that allele or genotype and the disorder. The odds ratio is often used as an approximation of relative risk although, strictly speaking, they are not identical.

The *relative risk* (*RR*) is also often reported as an indicator of a gene's contribution to a disease. The relative risk refers to the risk of disease or ADR in someone who possesses the risk-increasing allele versus someone who does not possess the risk-increasing allele. Assigning a risk-increasing allele a relative risk of 2.0 means that people who possess that allele are twice as likely to develop the disorder or exhibit the ADR as people who do not possess that allele.

Another term that is sometimes used to describe the degree to which a risk-increasing allele contributes to a disease is *attributable risk*. The attributable risk is calculated simply by subtracting the frequency of the disease in people without the risk-increasing allele from the frequency of the disease in people with the risk-increasing allele.

3.18.4 Many of the SNPs That Are Used in GWA Studies Are Not Themselves Functional Polymorphisms, But Are Linked to Functional Polymorphisms

DNA sequence polymorphisms are often referred to as *markers*, for several reasons. Researchers who discovered the sequence of the human DNA molecule and mapped the position of all the genes on their chromosomes began using the term in the same way cartographers do. Because the sequence of the entire human DNA molecule has been published, we now know the exact location of each of these polymorphic sites on its respective chromosome. If you know the location of a sequence on its chromosome, that sequence is capable of marking that position on the chromosome, the same way a mile marker marks position on a map. If a researcher who is trying to determine the location of a disease-related gene in the human genome sequences a long stretch of DNA and finds that it contains both a known marker and the gene of interest, this places the gene at that position on the respective chromosome's map.

Another reason why these polymorphisms are called markers is that one's status at a polymorphic locus can often be an indicator, or marker, that reflects some other aspect of the individual's status. For example, once researchers discover the genetic variants that increase an individual's risk for a specific disease, discovering that an individual possesses a risk-increasing allele marks that individual as someone who has an increased risk for that specific disease. The risk-increasing allele is often referred to as a marker for that particular disease, but given the limited penetrance that most risk-increasing alleles have, one should avoid using language that implies that possessing the risk-increasing allele necessarily means the individual will develop the disease.

Finally, a polymorphism can also serve as a marker when it reflects that individual's status for another, nearby polymorphism. As we discussed in Chap. 1 (see Sect. 1.2.4), when a marker and a gene lie far enough apart on the same chromosome, the recombination that occurs during meiosis will reshuffle the combinations of specific marker alleles and gene alleles that appear together on the chromosomes. Consequently, there will be a great many different combinations of specific marker alleles and specific gene alleles present together on the chromosomes that exist in the population. Knowing which alleles the individual has for the marker does not tell the analyst which specific gene alleles the individual has. In contrast, however, if the marker and the gene lie close enough together on the same chromosome that recombination happens very rarely between them, even though there may still be the same number of different marker alleles and gene alleles in the population, there will be many fewer combinations of the different specific marker and gene alleles on the chromosomes that are seen in the population.

When two sequences lie close enough to each other on their respective chromosome that they often stay together through meiosis and end up packaged together on the same chromosome in that individual's sperm or eggs, they are said to be *genetically linked*, or in *linkage disequilibrium*. If a marker is tightly linked to a polymorphic site within a gene's sequence, if a chromosome has one of the alleles of the marker on it, it is also likely to have a specific allele of the gene on it as well. Consequently, if you determine the individual's allele status at the linked marker, you are also likely to know what alleles the individual possesses for the functional polymorphism of interest.

The information from a linked marker may not always be completely reliable; recombination may occasionally occur between the marker and the functional polymorphism, even when they are tightly linked. Because there is always the possibility that genetic recombination can reshuffle allele combinations, a test that assesses the individual's status at even a tightly linked marker is always a less reliable source of information than a direct test of the functional polymorphism itself. When a marker is described as "linked to" a gene or a functional polymorphism, one must keep in mind that there is some probability that the result of the marker test will be deceiving with respect to the individual's status for the critical functional polymorphism. For this reason, tests that directly assess the presence or absence of the risk-increasing allele are used in a clinical setting. Linked markers, however, are important research tools, and are also included in the genome-wide SNP screens that commercial companies advertise directly to consumers.

The frequency with which recombination confounds the interpretation of these tests can be specified as the *recombination frequency* (*Rf*) between the linked marker and the critical gene/polymorphism. If a risk-increasing allele is detected at a marker that is linked to a critical gene polymorphism, unless you know the Rf between the tested marker and the critical polymorphism, you don't know how reliable that information is. The higher the Rf is between the marker and the polymorphism, the greater the probability is that the patient possesses the "risk-increasing" allele of the linked marker, but not the true risk-increasing allele of the critical gene polymorphism.

Another drawback to the use of linked markers involves the fact that the risk-increasing allele of the linked marker may be associated with a different allele of the functional polymorphism in different ethnic groups. Because the two polymorphisms are tightly linked, if different allele combinations lie together on the respective chromosome in different ethnic groups, these marker allele-gene allele associations will be maintained through the lack of recombination between the polymorphisms. Two studies that investigated a possible association between an allele of the linked marker and a disorder or drug response using subjects from different ethnic groups may produce directly contradictory results if the marker allele that accompanies the risk-increasing allele of the critical polymorphism in one ethnic group is associated with a typical-risk or risk-reducing allele of the functional polymorphism in the other ethnic group.

In addition to these long-recognized limitations, it has recently been suggested that one of the assumptions that is commonly made regarding the usefulness of linked markers is flawed. When one uses a linked marker for a genetic test, one usually assumes that the marker is linked to a functionally significant polymorphism that lies close to the marker; it is believed that the linked marker is providing information about a single critical polymorphism. When a research study reports an association between an allele of a linked marker and a disorder, it is assumed that, if you sequenced the DNA of the individuals who have the risk-increasing marker allele and the disorder, they would all have the same variation in a locus near the linked marker. Recent findings have suggested that it is possible, however, that different individuals may actually possess different sequence variants in the vicinity of the linked marker. Although the specific sequence variant that is associated with the linked marker's risk-increasing allele is different from one person to another, all these sequence variants alter the activity of the protein, and all therefore increase the individual's risk for the disease. The risk-increasing allele of the linked marker, therefore, is not associated with a specific critical polymorphism, but with different critical polymorphisms in different individuals. Once the cost of whole-genome sequencing declines to where it can be used routinely, many of the limitations that are inherent in the use of linked markers will be overcome.

3.18.5 Haplotype Blocks in the Human Genome Increase the Efficiency of GWA Studies

A *haplotype* is a set of genetic markers that lie close enough to each other that recombination very rarely occurs among them during meiosis. Recent research has shown that the human genome is arranged in *haplotype blocks*. A haplotype block is a stretch of contiguous markers among which there is little to no recombination; haplotype blocks are interspersed with regions in which recombination is more frequent. A haplotype block may contain several polymorphisms from within the same gene, or a set of polymorphisms that lie in two or even more neighboring genes, along with any polymorphisms that lies in the sequence(s) between the genes.

The only requirement that exists for two markers to be included in the same haplotype block is that there is very little recombination between them during meiosis. Note, however, that the term "haplotype" is occasionally used to refer to testing panels that include markers from different genes that lie on different chromosomes. These multigenic marker panels are extremely useful, especially if they include markers that indicate the functional status of several genes whose proteins work together in a single biochemical pathway. Technically speaking, however, these marker sets are not haplotypes; these testing panels are better referred to as "multigenic marker panels."

Haplotype analysis can be very useful when a family has a single-gene disorder running through it that is known to be caused by a mutation in a particular gene, but the causative mutation for this particular family's disorder has not yet been found. If one tests a haplotype of several markers that span the region within and immediately surrounding the gene, one can track the transmission of the risk-increasing haplotype (in this case, the mutation-bearing, or disease haplotype) through the family. Because it is a single-gene disorder, often the affected family members will all have inherited the same risk-increasing haplotype, with no recombination happening within the region bounded by the tested markers. One might also see recombination within the haplotype region in a rare individual; observing whether that individual is affected with the disorder or not may help localize the causative mutation within the gene.

The fact that the human genome is arranged in haplotype blocks enables one SNP (or information about any one nucleotide in the haplotype block) to provide information regarding all the other nucleotides in its haplotype block. This has the potential to greatly increase the efficiency of a whole-genome screen. The International HapMap Project[3] represents an international effort to develop a panel of *tagging SNPs* that contains at least one SNP that lies within each of the haplotype blocks that exist in the human DNA sequence. Because each of the haplotype blocks is represented in the set of tagging SNPs, a properly selected set of tagging SNPs will theoretically provide information on the individual's entire DNA sequence. It has been estimated that the entire human genome can be covered by testing 300,000–600,000 properly selected tagging SNPs. This is well within the capacity of current microarray technologies (see Sect. 3.19.2).

3.18.6 Internet Resources That Summarize Findings from GWA Studies

There are several Internet resources available to those who want to research the genetic association studies that have been conducted. The National Human Genome Research Institute maintains a Catalog of Published Genome-Wide Association

[3] HapMap is an abbreviation for haplotype mapping.

Studies on its website (http://www.genome.gov/gwastudies/). You can search this database by disease, either by putting in your own search term or choosing from a list, or by gene, chromosome region or specific polymorphism. In addition to the association studies, the website also has a link to recent publications that discuss various aspects of GWA studies in general.

Similarly, the HuGE Navigator (http://www.hugenavigator.net/) maintains a continuously updated database of genetic association studies, and provides links to web pages that are devoted to particular genes and disorders. In addition to these resources, the CDC's National Office of Public Health Genomics developed the Genetic Association Information Network (GAIN), which conducted three reviews of GWA studies, and has made its findings accessible to researchers by entering them into the National Library of Medicine's database of Genotype and Phenotype (dbGaP; http://www.ncbi.nlm.nih.gov/gap).

In addition to these sources, the National Human Genome Research Institute sponsors the Genes, Environment and Health Initiative (GEI; http://www.gei.nih.gov/). The GEI not only conducts GWA studies, but also includes an Exposure Biology Program that aims to develop better methods for monitoring environmental exposures that contribute to disease.

3.19 A Brief Introduction to the Most Important Technological Advances

3.19.1 The Polymerase Chain Reaction (PCR) Allows One to Isolate the Sequence of Interest; Many Assays Begin with the PCR

It is not necessary for you to understand the details of the mechanism whereby PCR works to understand its usefulness. It is enough to understand that the PCR acts like a molecular photocopier, and makes several hundred thousand to several million copies of any stretch of the DNA sequence the analyst desires (referred to as *amplification* of the sequence). The PCR's unique usefulness lies in its ability to amplify any stretch of sequence the analyst wants to analyze, allowing the target sequence to be the predominant molecule in the PCR product, and preventing the other, non-target sequences in the organism's DNA from interfering with the analysis. The analyst designs a pair of primers, which are single-stranded oligonucleotides approximately 20–25 bp long, with sequences that allow each primer to bind to a stretch of the DNA on one side of the target sequence. The PCR replicates the DNA that lies between the primers. Therefore, the analyst's choice of primer sequences determines the specific stretch of the DNA that will be amplified. The PCR is used as the first step in many genetic analyses; it provides material for direct sequencing, enzymatic digestions, array hybridizations and other mutation-detection techniques.

3.19.2 Microarray Analyses Have Greatly Accelerated the Pace of Discovery

One technology that has greatly accelerated the rate of discovery is the microarray. In a microarray, hundreds of thousands of probe sequences are micro-spotted onto a glass slide. The probes can be pieces of DNA of varying sizes, depending on the specific goal of the assay. For example, the microarrays that are used for GWA studies use oligonucleotides.[4] The oligonucleotides are single-stranded, and can therefore bind to other single strands of DNA whose sequences are complementary to theirs. Once the oligonucleotides have been spotted onto the slide, the slide is incubated with the individual's DNA.[5] The individual's DNA has also been denatured, and will therefore bind to all the oligonucleotides whose sequences are complementary to the sequences present in the individual's DNA. By spotting oligonucleotides that contain all the possible nucleotide sequences for a functional polymorphism, the analyst can determine the individual's status for the polymorphism simply by determining which oligonucleotides were bound by the individual's DNA.[6] Because hundreds of thousands of oligonucleotides can be spotted onto a single slide (sometimes called a DNA chip or a gene chip), the analyst can determine the individual's status for thousands of different polymorphisms in one assay. Microarray capacity is being expanded; one microarray that is being beta-tested at present contains 2.7 million probes.

Given the sheer amount of information they can provide, microarray studies have ushered in an explosion of genetic and genomic information. Even more illuminating, gene expression microarrays have been developed that can indicate which genes are active and inactive in the tissue that is being tested. These arrays are critical for genomic and metabolomic studies, in which researchers determine which sets of genes and metabolic pathways get turned on and off by specific drugs or nutritional regimens, observe which metabolic pathways get turned on and off in malignant cells as a cancer progresses, or identify subpopulations of individuals within a particular disorder who share some critical metabolic disturbance. Some microarrays have even been adapted for quantitative analyses. They can provide the analyst an estimate of a specific gene's level of activity, or of the number of copies an individual has of a specific gene.

The first microarray to be approved by the FDA for clinical use is the AmpliChip CYP450 microarray (Roche Molecular Diagnostics, Alameda, CA, USA).

---

[4] Oligonucleotides are single-stranded stretches of DNA, ranging in length from approximately 17 to approximately 60 nucleotides long, depending on the specific application for which they are used.

[5] The individual's DNA is first prepared using one of several processes, each of which includes using the PCR to amplify either the entire genome or selected portions of it.

[6] The assay uses fluorescent labeling to indicate the oligonucleotides to which the DNA has hybridized.

The cytochrome P450 isoenzymes are a family of heme-containing enzymes that metabolize a wide variety of commonly prescribed drugs. The AmpliChip 450 determines the individual's status for several functional polymorphisms in two of the cytochrome P 450 proteins, cytochrome p450 2D6 and 2C19 (CYP2D6 and CYP2C19, respectively). Several of these alleles produce protein isoforms that work at a lower level of activity than most of the other isoforms that are present in the population do, and a few of them are null alleles that fail to produce any functional protein. These assays are particularly useful for identifying "poor metabolizers," who accumulate excessively high levels of certain drugs in their blood, and have a correspondingly increased risk for ADRs after being given those drugs (see Sect. 4.6).

In studies that compare the individual's CYP450 allele status against actual drug clearance data, it is clear that other factors besides one's status for the CYP2D6 and CYP2C19 genes control the clearance of many of these drugs. The presence of the null alleles successfully identifies virtually all the poor metabolizers, so the individuals who are at greatest risk for ADRs are identified reliably by the AmpliChip assay. In individuals who have typical-activity isoforms and isoforms that have slightly decreased levels of activity (the extensive metabolizers and intermediate metabolizers), however, the individual's status for the CYP gene alleles does not reliably correlate with the drug clearance data.

3.19.3 Sequencing Provides Maximum Information, and Will Revolutionize Clinical Diagnostics

Dideoxysequencing is often referred to as the "gold standard" for DNA analytical techniques, because sequencing the DNA nucleotide-by-nucleotide provides the maximum amount of information possible, and can detect variations that have not yet been discovered, even those that are unique to that patient. New generation technology is bringing the cost of sequencing down to a level once thought impossible. At the present time, full genome sequencing is still too costly to be used for all medical genomic tests. The cost of sequencing is declining rapidly, however, and several commercial laboratories are planning to use high-throughput sequencing for diagnostic testing soon.

It is common to use PCR and dideoxysequencing to determine the individual's status for mtDNA polymorphisms, or to determine which bacterium or virus the individual has been affected by. For the mtDNA analysis, the analyst simply amplifies a stretch of mtDNA surrounding the polymorphism, and sequences it. To identify a virus, one uses PCR primers that will only bind to sequences that are present in that virus' DNA. To identify a bacterium, the analyst usually amplifies a stretch of sequence from a gene than encodes one of the subunits of the ribosomal RNA, and either sequences it or subjects it to analysis by restriction endonuclease digestion (see Sect. 3.19.3).

3.19.4 There Are Several Techniques Commonly Used to Assess Status for a SNP

There are several techniques by which one can determine the status of a SNP. The choice usually depends on how focused the study in question is. Because they need to assess hundreds of thousands of SNPs, GWA studies require microarrays, as described in Sect. 3.19.8. If the study focuses on a single SNP or a few SNPs, the analyst can generate a PCR product that contains the SNP and then sequence it, or use one of several other PCR-based assays to determine which of the possible sequence alleles is(are) present.

Some tests use specialized PCR protocols that use two sets of primers, each of which will only amplify one of the two alleles of that gene that are known to exist, and a means to detect which set of primers amplified the DNA successfully. Another modification of the PCR that can be used for SNP analysis is the primer extension assay, in which a short single-stranded DNA primer is used that anneals to the DNA up to the nucleotide right next to the SNP. DNA polymerase is then allowed to incorporate one more nucleotide after the primer, and the analyst determines which of the four nucleotides (each labeled with a different dye) was incorporated by the DNA polymerase.

Restriction endonucleases (*REs*) can also be used in conjunction with the PCR to determine the status of a SNP. REs are enzymes that cut DNA. Each RE has a recognition sequence, and cuts the DNA everywhere it finds that sequence. If a SNP either introduces a cut site that did not exist before, or abolishes a cut site that was present in the wild-type sequence, the ability of the RE to cut the DNA will be altered. For example, consider the wild-type sequence and variant sequence below (only one strand's sequence is shown in each case):

wild-type = …AATGAC**T**ACGTACTGC…
variant = …AATGAC**G**ACGTACTGC…

To determine the individual's status for this SNP, the analyst could first use the PCR to amplify a stretch of the DNA that contains the SNP. For example, the analyst could generate a 250 bp-long[7] PCR product that has the SNP at position 100 in the 250 bp. The PCR product could then be digested by an RE that has the recognition sequence ACTA, and will cut the DNA between the C and the T inside any ACTA it finds. Because only the wild-type sequence contains the ACTA recognition sequence, the RE will only cut the wild-type sequence.[8] If the DNA contains only the wild-type sequence, the analyst will see fragments that are 100 bp and 150 bp in length after the PCR product is digested by the endonuclease. If the DNA contains only the variant sequence, the PCR product will not be cut by the endonuclease,

[7] The choice of PCR product size for this example is arbitrary; many different-sized PCR products could be generated in the first step of this analysis.

[8] We are assuming there are no other ACTA sequences in the 250 bp PCR product.

and a 250 bp fragment will be obtained. If the individual has the heterozygous genotype, and therefore possesses both wild-type and variant sequence, the analyst will obtain fragments of 100, 150 and 250 bp in length from that sample.[9]

3.19.5 PCR Simplifies Analysis of Indels and Length Polymorphisms and Detection of MSI and LOH

Determining whether the insertion or deletion allele is present at an indel polymorphism or detecting differences in the length of a repeated sequence is often simply a matter of amplifying the stretch of DNA surrounding the marker and observing the size of the PCR product. For a 63 bp indel polymorphism, for example, the insertion allele would produce a PCR product that was 63 bp longer than the PCR product produced by the deletion allele would be. Similarly, if two alleles of the gene differ by having a different number of repetitions of the repeated sequence, the size of that PCR product will vary accordingly. For example, if amplifying the four-repeat allele of the 86 bp repeat IL-1RA polymorphism produced a 600 bp PCR product, amplifying the two-repeat allele would produce a 428 bp PCR product, because the amplified sequence would be missing two repetitions of the 86 bp motif (172 bp).

MSI and LOH analyses are also performed by simply amplifying the region surrounding the microsatellite by PCR, then comparing the sizes of the PCR products obtained from malignant cells versus healthy cells from the same patient.

3.19.6 Hybridization Techniques Detect CNVs

There are a number of sequence-detection techniques that capitalize on the fact that complementary DNA sequences will bind, or hybridize, to each other (see Sect. 1.2.1). There are several ways in which this can be exploited to detect CNVs, either in single genes or large chromosome regions.

Comparative genomic hybridization (*CGH*) allows the analyst to detect CNVs throughout the genome. It can therefore detect the deletion or amplification of genes that were not previously known to be involved in that disorder, and is particularly useful for detecting multiple genomic changes in malignant cells. The original CGH techniques were chromosome-based. DNA was extracted from tumor cells and from the individual's normal tissues, and each was labeled with a different

[9] We do not include descriptions of these technologies, but there are several manual and automated methods for detecting the size of a PCR product or other DNA fragment. Most laboratories use capillary electrophoresis using specialized polymers; these machines can differentiate between fragments that differ by a single bp.

fluorescent dye. A microscope slide that had the individual's chromosomes spread out upon it was incubated with the two differentially labeled genomic DNAs, and the analyst observed the intensity of fluorescence from the two dyes in different chromosome regions.

Modern CGH analyses use microarrays (arrayCGH), on which millions of probes are immobilized. The use of small probes such as oligonucleotides (often 40–60 nucleotides long) provides a fine-resolution comparison, as well as easy coverage of the entire genome. The analyst can choose an array of probes that cover the entire genome, or focus on a specific chromosome, a defined chromosome region, a group of genes of interest or a combination of known disease genes and chromosome regions for which an association has been reported between a marker allele and a disorder. At the present time, SNP arrays are used most frequently for clinical array-based CGH analyses.

The interpretation of CGH data is complicated by the fact that many CNVs seem to be benign polymorphisms, or to have low penetrance as risk-increasing alleles. For example, one array-based CGH study reported finding CNVs in 97% of patients with unexplained mental retardation, but also reported that the majority of these CNVs had been inherited from phenotypically normal parents. Some have estimated that the average unaffected individual has 12 CNVs in his/her genome. Because of the many nonpathogenic CNVs that will be detected by whole-genome array-based hybridization, it may be more clinically useful to use arrays that focus on a group of genes whose proteins participate in pathways related to the disease or the drug's action.

Fluorescence in situ hybridization (*FISH*) has largely been replaced by CGH, but can still be used to detect CNVs in single genes. For a FISH assay, one cultures the cells of interest, using special culture medium additives such as colcemid to halt the cell cycle in metaphase, when the chromosomes are most condensed and easiest to see under a microscope. When the analyst drops the cultured cells onto a glass slide, the cells burst, and the chromosomes splay out into what is called a metaphase spread. The slide can then be immersed in solutions that denature the chromosomal DNA (separate the two strands and expose the bases).

The FISH technique takes advantage of the fact that single-stranded nucleic acids will readily bind to complementary sequences. The glass slide with the single-stranded chromosome DNA on it is incubated with a probe, which is a DNA fragment that contains a portion of the sequence of the gene of interest. The probe is tagged with a fluorescent molecule, then denatured, so it is also single-stranded. When the single-stranded probe finds a complementary single-stranded sequence on any of the chromosomes, it binds, or hybridizes, to the sequence. When viewed under the microscope, using the appropriate wavelength of light, the bound probe will fluoresce, and the analyst will see a bright spot on the chromosome. The more copies of the gene the individual possesses, the more spots will be seen on the slide.

Because one can use a probe that contains a unique stretch of sequence from a specific gene of interest, FISH is very effective for detecting specific gene amplifications. For example, FISH can be used to detect *HER-2* gene amplifications, which are associated with especially aggressive breast cancers.

3.19.7 Reverse Transcription-PCR Measures the Level of Activity in a Gene

One way of measuring the level of activity in a gene is to determine the amount of mRNA the gene is producing. The most common method for measuring mRNA output involves a combination of *reverse transcription* (*RT*) and the PCR method described above (*RT-PCR*). For RT-PCR, the analyst extracts the RNA from the tissue in question, then uses RT (catalyzed by the enzyme reverse transcriptase) to make a DNA copy of each of the mRNA molecules that are in the tissue's RNA pool. The reverse transcriptase is a DNA polymerase that reads an RNA strand as a template, and synthesizes a DNA strand that is complementary in sequence to the RNA template. Because each DNA molecule that is made by the RT has a sequence that is complementary to that of the mRNA molecule that served as the template for its synthesis, the products of RT are referred to as *complementary DNAs*, or *cDNAs*. The RT procedure can be designed to selectively transcribe the mRNA, but none of the other RNAs that are present in the cell (rRNA, tRNA). In that case, sequencing the cDNAs that are obtained from the cell should identify every gene that was active in the cell at the time the tissue was sampled.

Once the RT has generated the cDNA pool, the analyst then uses the PCR to amplify the cDNAs from specific genes. This can be used both to establish which genes are active in the body's different tissue types, as well as to illustrate disease processes or the way in which the body responds to a drug. For example, the analyst can compare tumor tissue against normal cells from that individual to see which genes' activities are upregulated or downregulated as part of the malignant transformation. In the research and development lab, a drug researcher can compare the same tissue before versus after drug administration to see which pathways are activated or inhibited in response to the drug.

Because many cancers involve upregulation of growth-related pathways, RT-PCR analysis of genes whose proteins are involved in cell cycle-related signal transduction pathways can help classify individuals with one type of cancer into subgroups for which different treatment strategies may be appropriate. For example, the Onco*type* DX® test (Genomic Health, Inc.) uses RT-PCR to assess the level of expression of 21 cancer-related genes (see Sect. 5.19.1), to classify female breast cancer patients according to whether or not they are likely to benefit from the drug herceptin™ (trastuzumab).

3.19.8 Immunohistochemical (IHC) Analyses Allow Direct Visualization of the Protein

Although it is not a genetic test, an IHC analysis is a genomic test, because it deals with the products of gene expression and how the gene's function is integrated into the function of the cell. IHC analyses capitalize on the fact that antibodies can be

created that will specifically recognize certain proteins, and that these antibodies can be labeled with fluorescent molecules, or enzymes that catalyze reactions that produce colored products, to facilitate visualization of the protein. Incubating the tissue sample with the labeled antibody allows the antibody to bind to its target. After that, the analyst examines the specimen under a microscope, and the fluorescence or the colored reaction product allows him/her to see exactly where the protein of interest is located in the tissue being examined, as well as determine the quantity of protein that is present.

Because of the antibody's specificity for its target protein, IHC analyses provide sensitive and specific assays that directly indicate the amount of the target protein that is present in the tissue, as well as its pattern of distribution throughout the tissue. For these reasons, IHC analyses of the estrogen receptor (ER) protein have replaced ER ligand binding assays as the assay of choice for predicting a premenopausal breast cancer patient's response to tamoxifen treatment. Similarly, IHC analysis of the mitogen-activated kinase MKK4 may predict survival rate in individuals with pancreatic cancer. IHC is also being used as a routine assay in some medical centers to identify individuals with colorectal or endometrial cancer who may have Lynch syndrome (see Sect. 5.14).

3.19.9 Expression Arrays Reveal Disease-Associated and Treatment-Associated Changes in Gene Expression

In addition to determining the individual's genotype, a microarray can also be used to measure the level of activity in a particular gene, or identify changes in gene activity that accompany drug treatment or the progression or remission of a disease. For an expression array analysis, the RNA is extracted from the tissue to be studied, and the RT-PCR technique described above is used to make cDNAs that correspond to the mRNAs that are present in the tissue at the time it was sampled. Because the mRNAs are used by the cell's ribosomes to make the cell's proteins, the mRNAs that are present in the cell reflect the genes that are active in that cell at that time. Because the microarray allows the analyst to spot hundreds of thousands of oligonucleotides on a single array, each array can illustrate the expression status of thousands of genes (referred to as the cell's expression profile).

These expression array investigations have been particularly useful in characterizing cancers at the molecular level, because malignant cells often undergo striking metabolic disturbances. Comparing the expression profiles of the malignant cells to normal cells from the same patient reveals valuable information regarding the metabolic changes that underlie the progression of the disease. Not all individuals with the same type of cancer demonstrate the same metabolic disturbances. Expression profile analyses can identify molecular subclasses within a single type of cancer, and thereby identify individuals who are likely to respond or not respond to certain therapies. In addition, comparing the expression profiles of malignant cells before and after administration of a drug can illustrate whether

the drug effectively reverses the metabolic alterations that occur during malignant transformation.

It is often useful to develop an algorithm that can predict a clinical outcome based on the observed changes in activity in a small subset of critical genes. For example, researchers investigating metabolic changes in non-small cell lung carcinomas (NSCLC) have developed a cDNA microarray system that can indicate activity status in 27,648 different genes. An NSCLC patient's response to gefitinib therapy can be predicted by analyzing the expression patterns of 12 genes, several of which encode proteins that participate in the epidermal growth factor receptor-mediated (EGFR) signal transduction pathways. Upregulation of the EGFR pathway, for example by over expression of the EGFR ligand amphiregulin, conveys resistance to gefitinib. These microarrays can identify the individuals who are likely to respond well to gefitinib, and have also identified several elements of the EGFR signaling pathway as potential targets for future drugs.

3.19.10 Some Epigenetic Factors Can Be Easily Assessed

Although we still have much to learn about how epigenetic factors regulate gene activity, it is clear that the level of methylation in certain C nucleotides in some genes' promoter regions can indicate the level of activity in that gene. In order to determine the level of methylation at a particular C nucleotide, a simple modification of the PCR is used. The patient's DNA is first treated with sodium bisulfite, which converts unmethylated Cs to Us, but leaves methylated Cs unaltered. The analyst then performs a PCR using two sets of primers, one of which will amplify the DNA if the C in question was methylated, and the other which will amplify the DNA if the C in question was unmethylated. There are several methods available to determine the relative yields of the two PCRs, and thereby estimate the level of methylation at that C nucleotide.

3.19.11 G-Banding Allows the Detection of Chromosome Rearrangements

While new technologies are being developed to detect all the other genetic risk factors that were discussed in this chapter, chromosome rearrangements are still effectively detected using *G-banding*, which was originally developed in the late 1960s and has experienced impressive improvements in resolution since. Briefly, metaphase chromosome spreads are prepared much as they are for FISH analyses (see Sect. 3.19.6), but then instead of being denatured and bound by probes, the chromosomes are digested with a protease, which digests some of the proteins that are associated with the DNA. The chromosomes are then stained with a dye called Giemsa, giving the

technique the name G-banding. G-banding stains each chromosome arm with a unique pattern of black, grey and white bands. A trained analyst reviews the chromosomes to determine if there are any portions missing, duplicated or rearranged.

Further Readings

Bruno DL, Ganesamoorthy D, Schoumans J, et al. (2009) Detection of cryptic pathogenic copy number variations and constitutional loss of heterozygosity using high resolution SNP microarray analysis in 117 patients referred for cytogenetic analysis and impact on clinical practice. J Med Genet Feb;46(2):123–31

Centers for Disease Control and Prevention. ACCE Project. www.cdc.gov/genomics/gtesting/ACCE/

Collins A, Tapper WJ (2011) Genome Variation: A Review of Web Resources. Methods Mol Biol 713:129–139

McGuire AL, Evans BJ, Caulfield T, Burke W (2010) Science and regulation. Regulating direct-to-consumer personal genome testing. Science 330(6001):181–2

Ruano Y, Mollejo M, de Lope AR, et al (2010) Microarray-based comparative genomic hybridization (array-CGH) as a useful tool for identifying genes involved in Glioblastoma (GB). Methods Mol Biol 653:35–45

Sudmant PH, Kitzman JO, Antonacci F, et al. (2010) Diversity of human copy number variation and multicopy genes. Science 330(6004):641–6

Wang K, Li M, Hakonarson H (2010) Analysing biological pathways in genome-wide association studies. Nat Rev Genet 11(12):843–54

Chapter 4
Toward the Safer and More Effective Use of Prescription Drugs: Pharmacogenetics

Abstract An individual's response to a drug is influenced by factors that regulate absorption of the drug, distribution of the drug throughout the body, the interaction between the drug and the drug's molecular targets, metabolism of the drug and excretion of the drug's metabolites. The specialized proteins that regulate these processes all have variable levels of activity in different individuals. At present, however, drugs are often prescribed in a disease-oriented manner; any patient who presents with that disease gets the standard first-line drug and dose. After that, the doctor engages in an iterative trial-and-error process to find the safest and most effective drug/dose combination for that patient. This process wastes time and money, may allow the disease to worsen, and sometimes puts the patient at risk for an adverse drug reaction. The pharmacogenetic tests that are currently available have significant limitations. As our understanding of genetics and genomics grows, however, the tests will improve. Ultimately, the predictive algorithms will include not only the presence or absence of certain gene variants, but also nongenetic factors and personal and clinical data. In addition, the algorithms will be modified to reflect the interactions that occur between genes, and between genetic factors and nongenetic factors. At present there are several gene variants that are known to influence the pharmacokinetics of many prescribed drugs. These tests, while limited, can help guide dosing strategies for some patients, thereby reducing the frequency of adverse drug reactions. Researchers are also identifying increasing numbers of gene variants that influence the pharmacodynamics of drugs.

Pharmacogenetics/pharmacogenomics may be a relatively new field, but its foundation dates back to the turn of the twentieth century. Archibald Garrod introduced the concept of individual differences in metabolism (which he referred to as "biochemical individuality") in a paper in which he demonstrated that some rare individuals exhibited alkaptonuria. These rare individuals excreted what was later found to be homogentisic acid, which is a byproduct of the metabolism of the amino acids tyrosine and phenylalanine, in their urine. Despite the fact that Garrod's concept was well received by the scientific community, the first application of this concept to medicine didn't appear until the late 1950s, when three

K.M. Sweet and R.C. Michaelis, *The Busy Physician's Guide To Genetics, Genomics and Personalized Medicine*, DOI 10.1007/978-94-007-1147-1_4,

independent groups published studies that illustrated individual differences in the metabolism and/or effects of several prescription drugs, and Friedrich Vogel coined the term "pharmacogenetics."

The principles of personalized medicine have already improved research and development practices, especially in the field of drug development. The list of drugs for which pharmacogenetics testing is either required, recommended or considered desirable by the FDA is growing rapidly. According to the *JAMA*, genetic testing will be used routinely to personalize drug prescriptions by 2020. Drug development programs now routinely design studies to take the genetic background of the subjects into consideration. Incorporating the principles of genetics into clinical trials has shortened several drug trials. In addition, it has enabled several drugs that were withdrawn from testing or from the market to be revived for use in an identified population, reducing the money that gets lost when research and development efforts fail to produce viable treatments.

4.1 Genetic Polymorphisms Affect Both the Pharmacokinetics and Pharmacodynamics of Many Prescription Drugs

There are five processes that influence an individual's response to a drug:

- absorption (from the gut or other mode of delivery)
- distribution throughout the body
- interaction with the drug's molecular targets (ex. receptor proteins, enzymes) and the resultant activity in signal transduction pathways and other effectors
- metabolism of the drug
- excretion or elimination of the drug or its metabolites

Each of these processes relies on specialized proteins, and all the relevant proteins have variable levels of activity in different individuals. The activity level of the transport proteins that absorb the drug from the gut and distribute it to the tissues, the metabolic enzymes that activate and degrade the drug, and the transport proteins that carry the drug's metabolites out of the body affect the pharmacokinetics of the drug. Variations in the levels of these proteins' activities cause different individuals to achieve different concentrations of the drug or its active metabolites (or toxic metabolites) after a standard dose. In contrast, the level of affinity of the target cells' receptor proteins for the drug and the activity level of the proteins that mediate the signal transduction pathways downstream from the drug's receptor protein will affect the drug's pharmacodynamics. Variations in these proteins will cause different patients to have quantitatively different responses to the drug, even if they achieve identical concentrations of the drug at the relevant target tissues.

Drug metabolism involves two major groups of chemical reactions. Although the two different groups of reactions are referred to as *Phase I and Phase II* reactions, not all drugs go through both phases of metabolism, and those that do go through both phases do not necessarily go through them in that order. Phase I reactions

include reactions such as oxidations, reductions and hydrolysis. These reactions degrade, and thereby inactivate, some drugs. For drugs for which the active compound is a metabolite of the parent drug, however, these reactions create the active metabolite from the parent compound. Phase II reactions, on the other hand, are *conjugation reactions*; they attach a chemical group to the parent compound. Conjugation reactions include methylation, acetylation, glucuronidation and sulfation, among others. Both Phase I and Phase II reactions serve to convert drugs from a lipophilic form, which is necessary to allow the drug to cross cell membranes, to a more hydrophilic form that can be excreted in the urine.

Having an atypical level of metabolism for a drug can involve different risks, depending on the drug. If one of the drug's metabolites is the active compound, a slow metabolizer may not accumulate a high enough blood level of the active compound for it to be effective. On the other hand, if the parent drug is the active compound, slow metabolism can predispose the individual to ADRs, because he/she may build up toxic levels of the drug after a standard dose. This is a particularly important concern for a drug with a low therapeutic index (a small difference between the therapeutic dose and the toxic dose). For example, patients with low-activity alleles of the dihydropyrimidine dehydrogenase gene may experience fatal CNS toxicity after taking fluorouracil, which has a relatively narrow therapeutic window.

Any panel of tests or multifactorial algorithm intended to predict an individual's response to a drug must include an assessment of the critical polymorphisms in proteins that participate in all five of the aforementioned processes. Further, one must know whether the high-activity and low-activity alleles the individual possesses at these polymorphic sites act to oppose each other, cooperate with each other, or interact synergistically with each other to influence the individual's response to the drug.

For example, imagine that an individual possesses a low-activity allele in a gene whose protein metabolizes the drug, as well as a low-activity allele of a gene whose protein participates in the signal transduction pathway that enables the drug to exert its effects on its target tissues. The low activity in the metabolic enzyme causes the individual to achieve greater plasma concentrations of the drug than most patients do, while the low activity in the signal transduction protein diminishes the effect the drug has on its target tissues. Because of this unfortunate combination of genetic variants, the individual will accumulate enough of the drug in his/her body to increase his/her risk for ADRs, but will not be able to reap the therapeutic benefits of the drug, because of the sluggish response of the critical signal transduction pathway to even high concentrations of the drug.

In contrast, imagine the patient has a low-activity isoform of the metabolic enzyme, but a high-activity isoform of the protein that participates in the signal transduction pathway. In this case, the combination of high drug concentration and strong response to the drug may cause the target tissues to have a much stronger response to the drug than expected. Whether this will result in the drug being especially therapeutic or especially dangerous for that patient depends on a number of factors, including the drug's therapeutic index.

There is considerably more information available on the effects of genetic polymorphisms on the pharmacokinetics of drugs than on their pharmacodynamics, primarily because it is easier to measure the concentration of a drug or a metabolite in an individual's blood than to measure the response of target tissues to the drug. Genetic polymorphisms may markedly affect the sensitivity of target proteins and tissues to their drugs, however, so any algorithm that seeks to predict a patient's response to a drug must include information on pharmacodynamically relevant functional polymorphisms in addition to pharmacokinetically relevant polymorphisms and nongenetic factors. In addition, it is often necessary to assess the individual's status for polymorphisms in genes whose proteins maintain normal function in the tissues and pathways that the drug interacts with.

4.2 Improving on the Disease-Oriented Approach to Prescribing Drugs

Surveys suggest that approximately 40–50% of patients who are prescribed a drug do not improve, or worse, suffer adverse drug reactions (ADRs). ADRs represent one of the leading causes of death and illness in the industrialized world. ADRs come in two general types: concentration-dependent and idiosyncratic. The likelihood of one of the concentration-dependent ADRs varies directly with the concentration of the drug that accumulates in the patient. The risk for concentration-dependent ADRs is therefore strongly influenced by gene polymorphisms that influence the pharmacokinetics of the drug. The most widely applicable example of this involves the family of *CYP450* genes, which encode the CYP450 enzymes that metabolize many of the drugs that are currently prescribed for a number of diseases.

Idiosyncratic ADRs, on the other hand, cannot be predicted by the concentration of drug the patient accumulates. They are rarer than the concentration-dependent ADRs, but often extremely serious. The risk of one of these ADRs is more strongly influenced by gene polymorphisms that affect the pharmacodynamics of the drug, or other aspects of the individual patient's biochemistry, than by pharmacokinetically relevant gene polymorphisms.

One of the most common examples of idiosyncratic ADRs involves drug-induced liver injury (DILI), which one American study suggests accounts for 20% of all patients who are admitted to a hospital with severe liver injury, as well as 50% of cases of acute liver failure (of which 75% required a liver transplant). Recent studies have suggested that the individual's status for polymorphisms in some of the *HLA* genes influences his/her risk for DILI, and further, that differences in *HLA* gene allele frequencies can explain the different incidences of specific ADRs in different ethnic groups.

For example, possession of the *DRB1*1501* allele increases the risk for DILI in patients who were given the antimicrobial agent co-amoxiclav, while possession of the *DRB1*0701* allele increases the risk for DILI after taking the anticoagulant ximelagatran. The *DRB1*0701* is more common among Europeans than

East Asians, consistent with the fact that the incidence of liver toxicity after ximelagatran is greater in Europeans than East Asians. In contrast, the incidence of liver toxicity after ticlopidine is greater in Japanese patients than in Europeans. This may be due to the higher frequency of the *HLA-A*3303* allele in Japanese people versus Europeans.

Some *HLA* gene alleles may be associated with more than one idiosyncratic ADR. For example, possession of the *HLA-B*5701* allele is associated both with an 80-fold increase in the risk of DILI in patients prescribed flucloxacillin, and an increased risk for immune system hypersensitivity reactions following the administration of abacavir. The FDA recommends testing for the *HLA-B*5701* allele in patients who are to be prescribed abacavir.

Immune system hypersensitivity reactions often manifest as skin rashes, which can be serious, but may also involve the liver, kidneys or other organs. For example, up to 10% of patients who are prescribed the anticonvulsant carbamazepine (CBZ) will experience skin rashes, including toxic epidermal necrolysis or the blistering skin reactions seen in Stevens–Johnson syndrome (SJS). Studies have reported a strong association between the *HLA-B*1502* allele and CBZ-induced SJS in Taiwanese and Thai patients. Other studies, which included European patients or Japanese patients, have failed to replicate this finding. This discrepancy may be due to the fact that the *HLA-B*1502* allele is considerably less frequent in these latter two populations. These studies may have included so few patients with the *HLA-B*1502* allele that the statistical tests that were used were unable to detect the association reliably.

Not all ADRs are associated with *HLA* genotypes. Some polymorphisms that affect the kinetics of drugs influence the risk for ADRs. For example, polymorphisms in the *SLCO1B1* gene, which encodes the organic anion transporter OATP1B1, influence the individual's risk for statin-induced myopathy. Other polymorphisms that influence the level of function in systems that maintain the drug's target tissues' normal functions affect the risk for ADRs. For example, polymorphisms in the *KCNE1* gene, which encodes the beta subunit of a cardiac potassium channel, influence the individual's risk for drug-induced prolonged QT syndrome.

It has been estimated that genetic factors account for between 20% and 95% of the variability in different drugs' metabolism and effects, and the data from clinical drug trials consistently illustrate the variability in drug-metabolizing phenotypes and drug-response phenotypes that exist in the human population. Unfortunately, what is known about the variability in drug responses has not yet been translated into rules of thumb the practicing physician can use. At best, it is known that some ethnic groups are more likely than others to experience ADRs in response to specific drugs. Even at that, there is enough variability within a single ethnic group that there are no rules of thumb that apply reliably enough to guide clinical practice.

For example, one reason why Asian patients have a tendency to require lower doses of tricyclic antidepressants (TCAs) than patients from other ethnic groups is that approximately 14–21% of Asians lack an active CYP2C19 protein, which plays a critical role in the Phase I metabolism of TCAs. In contrast, only approximately 3–4% of Caucasians and African American patients lack CYP2C19 activity.

Although 14–21% is a substantial percentage of patients to be at high risk for ADRs, the majority (79–86%) of Asian patients have some level of CYP2C19 activity, and that level of activity ranges from the same low to high levels as are seen in other ethnic groups. Beginning all Asian patients on a low dose of a TCA will avoid some ADRs, but will also result in many more patients failing to respond to the drug.

One of the greatest promises of personalized medicine, and the one that will touch more people than any of the others, is the ability to tailor one's choice of drug and dose to the individual. Because drug therapy is a pillar of modern medicine, applying the principles of personalized medicine to pharmacology will enable physicians in almost all specialties to maximize the benefits and minimize the risks for their patients. Fortunately, pharmacogenomics is one of the areas in which the most progress has been made, and in which there are tests already available that can personalize treatments, at least to some degree.

The pharmacogenomic tests that are currently available reflect our incomplete understanding of the genetic and nongenetic factors that influence a patient's response to a drug. In most cases, tests focus on one aspect of the response to the drug (ex. pharmacokinetics), and do not combine the genetic information with nongenetic factors, family history data or clinical data. For example, as discussed below, the CYP450 family of enzymes metabolizes many commonly prescribed drugs (Table 4.1), and CYP450 testing is the most widely applicable personalized medicine test available today. Although the predictive ability of these CYP450 tests is limited by their inability to differentiate between individuals with high and moderate rates of drug metabolism, testing CYP450 status can reliably detect patients who are slow metabolizers of certain drugs, and will have a greater risk for ADRs after standard doses than most patients have. Further, as described in Chap. 5 (see Sect. 5.19), genomic testing of malignant cells is currently guiding decisions regarding choices of drugs for patients with certain types of cancer.

These partial successes have clearly proven the principles behind personalized medicine, and illustrate the great potential benefits that can be derived from carefully designed pharmacogenomic tests. As the field matures, each succeeding generation of tests will include genetic variants that affect more aspects of the drug response (i.e. adding pharmacodynamically relevant polymorphisms to the known pharmacokinetically relevant ones), and combine genetic information with an increasingly longer list of nongenetic factors.

In time, properly designed pharmacogenomic tests will provide safer and more convenient methods for predicting the individual's drug response than the conventional tests do. For example genetic testing provides several advantages for classifying patients as slow versus fast drug metabolizers over the conventional method, which involves giving the patient a probe drug and monitoring the time course of appearance and disappearance of the drug and/or one of its metabolites in the blood.

For example, one common test involves monitoring the metabolic ratio between dextromethorphan (DEM) and its metabolite dextrorphan in the urine for 8 h after ingestion of a single 25 mg oral dose of DEM. In contrast, CYP450 testing is safer

Table 4.1 Drugs metabolized by CYP2D6, CYP2C19 and CYP2C9

	CYP2D6	CYP2C19	CYP2C9
Beta-blockers	Metoprolol carvedilol Timolol		
Class I antiarrhythmics	Flecainide Lidocaine Propafenone Encainide Mexiletine		
Tricyclic antidepressants	All TCADs	Amitriptyline Clomipramine Imipramine	Amitriptyline
Tetracyclic antidepressants	Mianserin		
SSRIs	Most – ex. fluoxetine paroxetine	Citalopram	
SNRIs	Venlafaxine		
MAO inhibitors		Moclobemide	
Opioids	Codeine, tramadol		
Antitussives	Dextromethorphan		
Antihypertensives	Debrisoquine		Irbesartan Losartan
Antipsychotics	Most, ex. haloperidol Risperidone Perphenazine Thioridazine Zuclopenthixol Remoxipride Aripiprazole		
Antiemetics	Ondansetron tropisetron Metoclopramide		
Beta blockers	Alprenolol, altenolol	Propranolol	
Stimulants	Amphetamine		
Drugs of abuse		Tetrahydrocannabinol	
Antidiabetic drugs	Phenformin	Gliclazide	Glibenclamide Glimepiride Glipizide Tolbutamide rosiglitazone
Antihistamines	Chlorphenamine		
SERMs	Tamoxifen		
Vinca alkaloids	Vincristine		
Antiepileptics		Nordazepam Diazepam Phenytoin Phenobarbital primidone	Phenytoin

(continued)

Table 4.1 (continued)

CYP2D6	CYP2C19	CYP2C9
Proton pump inhibitors	Esomeprazole Lansoprazole Omeprazole Pantoprazole Rabeprazole	
Antimalarial	Proguanil	
Anticoagulants	Clopidogrel warfarin	Warfarin (S enantiomer)
NSAIDs	Diclofenac indomethacin	Celecoxib Lornoxicam Diclofenac Ibuprofen Naproxen Piroxicam Meloxicam
Alkylating agents	Cyclophosphamide	
Antiretroviral drugs	Nelfinavir	
Chemotherapeutics	Teniposide	
Hormones	Progesterone	
Muscle relaxers	Carisoprodol	
Antifungals	Voriconazole	Terbinafine miconazole
Statins		Fluvastatin pitavastatin
Erectile dysfunction drugs		Sildenafil

(because no drug is given to the patient) and much more convenient (one blood sampling versus several over a long period of time) than a drug clearance study. In addition, because the CYP450 superfamily of enzymes metabolizes so many commonly prescribed drugs, a CYP450 test that includes polymorphisms in several CYP450 genes can predict the individual's response to many drugs, including drugs from different classes. The results of a direct drug kinetics test such as the DEM test may only be applicable to that drug or drugs that are metabolized by the same pathways.

The CYP450 tests will need to be made more inclusive, and the CYP450 information combined with other genetic and nongenetic factors, to produce a test that predicts the patient's response to drugs reliably enough to be used in clinical practice. Once the CYP450 tests reach that level of development, however, they will provide safer and more convenient ways of estimating the patient's risk for ADRs, and provide information relevant to a great many drugs. Further, once they become routine, wide-ranging genetic tests such as whole-genome sequencing or SNP microarrays will continue to provide new information of benefit to the patient,

without requiring new testing. As we learn more about the means by which our proteins influence our health and metabolize drugs, we will discover new ways to use information from past sequencing or microarray tests to predict the individual's susceptibility to an ever-expanding array of diseases or his/her response to an ever-expanding list of drugs.

4.3 Limitations of Genetic Testing

As discussed in Chap. 3 (see Sects. 3.4, 3.5 and 3.6), the conventional measure of the value of any new clinical test is whether it classifies the patient or predicts a clinical outcome better than the measures that are currently used to do so. At this point in time, few genetic tests classify patients or predict outcomes better than the conventional predictors do.

While genetic factors account for a considerable portion of the interindividual variability in drug response, because there are often many genetic factors that influence the individual's drug response, any single-gene test only tells one part of a many-faceted story. Unfortunately, most of the currently available tests focus only on one gene, or on a set of genes that influence only one aspect of drug response (usually pharmacokinetics).

While the partial successes that have been achieved by these tests have proven the principles behind personalized medicine, they have also proven that the influence our genes have over drug metabolism is much more complex than we thought. In order to properly assess the effects that genetic variation has on the response to any drug, it is necessary to assess thousands, perhaps tens of thousands, of polymorphisms in genes whose proteins influence all aspects of the drug's metabolism and effects on the body. It is also necessary to know whether there are interactions between different gene's alleles. Some genes exhibit *epistasis*, in which the effect of the patient's allele status for one gene will depend on his/her allele status for a second gene. For example, if a patient has an extremely high level of activity in a multidrug transporter protein that reduces the absorption of the drug by transporting the drug back out of the intestinal epithelial cells that absorb it, it may not matter whether the patient also has low-activity versus high-activity alleles in one of the CYP450 genes whose protein metabolizes that drug, because the patient may absorb so little of the drug that he/she cannot build up a therapeutic concentration, even if he/she has very low activity in the critical CYP450 protein.

It is only through the technological advances of the last (and next) decade that we will be able to design tests that include the number of gene variants that will be needed to accurately predict most patients' responses to a drug. In addition, the companies that offer tests must avoid the tendency to specialize in one type of information, and combine genetic information with clinical data, family history data, and information of the individual's diet, environment and lifestyle to produce the best possible predictive algorithms.

There are several factors that limit the clinical relevance of a pharmacogenetically relevant polymorphism. Foremost among them is the complexity of the mechanisms whereby drugs are processed in the body. Some drugs are administered in an inactive form, and metabolized to active compounds. Other drugs are administered as active compounds, and then metabolized into compounds that may or may not be active themselves. Most pharmacokinetic studies have measured the level of the drug, and perhaps one active metabolite, in the subjects' blood samples, but many have not measured the level of all that drug's active metabolites. If the drug has active metabolites, the total concentration of active compounds is a more relevant measure of therapeutic potential and risk for ADRs than the concentration of one compound is. In some cases it may be difficult to identify all the active metabolites of a drug, or predict the effects of a genetic polymorphism on the total concentration of active compounds that will result from drug administration.

Similarly, some studies do not differentiate between the R- and S-enantiomers[1] of the compounds they assay, despite the fact that different enantiomers may have different levels of biological activity. In addition, genetic polymorphisms may affect the metabolism of the different enantiomers of a compound differently. In order to classify patients according to their pattern of metabolizing a drug, not only must we measure the kinetics of all the active compounds that are associated with that drug, we must also understand the different effects a genetic variant may have on the different enantiomers of the active compounds.

This complexity is compounded by the fact that many drugs are capable of being metabolized by several different pathways, so a deficiency in one pathway may not change the total concentration of active metabolites in the patient's blood to a degree that will have clinical significance. There are examples, however, of drugs for which one metabolic pathway produces active compounds, but the others do not. In these cases, if the pathway that produces the active compounds is deficient, the patient may not achieve a therapeutic blood level of the active compounds. Alternatively, if one of the other pathways is deficient, more of the drug may be metabolized through the pathway that produces the active compounds, increasing that patient's risk for concentration-dependent ADRs.

Another factor that may limit the clinical relevance of a pharmacogenetic polymorphism is the fact that there is not always a clear relationship between the concentration of a drug or metabolite and either its therapeutic or adverse effects. In these cases, even knowing the polymorphism's effect on the kinetics of the active compounds does not allow the analyst to predict the clinical utility of the test. A pharmacogenetic test may also be of limited clinical utility if the drug in question has a high therapeutic index, and the patient is unlikely to suffer ADRs, even if he/she is slow to clear the active compounds.

---

[1] Some compounds exist in two mirror-image forms. These forms of the compound are not identical or interchangeable, but are mirror images of each other, similar to a person's left and right hands.

Finally, chronic administration and multidrug exposure introduce issues that are not taken into account in acute drug clearance studies. There are several reasons why the metabolism of a drug may slow down after repeated dosing. For some drugs, chronic exposure to the drug slows the metabolism of that drug, often through unknown mechanisms that inhibit the activity of the enzymes that metabolize it. In addition, some drug clearance systems are saturable; patients with different initial clearance rates may achieve similar steady-state drug levels after chronic dosing. Further, many drugs inhibit the enzymes that metabolize other drugs, which has the same effect of slowing the clearance in an individual who originally had a fast clearance rate.

When an individual who has a typical to high level of metabolism of a drug at first becomes a slow metabolizer after chronic exposure or addition of another drug to the regimen, the situation is referred to as a *phenocopy*, because a nongenetic factor (chronic drug exposure or addition of the other drug) has caused the patient to mimic a phenotype that was different from the one that his/her genotype originally endowed him/her with. In addition, when a drug inhibits its own clearance, the term *autophenocopying* or *autoinhibition* is often used to describe the situation.

The problems described above present formidable barriers to the clinical application of genomic information at the present time, but all the major problems can be addressed by more research, and the current explosion of genomic information will allow many of them to be addressed in the near future.

4.4 Dose-Calculating Algorithms Must Take Genetic and Nongenetic Factors into Account

The FDA has recently begun recommending that genetic testing be performed to inform the choice of dose for patients who are prescribed the anticoagulant warfarin. Warfarin inhibits activity in the VKOR complex, which regenerates reduced vitamin K from oxidized vitamin K. The individual's status for a functional polymorphism in the promoter region of the *VKORC1* gene (A-1,639G), which encodes the major subunit of the vitamin K oxidoreductase (VKOR) complex, plays an important role in determining the proper dose of warfarin for that individual. This is not the only factor that must be considered, however; warfarin is metabolized by the CYP2C9 enzyme, the activity of which, as mentioned above, is highly variable between different individuals. Warfarin occurs in both the R-enantiomeric form and the S-enantiomeric form. The S-enantiomer is 3–5 times more active than the R-enantiomer, and the S-form is metabolized by the CYP2C9 enzyme. The low-activity CYP2C9*2 and CYP2C9*3 alleles increase the individual's risk for potentially serious bleeding complications, by leading to buildup of toxic levels of warfarin after a standard dose.

One algorithm that has been suggested as the best means by which to calculate the proper dose of warfarin for the individual exemplifies the approach all

pharmacogenomic tests must take in the future. The algorithm takes several factors other than the status of the CYP2C9 and VKORC1 polymorphisms into account. The algorithm is constructed as follows:

$$\begin{aligned}\text{Square Root of Dose} = {} & 0.628 - 0.0135\,(\text{Age in years}) - 0.240\,(\text{CYP}*2) \\ & - 0.370\,(\text{CYP}*3) - 0.241\,(\text{VKOR}) \\ & + 0.0162\,(\text{height in centimeters})\end{aligned}$$

To input the genetic data, one enters 0, 1 or 2 to indicate the number of CYP2C9*2 and CYP2C9*3 alleles in the patient's genotype, and enters 1 if the individual has the GG genotype for the VKORC1 polymorphism, 2 if the individual has the GA genotype, and 3 if the individual has the AA genotype.

As you can see, the algorithm takes into account the multifactorial nature of the drug-metabolizing phenotype. It includes functional polymorphisms in genes whose proteins influence both the pharmacokinetics and pharmacodynamics of the drug, as well as nongenetic factors that influence drug response as well. As you can see from the formula, the greater the number of low-activity CYP2C9*2 and CYP2C9*3 alleles the individual possesses, the lower his/her recommended dose will be. In addition, the greater the number of low-activity A alleles the individual possesses for the VKORC1 polymorphism, the lower the dose will be. Further, older patients and shorter patients require lower doses than younger and taller patients do, although these factors are not weighted as heavily as the genetic factors are (coefficients of 0.24, 0.241 and 0.37 for the genetic factors, compared to 0.0135 and 0.0162 for age and height).

As we learn more about the means by which our genes influence our responses to drugs, these algorithms will become more complex, incorporate more genetic and nongenetic factors, and will become more accurate predictors of the individual's response to the drug. For example, gender will probably prove to be a critical factor in many predictive algorithms. The hormonal differences between men and women often constitute important sources of nongenetic factors that can influence drug response. In addition, if one of the critical functional polymorphisms lies in a gene that is located on the X chromosome, the algorithm must include an acknowledgment that the individual's gender will influence his/her susceptibility to the disorder or response to the drug.

For example, the cysteine-leukotriene receptor type I (*CYSLTR1*) gene lies on the X chromosome. A complete assessment of all the important functional polymorphisms in asthma-related genes includes an assessment of several polymorphisms in the promoter region and coding sequence of the CYSLTR1 gene. There is an unusually high degree of discrepancy between the results of studies that have measured the association between *CYSLTR1* gene markers and asthma, however. One reason for this may be that the association between the *CYSLTR1* gene alleles and asthma may be gender-specific, and many studies have not separately assessed the effects of *CYSLTR1* polymorphisms on male versus female patients.

One area in which there is much to be learned involves the potential for drug-drug interactions. Because most of the enzymes that metabolize prescription drugs

each metabolize many drugs, the speed with which the patient metabolizes a drug may be reduced if the patient is taking another drug that is the target of the same enzyme. It will therefore be important to know whether any other drugs the patient is taking can significantly alter the metabolism of the drug in question.

4.5 Epigenetic Factors Must Be Factored into Many Algorithms as Well

As we discussed in Chap. 1 (see Sect. 1.7), epigenetic factors such as methylation of the CG cytosines in the promoter region of the gene can have a strong influence on the level of activity in the gene. This can not only exert a strong influence on the individual's response to a drug, but it can also provide a test that can add considerable predictive power to the algorithm. Promoter methylation often results in a significant, if not complete, silencing of the allele. Adding promoter methylation status to a predictive algorithm could increase the algorithm's predictive power significantly, because it constitutes a reliable indicator that the individual possesses a true null allele (one that does not produce any functional protein) of the gene in question.

One example of the way in which epigenetic factors can be employed to improve predictive algorithms involves assessing the methylation status of the gene that encodes the DNA-repairing enzyme O^6-methylguanine-DNA methyltransferase (MGMT) to predict the response of patients with glial cell tumors to alkylating agents. The ability of MGMT to remove methyl groups from guanine nucleotides works against alkylating agents such as carmustine and temozolomide, which arrest the growth of tumors by methylating guanine nucleotides. Methylation of the nucleotides causes mispairing of the bases, and activates the mismatch repair system. Attempts to repair and re-replicate the DNA result in double-stranded breaks in the DNA, which activate apoptotic pathways.

The level of activity in the *MGMT* gene is regulated by methylation of the gene's promoter region. Methylation of the promoter leads to a reduction in the activity of the gene, and has been associated with improved response to both carmustine and temozolomide in patients with glial cell tumors. As discussed in Chap. 3 (see Sect. 3.15), methylation-specific PCR protocols exist that can reliably and quantitatively assess the degree of methylation at any given C nucleotide in the DNA. The results of these tests can sometimes be easier to interpret than the results of sequence polymorphism tests. It can be difficult to know exactly what effect one allele or another of a sequence polymorphism may have on the protein's activity, or whether the association that has been observed between drug response and a gene allele is due to that polymorphism being a functional polymorphism, or merely being linked to a functional polymorphism. In contrast, finding that the gene's promoter region is methylated can confirm that the gene is inoperative, allowing the analyst to know for certain that this represents a true null allele.

4.6 Polymorphisms in the CYP450 Genes Influence the Pharmacokinetics of Many Commonly Prescribed Drugs[2]

4.6.1 The CYP450 Enzymes Metabolize Many Commonly Prescribed Drugs

As mentioned earlier, it has proven easier to identify genes, and develop tests, related to the pharmacokinetics of drugs rather than their pharmacodynamics. The single most widely applicable discovery in the field of pharmacogenomics to date involves the discovery that the genes that encode many members of the cytochrome P450 (CYP450) superfamily of enzymes are polymorphic. The CYP450 proteins constitute a superfamily of microsomal enzymes that mediate the oxidative (Phase I) metabolism of up to 80% of the drugs that are prescribed today (Table 4.1), as well as a number of environmental chemicals. The CYP1, CYP2 and CYP3 families are the ones primarily responsible for the oxidative metabolism of drugs. A database of CYP alleles is maintained at http://www.cypalleles.ki.se/.

4.6.2 Defining the Metabolizer Phenotype by Assessing CYP450 Status

Individuals are classified by the CYP450 tests as either poor metabolizers (PM), intermediate metabolizers (IM), extensive metabolizers (EM) or ultrarapid metabolizers (UM). The EM phenotype constitutes the typical levels of metabolic activity seen in most individuals, and some studies combine subjects with the IM and EM phenotypes into the "typical activity" group. Patients with the UM phenotype clear the enzyme's target drugs quickly, and may not respond to a standard dose. Patients with the PM phenotype have a higher than average risk for ADRs, because their slow metabolism allows for the buildup of toxic levels of the drug or its metabolites. Mutations that inactivate the gene or the protein are responsible for the null alleles seen in patients with the PM phenotype. In contrast, copy number variation is responsible for some cases of the UM phenotype; some individuals with the UM phenotype have between 3 and 13 copies of the *CYP2D6* gene, instead of the usual two copies.

It is estimated that the CYP3A4 protein is responsible for approximately 50% of the CYP450 proteins' oxidation of prescription drugs. Although the level of activity in CYP3A4 varies between different individuals, the few genetic polymorphisms that have been reported to influence the protein's level of activity occur at frequencies

[2] Chapters 5, 6 and 7 contain information on the pharmacogenomics of drugs that are prescribed for the diseases that are covered in those chapters. In this chapter, we confine our specific examples to gene variants that affect the pharmacokinetics of multiple drugs, and are therefore applicable to many different fields of medicine.

that are too low to explain the level of variability. In addition, no null alleles have been found in *CYP3A4*, and therefore one cannot identify a class of patients with the PM phenotype, as one can with the other CYP450 genes.

The CYP3A5 protein also has variable levels of activity, but for this particular isoform, there are several relatively common null alleles. Consequently, all ethnic groups have a high frequency of individuals with extremely low levels of CYP3A5 activity. In Europeans, for example, the *CYP3A5*3* null allele is so common that it has been estimated that approximately 90% of people of European descent do not express any of the CYP3A5 protein. It is uncertain how much this contributes to interindividual variability in drug metabolism. This will be a difficult question to answer, in part because many of the drugs that are metabolized by CYP3A4 are also metabolized by CYP3A5.

Of all the CYP450 proteins, CYP2D6 is probably the most extensively studied with respect to pharmacogenomics. Interestingly, all the CYP450 proteins' activities are inducible by a number of drugs except that of CYP2D6, which means one can expect genetic variation to influence the level of activity in the CYP2D6 protein to a greater degree than the other CYP450s.

Over 80 different *CYP2D6* alleles have been reported. *CYP2D6* has several null or low activity alleles, with markedly different distributions between different ethnic groups. Approximately 7–10% of European Caucasians are *CYP2D6* PMs, primarily due to the high frequency of the *CYP2D6*3*, **4* and **5* alleles, which produce either no protein or an inactive one (the *CYP2D6*4* allele frequency is approximately 21%). These alleles are almost absent in the Asian populations that have been studied, however. Consequently, only 1% of Chinese, Japanese and Koreans are *CYP2D6* PMs. Interestingly, however, the average level of CYP2D6 activity is lower in Asians than in European Caucasians, primarily because the low-activity allele *CYP2D6*10*, which is virtually absent in European Caucasians, has a frequency of approximately 50% in the Asian populations. The frequency of the *CYP2D6* gene duplications that are often associated with the UM phenotype also varies greatly between ethnic groups. Gene duplication has been reported in 29% of black Ethiopians, but only 7% of white Spaniards, and 1% of white Swedes.

The results of studies assessing the effects of *CYP2D6* status on antidepressant response illustrate both the promise and limitations of pharmacogenomics. Several studies have suggested that *CYP2D6* status influences response to TCAs and other antidepressants. One study reported that 29% of individuals who had experienced an ADR after antidepressant administration were classified as PMs according to their *CYP2D6* genotype, while 19% of those who had failed to respond to their antidepressant medication exhibited gene duplications that classified them as UMs. These frequencies are four and five times greater, respectively, than the frequencies one would expect from the population's allele frequencies. In another study, the rate of gene duplication that was observed in patients who failed to respond to an antidepressant drug was ten times the expected frequency.

As encouraging as findings such as these are, it is also clear that the individual's drug response cannot be predicted by the *CYP2D6* genotype alone. The aforementioned research findings notwithstanding, the majority of individuals who do not

respond to their antidepressants or exhibit ADRs have the EM *CYP2D6* genotype. This is not surprising, given the high frequency of individuals with the EM genotype, but it illustrates the point that *CYP2D6* genotype is only one factor that influences response to these drugs.

SSRIs generally have higher therapeutic indices than TCAs, but the side effects that have been reported are more severe, including the sometimes fatal serotonin syndrome. The effects of *CYP2D6* status on SSRI metabolism are complicated by the presence of multiple enzymes capable of metabolizing several commonly prescribed SSRIs, as well as the fact that chronic exposure saturates the clearance mechanisms and leads to phenocopying. In addition, the relationship between the concentration of the drug and its effects is not always clear.

One reason for the lack of a clear relationship between *CYP2D6* status and SSRI response is the different effect the CYP450 isoforms have on the different enantiomers of the SSRIs. *CYP2D6* status clearly influences the clearance of fluoxetine and its major metabolite norfluoxetine. The clinical significance of this is complicated, however, by the fact that the S-enantiomer is approximately 20 times as potent as a serotonin reuptake inhibitor as the R-enantiomer is. The patient's *CYP2D6* status influences the metabolism of S-norfluoxetine to a greater degree than R-norfluoxetine, but most studies have not differentiated between the two enantiomers in reporting the effects of *CYP2D6* genotype on the kinetics of norfluoxetine.

One study that further highlights the complexity of the relationship between a gene variant's effects on drug concentration and its clinical significance is the study that reported a greater frequency of moderate to severe ADRs after the atypical antipsychotic risperidone in *CYP2D6* PMs. The usual challenge is that pharmacokinetic studies report significant effects of CYP2D6 status on drug concentration, but limited effects of *CYP2D6* status on clinical outcomes. In this case, pharmacokinetic studies report minimal effects of *CYP2D6* status on the sum of the concentrations of the drug and its active metabolite, but one study reported a three-fold increase in moderate to severe ADRs in PMs.

Less is known about the *CYP2C19* gene than about *CYP2D6*, but it appears that there are seven *CYPC19* null alleles (*CYP2C19*2 - *8*) that have been reported in the literature. The **2* allele is relatively frequent, having a frequency of 17% in African-Americans, 30% in Chinese and 15% in Caucasians. The **3* allele is the next most common; it has a frequency of 0.4% in African-Americans, 5% in Chinese and 0.04% in Caucasians. Testing for the **2 - *8* alleles identifies virtually all PMs in the Caucasian population. In addition, the *CYP2C19*17* allele has been reported to produce a CYP2C19 isoform with an atypically high level of activity. This allele is also relatively common in some ethnic groups; it has a frequency of approximately 18% in Swedes and Ethiopians and 4% in Chinese.

With respect to the *CYP2C9* gene, there are a number of null or low-activity alleles of *CYP2C9*, but the *CYP2C9*2* and **3* alleles are the most clinically relevant. Approximately one-third of Caucasians carry one copy of either the *CYP2C9*2* or **3* allele. Interestingly, these alleles are both very rare in African-Americans and Asians.

The currently available CYP450 tests are imperfect predictors of metabolic status. One of the more important reasons for this is that all active alleles are treated as if they produced protein isoforms with the same level of activity. This is unlikely to be true; slight differences in the sequences of these alleles will undoubtedly result in the different isoforms of the protein having slightly different levels of activity. In addition, the CYP450 genes contain other important functional polymorphisms that affect CYP450 activity. Further, there are several drugs that can be metabolized by more than one CYP 450 protein, making it difficult to gauge the degree to which possessing a low-activity CYP allele will affect drug metabolism. For example, approximately 75% of white patients and 50% of black patients possess a null allele for the CYP3A5 gene, which does not produce an active form of the protein. Many of the drugs that are metabolized by CYP3A5, however, are also metabolized by CYP3A4, for which there seems to be very few null alleles in any population that has been studied to date.

In addition to these genetic factors, several nongenetic factors such as age, gender, hormonal status, hepatic disease, inflammation, nutritional status, pregnancy and environmental factors can affect CYP protein activity as well. For these reasons, it is considerably easier to identify patients with no CYP activity than to predict the level of activity in patients with active alleles. The currently available CYP450 tests identify virtually all patients with no or very low CYP activity (PMs), because they possess null alleles that do not produce an active protein. In contrast, it is a much less reliable predictor of drug metabolism phenotype for patients with one, two, or even more copies of active alleles.

One of the factors that make it difficult to predict the level of activity in individuals who possess more than two copies of the CYP2D6 gene is that it is as yet unknown whether all the extra copies of the gene are functional. In fact, only approximately 20% of people with extra CYP2D6 copies have the UM metabolizing phenotype. Gene amplification is also not the only mechanism whereby one achieves UM status. While many with the UM phenotype have extra copies of the CYP2D6 gene, many do not appear to.

4.7 Other Functional Polymorphisms That Affect the Pharmacokinetics of Multiple Drugs

4.7.1 N-Acetyltransferase 2

One of the seminal studies that launched the field of pharmacogenetics reported that there were interindividual differences in the acetylation of isoniazid, and that tuberculosis patients who were slow acetylators had a higher than average risk for peripheral neuropathy after isoniazid. The enzyme responsible for acetylation of isoniazid is N-acetyltransferase 2 (NAT2). The activity of NAT2 is trimodally distributed, with considerable differences between ethnic groups in the distribution

of activity. Only approximately 10% of Japanese and 20% of Chinese individuals are slow acetylators, while 40–70% of African-Americans and Caucasians are.

NAT2 status also significantly affects the individual's response to the antihypertensive drug hydralazine. Several studies have reported a greater response to standard doses in slow acetylators versus fast acetylators and/or suggested that fast acetylators require higher doses of hydralazine to achieve the desired therapeutic effect. Although the bulk of the evidence supports this claim, there are a few studies, conducted using different methods, that do not confirm these findings. Several studies also suggest that slow acetylators have an increased risk for a lupus-like ADR after hydralazine.

Slow acetylation increases the individual's risk for ADRs to a number of drugs. For example, most patients who are prescribed procainamide develop antinuclear antibodies, and 10–20% develop a lupus-like reaction. Both these effects may occur more rapidly in slow acetylators than in fast acetylators. Similarly, slow acetylators have an increased risk for hypersensitivity reactions after sulfonamide antibiotics.

NAT2 acetylates sulfasalazine, but conflicting results have been reported in studies that have assessed the effect of NAT2 status on sulfasalazine's therapeutic effect. There is more agreement, however, that slow acetylators have an increased risk for ADRs after sulfasalzine, including cyanosis, hemolysis and reticulocytosis.

4.7.2 Butyrylcholinesterase

Butyrylcholinesterase (BChE) was the subject of one of the seminal studies that launched the field of pharmacogenetics, in which it was reported that a noticeable minority of patients were at risk for prolonged apnea after being prescribed succinylcholine. BChE does not metabolize many prescription drugs, so BChE phenotype is primarily important for patients being administered anesthetics. For example, low levels of BChE activity convey an increased risk for prolonged paralysis after administration of two neuromuscular blockers, suxamethonium and mivacurium. In addition, the effect of epidural 2-chloroprocaine is enhanced in patients with low BChE activity. In dental patients receiving procaine, a higher frequency of ADRs, including weakness, nausea, dyspnea and unconsciousness due to hypoxia, have been reported in patients with low BChE activity.

4.7.3 Functional Polymorphisms in Drug Transporter Genes also Affect the Response to Many Drugs

Approximately 10% of a cell's proteins are transporters. They control the absorption of drugs from the gut (or other mode of delivery), the distribution of drugs throughout the body and the efflux of drugs from cells. The levels of activity in

these transport proteins are critical determinants of the concentration the individual will achieve at the relevant target tissues after any given dose.

The ATP binding cassette proteins (ABCs), also called multidrug resistance proteins (MDRs), actively transport many commonly prescribed drugs out of cells. For example, the p-glycoprotein (aka ABCB1, aka MDR1) transporter transports bilirubin, several anticancer drugs, cardiac glycosides, immunosuppressive agents, glucocorticoids and HIV type I protease inhibitors, and limits the accumulation of several drugs in the brain, including digoxin, ivermectin, vinblastine, dexamethasone, cyclosporine, domperidone and loperamide. In addition, the ABCC4 (aka MRP4) protein transports nucleoside antiretroviral drugs, such as zidovudine, out of the cells.

The most commonly tested polymorphism in the *ABCB1* gene is a SNP in exon 26 (C3435T = synonymous SNP I1145I). Although this SNP does not alter the amino acid content of the protein, possessing the TT genotype of the C3435T polymorphism results in greater than 50% reduction in p-glycoprotein protein level in the duodenum, greater retention of the p-glycoprotein substrate rhodamine in CD56+ killer cells *in vitro*, greater oral bioavailability of digoxin, and better CD4 cell recovery in HIV-infected patients treated with nelfinavir and other anti-retroviral drugs.

Another *ABCB1* polymorphism, the G2677T SNP, affects the amino acid content of the protein; possessing the C allele causes alanine to be incorporated as the 893 rd amino acid in the protein, while possessing the T allele causes serine to be incorporated in that position. Interestingly, the effects of the G2677T genotype on drug metabolism may depend on the drug itself. The TT genotype apparently results in lower plasma levels of fexofenadine, but higher plasma levels of digoxin.

Other ABC transporters, such as ABCC2, ABCB4 and ABCB11, have also been the subject of pharmacogenetic studies. Emerging evidence suggests that polymorphisms in both the *ABCB4* and *ABCB11* genes influence the individual's risk for cholestasis.

Another family of transporter proteins that transport drugs and their metabolites out of cells is the organic anion transporter family of proteins. Both the ABCC1 member of the ATP-binding cassette family of proteins and the solute carrier organic ion transporter family, member 2B1 (SLCO2B1) transport leukotriene (LT) receptor antagonists. One SNP in the *SLCO2B1* gene has been shown to influence both plasma levels of montelukast and response to the drug after 1 month and 6 months of treatment. In addition, one SNP in the *ABCC1* gene influences the patient's response to montelukast.

The *SLCO1B1* gene codes for the organic anion transporter OATP1B1, which transports anionic drugs into liver cells. The individual's *SLCO1B1* status influences the pharmacokinetics of several statin drugs, and may influence the individual's risk for ADRs such as statin-induced myopathy.

The organic cation transporters also transport some of the more commonly prescribed drugs, and are important determinants of the concentrations of drug that are achieved in organs such as the liver, kidney, heart and brain. For example, variants in the organic cation transporter type 1 and type 2 (*OCT1* and *OCT2*, respectively) genes influence absorption and clearance of the antihyperglycemic drug metformin, which

is commonly prescribed for patients with Type 2 diabetes. The OCT1 protein regulates the uptake of metformin into the liver, while OCT2 regulates the renal clearance of the drug. Patients who have low-activity isoforms of OCT1 have a limited ability to take metformin up into the liver, and because of this achieve a smaller reduction in plasma glucose levels after metformin than do patients with typical-activity isoforms of OCT1. In contrast, patients who possess a low-activity isoform of OCT2, which is primarily responsible for the renal clearance of metformin, achieve higher blood metformin levels than patients with other *OCT2* genotypes do.

In addition to influencing the acute actions of metformin, polymorphisms in the organic cation transporter genes can also influence its interactions with other drugs. Metformin interacts with several other prescription drugs, including cimetidine, the histamine H2 receptor antagonist that is commonly used in the treatment of heartburn and peptic ulcers. Cimetidine reduces renal clearance of metformin, and one study has shown that cimetidine reduces the clearance of metformin to a significantly lesser degree in patients with the TT genotype for the G808T polymorphism in the *OCT2* gene than in patients with other *OCT2* genotypes.

Polymorphisms in the *SLC22A2* gene, which encodes the organic cation transporter OCT2, also influence a patient's risk for nephrotoxicity after cisplatin. The *SLC22A2* gene has a G808T (serine270alanine) polymorphism, the T (alanine) allele of which exhibits reduced expression. Because OCT2 transports cisplatin into kidney cells, the low-activity T allele reduces renal clearance of the drug, but also reduces the patient's risk for cisplatin-induced nephrotoxicity.

4.8 Polymorphisms in the Genes Encoding Beta-Adrenergic Receptors Influence the Pharmacodynamics of Beta-Blockers

Drugs that antagonize the beta-adrenergic receptors are prescribed for patients with hypertension as well as cardiac arrhythmias. Because these drugs act through the beta-adrenergic receptors, it is no surprise to see that polymorphisms in the genes that encode these receptor proteins affect the patient's response to beta-blockers. The *ADRB1* gene encodes the beta-1 adrenergic receptor, and has a single nucleotide substitution that causes either arginine or glycine to be incorporated as amino acid number 389 (R389G). Although larger clinical studies are needed to confirm this, current research suggests that patients who are homozygous for the glycine variant show a poorer response to beta-blockers, and may require higher doses than patients with the other *ADRB1* genotypes.

Many of the older beta-blockers have affinity for both the type 1 and type 2 ADRB receptors. There are polymorphisms in *ADRB2* that influence the level of activity in the ADRB2 receptor protein, but recent research suggests that the *ADRB2* polymorphisms do not influence the response to beta-blockers as much as *ADRB1* polymorphisms do. It is particularly unlikely that *ADRB2* polymorphisms

will influence the effectiveness of the newer beta-blockers, which have greater affinity for the ADRB1 receptor than the ADRB2 receptor.

4.9 Keeping up to Date with FDA Approvals and the Status of the Field

Much of the information that was presented in this chapter is still in the research stage. To determine whether any specific gene test is available through a commercial testing company or research laboratory, consult the NIH's GeneTests website (http://www.ncbi.nlm.nih.gov/sites/GeneTests/); this site provides the most comprehensive and current listing of laboratories that offer genomic testing.

Approximately 10% of all drugs that are currently approved by the FDA contain information on their labels regarding pharmacogenomic tests that can be performed to help determine the best starting dose for each individual patient, or to classify patients according to the likelihood they will respond favorably. The current version of the FDA's Table of Valid Genomic Biomarkers in the Context of Approved Drug Labels can be found at (http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm). Alternatively, one can go to the FDA's homepage (http://www.fda.gov/) and enter the term "genomic biomarkers" in the intrasite search window. Although this information is useful, the vast majority of these drug labels merely provide pharmacogenomic information, rather than suggest a course of action. Only a few of these labels classify the test as "recommended," and even fewer cite the test as "required."

The FDA table lists only those genetic/genomic tests that have been approved by the FDA. Thousands of other genetic polymorphisms are being investigated, and this list will undoubtedly grow steadily in the next few years. In order to stay current, you can subscribe to receive email updates from the FDA on a host of topics. From the FDA website, click on "Email Updates" in the "Get Updates" window, then select from the long list of specific offerings. Of greatest relevance to pharmacogenetics and pharmacogenomics is the section entitled "Drugs." Under this heading there are subsections on "Genomics at FDA" as well as other issues such as patient safety bulletins. In addition, under "News and Events," you will find "FDA Updates for Health Care Professionals." This section informs the practitioner on a wide variety of topics, including new medical product approvals.

One easy way to access the information that is included on specific drug labels and package inserts is to consult the website for the DailyMed project (http://dailymed.nlm.nih.gov/dailymed/about.cfm). The DailyMed project is administered by the NIH and the FDA. Its website provides a continuously updated catalog of the information that is included on labels and package inserts for many prescription drugs. The website is very easy to navigate; one merely puts the drug name in the intrasite search window, and the site provides a complete and current listing of the information that is distributed along with the drug.

If a physician wants to enroll a patient in a clinical trial, the NIH maintains a website listing ongoing clinical trials, not just in pharmacogenomics but all areas of medicine (http://clinicaltrials.gov/). The listing explains the purpose of the study, the eligibility requirements, whether the study is recruiting patients or not, and provides contact information for those interested in participating in a study.

In addition, there are many research laboratories that can perform testing that reveals the individual's status for critical polymorphisms as part of their ongoing study of that gene or disease. These laboratories may not be formally certified to perform these tests, but the information these labs generate is often both valid and useful, and can sometimes be confirmed by a CLIA-certified laboratory. The individual practitioner will need to consider the relative risks and benefits of acquiring this information before it has been formally approved for clinical use.

Further Readings

Daly, AK (2010) Pharmacogenetics and human genetic polymorphisms Biochem J 429, 435–449

Department of Health and Human Services (2008) Report of the Secretary's Advisory Committee on Genetics, Health and Society. http://oba.od.nih.gov/SACGHS/sacghs_documents.html

Gardiner SJ and Begg EJ (2006) Pharmacogenetics, Drug-Metabolizing Enzymes, and Clinical Practice. Pharmacological Reviews 58:521–590

Zhou SF, Di YM, Chan E, Du YM, Chow VD, Xue CC, Lai X, Wang JC, Li CG, Tian M, Duan W (2008) Clinical pharmacogenetics and potential application in personalized medicine. Curr Drug Metab Oct;9(8):738–784

Chapter 5
Taking a Personalized Medicine Approach to Breast and Colon Cancer

Abstract One's susceptibility to cancer is influenced by many proteins, including those involved in cell signaling and proliferation pathways, repair of DNA damage, the mitotic cycle and apoptosis. Several relatively rare but highly penetrant alleles that increase one's risk for breast or colon cancer have been identified. In addition, a significant number of more common, but less penetrant, risk-increasing alleles have been identified as well. The chapter includes an extensive discussion of hereditary breast-ovarian cancer syndrome, *BRCA* gene testing, and a case report that illustrates how *BRCA* testing can be used to estimate risk for your patient and his/her family members. The chapter also includes discussions of Lynch syndrome and familial adenomatous polyposis, and the genetic and immunohistochemical tests that can be done to diagnose and predict risk for family members. Genetic tests have already been developed that classify patients with certain cancers into subgroups according to their molecular pathology. These tests are guiding the choice of treatments for patients with certain cancers. In addition, several gene variants are known to influence the pharmacokinetics or pharmacodynamics of specific anti-cancer drugs. The chapter also includes a list of Internet websites that will analyze family history and personal data and calculate one's risk for developing breast cancer.

5.1 Cancer Is a Complex Genetic Disease

Cancer is a term that encompasses a complex group of more than 100 different diseases that all share the primary characteristic of uncontrolled cell growth. Each individual type of cancer arises from a single cell that requires varying numbers of gene mutations for progression to the invasive state. Cancer is therefore a genetic disease, although only a small portion of all cancers (10–25%) are caused by inherited (germline) gene mutations. More often the cumulative effect of genetic damage (somatic mutations) that has been acquired over the individual's lifetime leads to malignancy. Genes involved with cell signaling and proliferation pathways, repair of normal DNA damage, the mitotic cycle and apoptosis all contribute to the risk for cancer. Another reason why cancer constitutes a complex disease is the fact that

K.M. Sweet and R.C. Michaelis, *The Busy Physician's Guide To Genetics, Genomics and Personalized Medicine*, DOI 10.1007/978-94-007-1147-1_5,

numerous nongenetic factors, such as environmental radiation exposure, carcinogens from cigarette smoke, and even age are contributing factors.

The two best-studied classes of cancer genes are oncogenes and tumor suppressor genes. Oncogenes are produced from mutated proto-oncogenes, whose proteins normally perform functions that stimulate cell proliferation and increase blood supply. These genes control the speed with which the cell goes through its cycle of replication and division. Most proto-oncogenes are only active in a subset of tissues, or during certain periods of the lifespan. Aberrant activation of one or more of these genes can cause a cell to resume or ramp up the cycle of replication and division after it should have stopped replicating and dividing and activated the apoptosis (programmed cell death) pathway.

Proto-oncogenes are transformed into oncogenes in one of three ways: gain-of-function point mutations, translocations or gene amplification (see Sect. 2.8). In each situation, mutation in a proto-oncogene results in over-expression of the gene, leading to disruption of the normal feedback loops within a cell and unregulated cell growth. Because the mutation is a gain-of-function mutation, and the other normal allele cannot do anything to reign in its overactive partner, mutation of a single allele of a proto-oncogene is all that is necessary to change the cellular phenotype. Examples of oncogenes include *KRAS, MET, RET, WNT, MYC,* and *TRK*.

Tumor suppressor genes also help regulate the cell's cycle of replication and division, and the process whereby cells go into apoptosis. Tumor suppressor proteins induce the cell to enter the apoptosis pathway if the cell has suffered DNA damage, which might lead to unregulated replication and division. Just like the gain-of-function mutations that activate oncogenes, loss of function mutations in tumor suppressor genes also predispose the individual to cancer. However, because the situation involves loss-of-function mutations, and the remaining normal allele can often produce enough protein to support normal cell function, a mutation in a single copy of a tumor suppressor gene is usually not sufficient to cause cancer. Instead, both alleles of a tumor suppressor gene must be altered by loss-of-function mutations.

This principle underlies the well-known "two-hit" model of carcinogenesis. Most hereditary cancer syndromes are associated with germline mutations in tumor suppressor genes such as *BRCA1*, *BRCA2, APC, MEN1* and *RB*. The affected individual will often have a germline mutation in one copy of the tumor suppressor gene, and have a somatic mutation (see Sect. 3.17) occur in the other copy of the gene, causing that cell to continue repeating the cycle of replication and division after it should have stopped cycling and gone into apoptosis.

There are a great many genes involved in the regulation of the cell cycle and apoptosis. The effort to find the genetic variants that influence cancer risk and the response to treatment has been complicated by the number of potential candidate genes, and by the fact that somatic mutations are an important cause of many cancers. It is often necessary to determine the genetic status of the malignant cells themselves in order to guide treatment decisions.

In 2006 the National Cancer Institute and the National Human Genome Research Institute started the Cancer Genome Atlas project (http://www.genome.gov/17516564),

which systematically catalogs mutations occurring in many common tumors. The original goal of the project was to provide systematic, comprehensive genomic characterization and sequence analysis of three types of human cancers: glioblastoma multiforme, lung cancer and ovarian cancer. Over the next 5 years the project will provide genomic characterization and sequence analysis for an additional 25 different tumor types. Further defining these molecular pathways will allow for even more targeted and effective cancer treatments.

5.2 Breast Cancer Gene Variants with Low Penetrance

Breast cancer is the most common cancer in women worldwide, with global incidence of more than 1.2 million new cases diagnosed each year. The incidence is highest in industrialized countries, particularly among populations of Northern European origin. In the United States, the annual incidence rate is approximately 125 cases per 100,000; in 2008 more than 180,000 new cases of invasive breast cancer were diagnosed. An estimated 68,000 ductal- and lobular-carcinoma *in-situ* cases were also diagnosed. For most US women the lifetime likelihood of developing breast cancer is 12.3% (one in eight). However, given that breast cancer tends to cluster in families, the disease is approximately twice as common among first-degree relatives of affected individuals as among women in the general population.

Genome-wide association studies have identified at least 17 variants that constitute risk-increasing alleles for breast cancer. Unfortunately, these discoveries do not automatically translate into clinically useful tests. Many of these risk-increasing alleles are either rare enough in the population that they do not contribute to many cases, or of low enough penetrance that possessing the risk-increasing allele does not significantly increase the individual's risk for the disease.

The highly penetrant gene variants that are discussed in the next section are associated with high relative risks, but they account for at most 10% of the inherited component of breast cancer, primarily because they are relatively rare in the population. The lion's share of the inherited component of breast cancer lies in variants that are considerably more frequent, but have relatively low penetrance. For example, although the risk-increasing allele of the rs2981582 SNP in the *FGFR2* gene is relatively common (frequency of 38% in the general population), it only accounts for approximately 2% of the genetic risk of breast cancer.

Table 5.1 illustrates the nine breast cancer gene variants that are commonly used in the commercially available tests that predict a woman's risk for breast cancer. Some companies use most or all of these markers, others just a few. The most common of these risk-increasing alleles is the D allele of the D302N SNP in the *caspase 8* (*CASP8*) gene. As the table illustrates, this allele has a frequency of 0.86, but only conveys a relative risk of 1.13. As discussed in Chap. 3 (see Sect. 3.7.2), a relative risk of 1.13 means that an individual who possesses the risk-increasing allele is 1.13 times as likely to develop the disorder in his/her lifetime than an individual who does not possess the risk-increasing allele.

Table 5.1 Frequencies of and relative risks associated with the commonly tested SNPs for breast cancer

dbSNP No.	Gene	Chromosome	Risk-increasing allele frequency	Relative risk per allele
rs2981582	*FGFR2*	10q	0.38	1.26
rs3803662	*TNCR9*	16q	0.25	1.20
rs889312	*MAP3K1*	5q	0.28	1.13
rs3817198	*LSP1*	11p	0.30	1.07
rs1053485	*CASP8*	2q	0.86	1.13
rs13281615	*Unknown*	8q	0.40	1.08
rs13387042	*Unknown*	2q	0.50	1.20
rs4415084	*MRP530*	5p	Unknown	1.03
rs1219648	*FGFR2*	10q	Unknown	Unknown

You will note that a few of these SNPs are not located in known genes. At least one allele of each of these markers has been empirically associated with a disorder through a genome-wide association study, however. This suggests that these SNPs are either located inside genes that have yet to be identified, or that they lie close to, i.e. are genetically linked to (see Sect. 3.3.1), important functional polymorphisms.

For example, the rs13281615 SNP lies within 1 Mb (1,000,000 bp) of both the *MYC* and *PVT1* genes. Activation of both *MYC* and *PVT1* have been implicated in some cases of Burkitt lymphoma, especially in patients who have the chromosome translocation t (2;8) that is frequently associated with this disorder. Similarly, the rs13387042 SNP lies within 1 Mb of both the *NCK* and *TGFBRAP1* genes. Because both the NKC and TGFBRAP1 proteins serve as part of the mechanism whereby serine-tyrosine kinase receptors transduce extracellular signals into changes in intracellular metabolism, these genes could have some implication for breast cancer as well.

The relative risk associated with a risk-increasing allele is often dose–dependent and complex; possessing two copies of the risk-increasing allele often conveys more than twice the risk that possessing one copy of the risk-increasing allele does. For example, for the rs2981582 SNP in the *FGFR2* gene, the genotype specific relative risk of breast cancer (individual's risk compared to the risk in the general population) is 0.83 for carriers of two risk-reducing alleles (common allele homozygotes), 1.05 for carriers of one risk-increasing and one risk-reducing allele (heterozygotes) and 1.38 for carriers of two risk-increasing alleles (rare allele homozygotes). This projects a lifetime risk of 10% for women who carry one risk-increasing allele and approximately 13% for women who carry two risk-increasing alleles.

5.3 Further Research Will Increase Accuracy and Standardize Risk-Estimating Algorithms

The commercial testing companies provide breast cancer risk estimates that are based on the number of these low-penetrance alleles the individual possesses. Each company uses a different specific algorithm to combine the genotype data

and produce their risk estimate. For example, the deCODE BreastCancer™ test determines the individual's status for seven of the risk alleles listed in Table 5.1. They multiply risk quotients at each of the seven markers to define risks from 0.4- to 4.0-fold relative to the general population. deCode notes in their analysis that approximately 5% of women have a two-fold increase in risk on average for breast cancer compared to the general population, while approximately 1% have a three-fold increase in risk. This translates to a lifetime risk of 24% and 36%, respectively, versus the average lifetime risk of 12%. In contrast, 23andMe analyzes two of the SNPs listed in Table 5.1: rs1219648 and rs3803662. Based on these two markers, the company reports lifetime risks of breast cancer from 12% to 27%.

In time, research will identify the critical genetic and nongenetic factors, and the way in which they must be combined to produce the most accurate estimate possible of an individual's risk for cancer. Different companies' risk-estimating algorithms will undoubtedly become more similar as time goes on. For the near future, however, different companies' tests may produce different estimates of the same individual's risk.

The fact that different companies use different panels of markers for their calculations is only one reason for this difference. Another reason is that different companies may use different estimates for the general population's risk for cancer. The individual's lifetime risk is calculated by beginning with a baseline (general population) risk, then multiplying that risk by the relative risks (RR) that are associated with the risk-increasing and risk-decreasing SNP alleles that were identified. An individual's baseline risk for cancer is influenced by his/her ethnic background, however. One company may use a single figure for all women's baseline risk, while another may use different baseline risk estimates for women of different ethnic groups.

At this point in time, it is also impossible to determine which of any two different companies' risk estimates is the more accurate. The accuracies of all these risk-estimating algorithms are limited, because there are many important risk-influencing factors that do not get taken into account by any of the commercial companies' risk-estimating algorithms. For example, at this point the nongenetic factors that influence cancer risk are virtually ignored by these algorithms. Risk estimates are usually calculated by multiplying a baseline risk by a factor that reflects the individual's allele status for the tested markers. Baseline risk estimates are based on large population data, and do not take into account the fact that within any single ethnic group, there are subgroups with different levels of exposure to the critical nongenetic factors, and therefore different levels of risk.

Our current understanding of gene-gene interactions remains limited as well, so there is no way to know whether different SNPs' data should be combined in an additive, multiplicative or synergistic fashion. It is likely that research will ultimately show that all three of these models apply, depending on the specific markers being analyzed. Until further research can identify the critical factors and illustrate how they can best be combined in a risk-estimating algorithm, the accuracy of these risk-estimating tests will be limited, and different companies may provide different estimates for the same individual.

5.4 Useful Online Programs to Calculate Breast Cancer Risk

Several different programs are available online for calculating a woman's breast cancer risk. As described above, these risk-estimating algorithms will be improved by further research, and will someday incorporate genetic data, family history, diet, medications, environment and lifestyle factors in the formula. In addition, it has been said that age is the most common carcinogen; a complete model must include the individual's age, and perhaps age at first menstrual period or age at first childbirth as well.

Despite their limitations, the currently available risk-estimating algorithms are based on a substantial amount of empirical data, and will help you provide as accurate an estimate as you can at the present time. Among the risk-estimating programs that are available online are:

The Gail Model at the National Cancer Institute's website (http://www.nci.nih.gov)

The Claus model (BreastCa for Palm, version 1.0, copyright 2001) http://www.palmgear.com/index.cfm?fuseaction=software.showsoftware&prodID=29820

The CancerGene 5.1 model from the U.T. Southwestern Medical Center http://www4.utsouthwestern.edu/breasthealth/cagene/default.asp

These models each utilize different combinations of risk factors derived from different data sets, and each has its own strengths and limitations. As a result, they may generate different breast cancer risk estimates for each individual. Although all three models incorporate family history of breast cancer in the risk calculation, the Gail model is limited because it does not incorporate second-degree relatives (e.g. aunts, grandmothers), paternal relatives with breast cancer or the age of onset of breast cancer in the family. Thus it is limited in its ability to identify women who are at moderate risk (>20–25% lifetime) for breast cancer due to family history. The current American Cancer Society's guidelines for breast cancer screening recommend that women with >20–25% lifetime risk of cancer have MRI screenings, so the Gail model's limited ability to identify members of this group is a significant limitation. The Gail model is useful in the clinical setting to identify women for whom chemoprevention of breast cancer with tamoxifen (5 year risk >1.7%) may be appropriate, however.

The Claus model incorporates more family history than the Gail model does, but is still incomplete in that respect. The Claus model incorporates up to two relatives with breast cancer, taking into account their ages of onset, but does not take into account nongenetic risk factors such as age at menarche. It also does not include ethnicity information in calculating risk.

Another notable limitation of both the Gail and Claus models are the fact that they are not useful at estimating if an individual woman may have a highly penetrant breast cancer gene mutation (e.g. *BRCA*) in their family. A third model, the CancerGene 5.1 model (aka BRCAPro model) from the U.T. Southwestern Medical Center at Dallas and The BayesMendel Group at Johns Hopkins includes *BRCA1*

and *BRCA2* gene mutation probabilities as well as modeling for mutation probability for a hereditary form of colon cancer, Lynch syndrome, which will be discussed later in this chapter. CancerGene also incorporates both the Gail and Claus models in its risk calculation.

In order to provide the best possible estimates of an individual's lifetime risk for breast cancer, models must be developed that incorporate the important nongenetic factors, the individual's status for the critical functional polymorphisms, and any adjustment factors that are necessary to account for the reduced penetrance of risk-increasing alleles, gene-gene interactions such as epistasis, epigenetic factors and interactions between genetic and nongenetic factors. At this point in time, although a number of critical functional polymorphisms and linked markers have been identified, little is known regarding the way these gene variants modify each other's influence, or interact with dietary, environmental or lifestyle factors to influence the individual's overall risk.

5.5 Highly Penetrant Breast Cancer Gene Variants

Although 20–25% of all breast cancers are familial (i.e. associated with a family history of breast cancer), only 5–10% are strongly hereditary, with high penetrance and an autosomal dominant single gene determinant (Table 5.2). Because of their high penetrance, these variants will almost always cosegregate with the disease in families. Thus, one will see multiple generations of affected family members, the gene mutation will be found equally in males and females, and most people who possess the mutation will have a heterozygous genotype, which means they will have a 50% risk of passing on the gene mutation to each of their offspring.

Because the relationship between the high-risk allele and the disorder is quite consistent, the corresponding tests have good clinical utility; the result of the test often influences the strategy for treating the patient. In addition, because the disorder cosegregates so consistently with the risk-increasing allele, many of these highly penetrant

Table 5.2 Highly penetrant genes for breast cancer

Gene	Syndrome	Inheritance	Protein function	Penetrance
BRCA1	HBOC	Autosomal dominant	Tumor suppressor	60–85%
BRCA2	HBOC	Autosomal dominant	Tumor suppressor	60–85%
PTEN	Cowden (part of the *PTEN* hamartoma tumor syndrome which also includes Bannayan-Riley-Ruvalcaba syndrome and Proteus-like syndrome)	Autosomal dominant	Tumor suppressor	70–80%
P53	Li-Fraumeni	Autosomal dominant	Tumor suppressor	>90%

risk-increasing alleles have been identified by means of family-based linkage studies. Because the relationship between a low-penetrance allele and the disorder is less consistent, it often requires data from a large population in order to demonstrate an association between the risk-increasing allele and the disease. Because a family-based linkage study only requires one to gather information from one or a few families, it is often easier to conduct these studies than population studies.

Germline mutations in the *TP53* tumor suppressor gene cause Li-Fraumeni syndrome (LFS). These germline mutations are relatively rare, but the *TP53* gene is one of the key gatekeepers for cell cycle maintenance and regulation in all the body's cells. Possessing even one mutant copy of *TP53* conveys a very high risk for single and multiple primary cancers. *TP53* mutations have almost complete penetrance; the lifetime risk for a primary cancer in someone with one mutant copy of the *TP53* gene is close to 90%.

Osteosarcomas and soft tissue sarcomas are the signature cancers in patients with *TP53* mutations, although early onset breast cancer, brain tumors, leukemias, lymphoma and adrenal tumors are also seen regularly. Often one may see unusually early onset of cancers associated with *TP53* mutations. Approximately 2–7% of all women who develop breast cancer before age 30 harbor a *TP53* mutation, regardless of their family history. In addition, several studies have suggested that approximately 4% of breast cancer patients who do not have BRCA1/BRCA2 mutations may have *TP53* mutations, again independent of their family history.

Cowden syndrome (CS) is caused by germline mutations in the *PTEN* gene. As a tumor suppressor gene, *PTEN* is a key regulator of cell division and apoptosis; disruption leads to both benign and malignant cellular overgrowth. Mutation carriers are at increased risk for breast and thyroid cancer (30% and 10% increase in risk, respectively). CS is also part of the more comprehensive PTEN hamartoma tumor syndrome, which also includes the Bannayan-Riley-Ruvalcaba syndrome and Proteus-like syndrome.

Clinical gene testing for LFS and CS is available through a number of laboratories (listed at GeneTests.com http://www.ncbi.nlm.nih.gov/GeneTests/), for approximately $1,200 per gene. Testing is recommended when a clinical diagnosis is suspected. You should refer to the annually updated criteria established by the National Comprehensive Cancer Network (NCCN; www.nccn.org), which has established guidelines for testing for LFS and CS. As was discussed in Chap. 3 (see Sect. 3.2), possible results include no variants, one or more deleterious variants, or one or more variants of uncertain significance. Genetic counseling must be provided to help the patient properly interpret the results and understand how the test result influences his/her overall lifetime risk of cancer.

5.6 Hereditary Breast-Ovarian Cancer Syndrome

Hereditary Breast-Ovarian Cancer (HBOC) syndrome is caused by germline mutations in either the *BRCA1* or *BRCA2* tumor suppressor genes. *BRCA1* mutations are responsible for a significant proportion (45%) of hereditary breast cancer cases, and

up to 90% when one or more cases of ovarian cancer are present in the family. An additional one-third of multiple-case breast cancer families are due to *BRCA2* mutation. Most families with four or more cases of breast cancer can be accounted for by mutation in one of the *BRCA* genes, especially if ovarian cancer is present.

In cases of HBOC, the patient's ethnicity influences the physician's strategy for *BRCA1/BRCA2* testing, because there are several "founder mutations" that appear with unusually high frequency in certain ethnic groups. There have been numerous occasions during human history when either natural or human events caused small groups of people to be uprooted from their homes and migrate in small groups to new homelands, where they were either geographically or socially isolated from other ethnic groups. These groups were relatively small, and often a significant percentage of the members of the group were related to each other. Because the effective gene pool in the group was so small, and because the group mingled little with surrounding ethnic groups, any mutations that were present in the original founder population are currently present in unusually high frequencies in the members of that ethnic group. For example, there are three founder mutations in the *BRCA* genes that are found in ~2.5% of individuals of Ashkenazi Jewish descent: the 185delAG and 5382insC mutations in *BRCA1*, and the 6174delT mutation in *BRCA2*. These three mutations make up approximately 95% of the *BRCA* mutations that are found in patients of Ashkenazi Jewish descent. Twelve percent of breast cancer cases in Jewish patients, including 28% of those diagnosed below age 50 years, have been found to carry one of the three Ashkenazi Jewish *BRCA* founder mutations.

BRCA founder mutations have been identified in several other ethnic groups, including Icelanders and Russians. In Iceland, a single *BRCA2* mutation (999del5) is found in 24% of women diagnosed with breast cancer below age 40. In Russian studies, the 5382insC mutation in *BRCA1* has been seen in up to 17.2% of patients with ovarian cancer. Because these founder mutations are present in so many members of these ethnic groups, individuals of Icelandic, Russian or Ashkenazi Jewish descent should always be tested for their respective founder *BRCA* mutations, even when another *BRCA* gene mutation has already been discovered.

Family history analysis and genetic counseling can provide more definitive estimates of risk and assist in decision making for those individuals who seek *BRCA* gene testing. According to the National Comprehensive Cancer Network (NCCN, http://www.nccn.org), the following might indicate that a person (or family) could have a significant genetic load for breast (and ovarian) cancer, and therefore be someone for whom *BRCA* gene testing has clinical utility:

- Having a known *BRCA1* or *BRCA2* gene mutation in the family
- Personal history of female breast cancer < age 45 (regardless of family history)
- Personal history of male breast cancer (regardless of family history)
- Personal history of epithelial ovarian/fallopian tube/primary peritoneal cancer
- Multiple cases of early onset breast cancer (< age 45) in a family
- Breast and ovarian cancer in the same person (or breast and ovarian cancer on the same side of the family)
- Bilateral breast cancer, especially when the first breast cancer occurred prior to age 50

In taking a family history, consider that many individuals may not know the details necessary for accurate analysis. They often may not know the specific sites of tumors, the age of onset or how the tumor was treated. This historical information needs to be verified in order to accurately assess risk, so having your patient collect medical records or death certificates, or having them sign appropriate release forms to allow access to family medical records, is important. Family histories are also dynamic, and with the passage of time, additional diagnoses may have been made. These changes in diagnosis may affect the likelihood of a hereditary cancer syndrome, so family history needs to be updated every few years.

The age of cancer onset is quite variable in female *BRCA* variant carriers. No clear explanation exists for why some carriers develop multiple primary cancers before age 50 years, while others may not develop cancer until after age 70 years, if at all. Thus, although penetrance is typically quite high (60–85%), the *BRCA* gene mutation may appear to skip generations. The mutation will be present in that generation, but the individual who possesses it will represent a case of incomplete penetrance, and not develop the disease.

Other common misconceptions include the notion that cancers in the paternal lineage need not be factored into the risk assessment. The truth is that half of all women with a *BRCA* mutation have inherited their mutation from their father. Always remember to consider the presence of ovarian cancer in a close relative an important indicator of hereditary risk, although it is not always present even in *BRCA* mutation-positive families. In addition, it is not uncommon for women in higher risk breast-ovarian cancer families to have chosen to have prophylactic surgeries as a means of lowering their risk for breast or ovarian cancer, or for other medical reasons. This may cause you to underestimate the individual's risk based on family history data. Lastly, the age of onset of breast cancer is more important than the number of women in the family who have the disease.

5.7 *BRCA* Gene Testing

The likelihood of finding a *BRCA* mutation upon testing depends upon the patient's ethnicity, the age of the patient at cancer diagnosis, the relationship of the affected family members to the patient, the presence of ovarian along with breast cancer, and the number of family members who are affected. It is important to include relatives in each lineage separately when applying risk analysis.

You should also keep in mind that limited family structure may cause you to underestimate the individual's level of genetic risk. For example, if a woman has a *BRCA* gene mutation in her family, but has an abundance of male relatives, the small number of affected individuals in that family may cause you to underestimate the probability of a *BRCA* gene mutation in that family. As mentioned above, there is an online computer program, CancerGene, which can help you compute the likelihood of a *BRCA* mutation based on personal and family cancer history information. It is advisable to have a detailed knowledge of mathematical risk models and cancer genetics to make full use of the CancerGene software, however.

If any of the following are seen in a family, referral for genetic counseling is appropriate to enable complete risk assessment, informed decision making, illuminate necessary facts and to anticipate consequences based on the results of *BRCA* testing:

Family cancer history	Probability for BRCA mutation
Breast cancer in a woman <age 50	~10%
Plus one relative with breast cancer <age 50	~20%
Plus one relative with ovarian cancer	~40%
Bilateral breast cancer or ovarian cancer in a woman plus one relative with breast cancer <age 50	~60%
Bilateral breast cancer or ovarian cancer in a woman plus one relative with breast cancer <age 50 plus one relative with ovarian cancer	~80%

Because of intellectual property issues, at this time full sequencing of the *BRCA* genes is commercially available only through Myriad Genetic Laboratories in the United States. The Comprehensive BRACAnalysis test provides full sequence analysis of the *BRCA1* and *BRCA2* genes (5,400 nucleotides and 10,200 nucleotides, respectively). It also detects five specific large genomic rearrangements involving the *BRCA1* gene. The current cost is US$3,280, and the test requires a blood or saliva sample and signed informed consent by the consumer and provider. Genetic counseling is necessary to enable informed decision making, illuminate necessary facts and to anticipate consequences based on the results of the testing. Results take approximately 2–4 weeks.

As discussed in Chap. 3 (see Sect. 3.2), the report will characterize any sequence variants that were found according to the predicted effect of that sequence variant on the corresponding protein's function. Those sequence variants that are believed to have no effect on the BRCA protein are characterized as "genetic variant-favor polymorphism." Those sequence variants that are believed to disrupt the function of the respective BRCA protein are characterized as "deleterious mutations." Finally, some sequence variants are characterized as "variant of uncertain significance." While these tests are not "negative," they cannot be considered "positive" either, because they have not detected a risk-increasing variant. If a variant of uncertain significance is detected, it does not help you estimate the individual's cancer risk, nor will anything be gained by having other unaffected family members tested for that variant. However, additional research studies may at times be initiated to see if the variant cosegregates with the disease in a given family.

If a healthy individual with a positive family history has a negative *BRCA1* and *BRCA2* gene test, it may be because the test failed to detect the mutation that was present, he/she has not inherited the *BRCA* mutation that runs in the family, or it may be that the family cancer history is not due to a *BRCA* mutation. The same could be true if an individual with breast, ovarian or other cancer has a negative test. As mutations in either *BRCA1* or *BRCA2* account for only a portion of the known hereditary and familial causes of breast cancer, one needs to consider the possibility of other underlying genetic causes for the cancer in an individual or family.

An individual's breast and other cancer risk may still be elevated above the general population, even if they have a negative *BRCA* gene test. Sometimes testing other individuals in a family, especially if there are other family members with breast or ovarian cancer, can better clarify a negative test. Other times, using the personal and family history to predict breast and other cancer risk for management is the best option.

Myriad Genetic Laboratories has developed the BRACAnalysis Rearrangement Test (BART), which serves as a supplement to the standard Comprehensive BRACAnalysis. This analysis looks for larger rearrangements (duplications, deletions) that may not be identified on the standard analysis. It is believed that the causes of an additional 1–3% of hereditary breast cancer cases may be explained by the results of the supplementary BART test, which currently is available for $750.00. This additional analysis is considered when the familial risk for a mutation is substantial.

There are two additional *BRCA* related tests offered by Myriad. The single site BRACAnalysis is performed when there is a known deleterious mutation in a family. The cost is approximately $450.00. It is a definitive test for a specified mutation. If an individual has the familial mutation, this is a true positive result, and they have the *BRCA*-associated cancer risks. As a heterozygous carrier of a *BRCA* gene mutation, they have a 50% probability of passing the mutation to each of their offspring. The converse is true if an individual does not have the familial *BRCA* mutation. This is considered a true negative result, as they do not carry the known familial gene mutation. Although other risk factors need to be considered, their cancer risks are most likely to be lower than those of their family members. Importantly, they will also not pass the familial *BRCA* mutation to any of their children.

As described above, there are founder *BRCA* mutations in a number of ethnic populations, and the standard Comprehensive BRACAnalysis will identify these mutations. However, Myriad also offers the Multisite 3 BRACAnalysis (Ashkenazi Jewish panel) for individuals of Ashkenazi Jewish descent. This test focuses on the three founder mutations that account for 95% of identifiable *BRCA* mutations in this ethnic population (note that recent research suggests that the rearrangements that are detected by the BART test are not unusually common in Ashkenazi Jews). Approximately 2.5% of Ashkenazi Jewish individuals are carriers for one of these three *BRCA* founder mutations. The cost is approximately $500, and both test application and interpretation should be taken in the context of the personal and family history.

5.8 Effects of Possessing Risk-Increasing BRCA Alleles

Women with deleterious mutations in the *BRCA* genes are predisposed to breast and ovarian cancer, typically at early ages, and can develop more than one primary cancer. Most studies estimate that the lifetime risk of developing breast cancer in female carriers of the risk-increasing *BRCA* alleles is between 46 and 85%. For female carriers

of the *BRCA1* variants, at least 14% will develop breast cancer by age 40, and 28% by age 50. In *BRCA2* carriers, the risks are at least 5% by age 40 and 13% by age 50. In female *BRCA* mutation carriers who have had breast cancer, the risk for a contralateral cancer increases by 3–5% per year to a maximum of 40–60% over the remainder of her lifetime, which is significantly higher than the risk for contralateral cancer in women with sporadic breast cancer (which increases by 0.5–1.0% per year). The risk of developing ovarian or fallopian tube cancer is also considerable, with *BRCA1* carriers having a higher lifetime risk than *BRCA2* carriers (39–63% vs. 11–31%, respectively as compared to the lifetime population risk of 1–2%).

Although the majority of *BRCA* mutation carriers develop ovarian cancer at typical ages, possessing a *BRCA* mutation does convey an increase in the risk for earlier-onset disease. A woman who possesses a *BRCA* mutation has an approximately 3% risk of developing ovarian cancer by age 40, and a 10% risk of developing ovarian cancer by age 50. In addition, these cancers are usually of the serous papillary subtype.

Male *BRCA* variant carriers also face increased cancer risks, specifically for male breast cancer and prostate cancer. Most studies show 5–6% lifetime risks for breast cancer in male *BRCA* variant carriers, a cancer that is quite rare in the general population (~1,500 cases per year in the U.S.). Most studies suggest that the lifetime risk for prostate cancer is two- to three-fold above the general population risk. There is also some increased risk in male and female *BRCA2* variant carriers for skin melanoma and pancreatic cancer, and possibly other cancers.

Because of the high penetrance of these *BRCA1* and *BRCA2* variants, these tests have significant clinical utility. An individual who possesses one or more high-risk *BRCA* alleles will benefit from more aggressive surveillance and treatment, such as earlier and more frequent mammograms and the addition of breast MRI, or preventive measures such as prophylactic surgery and chemoprevention; female *BRCA* variant carriers may benefit especially from these measures.

5.9 Cancer Screening and Prevention Measures for Female *BRCA* Variant Carriers

Given the high cancer risks, women genetically predisposed due to *BRCA* mutations may benefit from risk reduction strategies. Increased cancer surveillance is one option. For ovarian cancer risk, the role of surveillance remains limited, as available modalities (pelvic examination, serum CA-125 and transvaginal ultrasound with color flow doppler) have not been particularly effective at reducing mortality. In contrast, breast cancer surveillance with contrast-enhanced MRI in combination with physician exam, mammography and sonography is fairly effective in detecting early-stage breast tumors and saving lives.

A few studies have shown that chemoprevention with tamoxifen is effective in preventing breast cancer in *BRCA* carriers. One such study showed a 50% reduction in the risk for contralateral breast cancer in *BRCA* carriers when tamoxifen was

given as treatment for the initial breast cancer. Any benefit of tamoxifen, however, must be weighed against the possibility of adverse side effects associated with long-term use. Similarly, oral contraceptive use is associated with a significant risk reduction for ovarian cancer in *BRCA* carriers (*BRCA1* odds ratio 0.56; *BRCA2* odds ratio 0.39). Concern remains that long-term use of oral contraceptives may lead to a modest increase in the risk of breast cancer. However, much of the data that support this claim come from studies of women taking oral contraceptives before 1975, when the hormone levels that were contained in these preparations were significantly higher than those that are administered today.

The most effective risk-reducing strategy is prophylactic surgery. Prophylactic bilateral salpingo-oophorectomy (PBSO) has become widely accepted as an effective means of reducing ovarian cancer risk in *BRCA* carriers. Some studies suggest that PBSO may reduce the risk for ovarian cancer by as much as 95–98% (although there remains a risk for peritoneal cancer). Typically, this surgery is offered by age 35 or after childbearing is complete in *BRCA* carriers.

Given the higher likelihood of occult cancer in *BRCA* mutation carriers, the American College of Obstetricians and Gynecologists emphasizes the need for pathology examinations. This process includes thorough visualization of the peritoneal surfaces with pelvic washings, as well as the removal of all tissue from the ovaries and fallopian tubes, with complete serial sectioning and microscopic examination for occult cancer. Importantly, it has been shown that PBSO can lower the risk of breast cancer by 56% and 46% in *BRCA1* and *BRCA2* carriers (respectively), with the greatest effect in women who had PBSO before age 40.

Prophylactic mastectomy also significantly reduces breast cancer risk (>90%). Women who undergo both PBSO and prophylactic mastectomy benefit most (95% risk reduction). Although prophylactic surgery remains the most effective means of risk reduction for women with high risk, it is invasive and irreversible and can have adverse physical and psychosocial side effects, including issues of body image and quality of life.

5.10 *BRCA* Mutation Positive Case Study

Mrs. C has been coming to your internal medicine practice for 6 years. She is a college educated, Internet-savvy mother with two daughters, ages 19 and 21. When she turned age 40 earlier in the year, she underwent a baseline mammogram, which was normal. However, she is back in your office, for the second time in the last 3 months, because there is a worrisome palpable mass in her left breast on exam. Mrs. C states that the area has been tender. Mrs. C has used the Family Healthware tool to construct her family tree and forwarded you a copy by email for review (Fig. 5.1). You are aware of her significant family history of breast cancer, primarily on her father's side of the family.

You review the family history again with Mrs. C today and note that she had a paternal aunt who died of breast cancer in her 40s. In addition, according to family

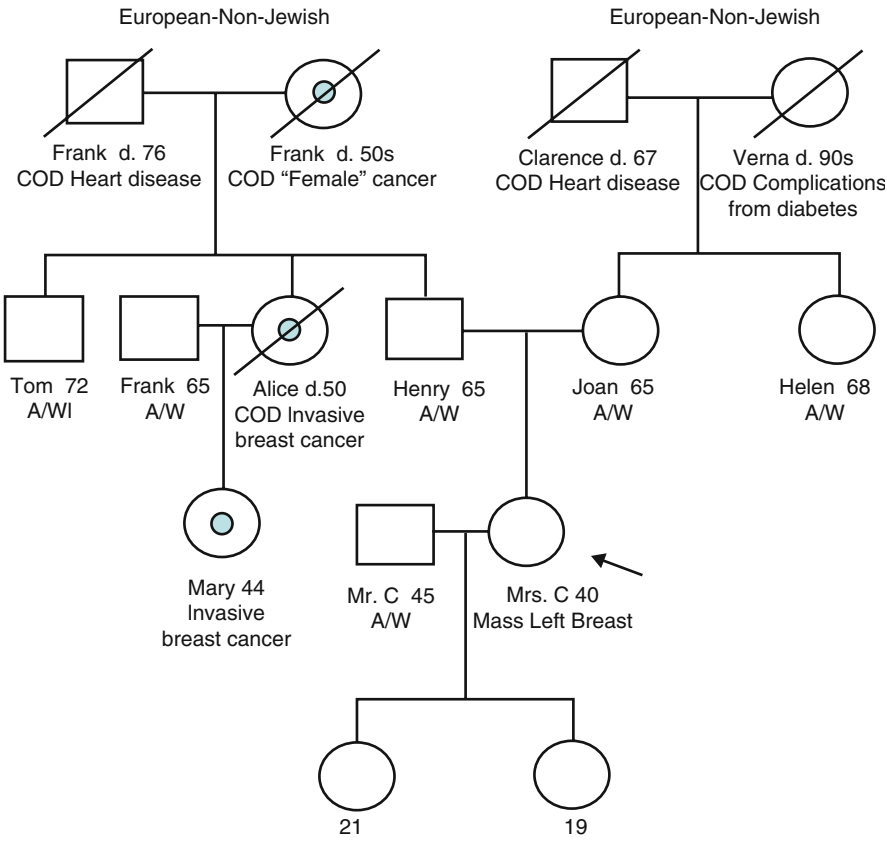

Fig. 5.1 Pedigree illustrating the members of Mrs. C's family, along with the age and cause of death (COD), where applicable. *A/W* alive and well

lore, she had a paternal grandmother who died in her 50s of a "female" cancer. Mrs. C also mentions that she had recently heard that a cousin on her father's side of the family, Mary, is currently undergoing treatment for breast cancer. The family ancestry is European, and not Jewish. You recall that some of the key risk factors for hereditary breast-ovarian cancer syndrome (HBOC) include multiple generations of affected women and earlier than expected age of cancer onset (40s). You suspect that the "female" cancer in the grandmother may be relevant. The fact that the family history presents from the paternal lineage should not be dismissed, as *BRCA* mutations can be passed in an autosomal dominant manner through either the maternal or paternal lineage.

You are also aware that Mrs. C has an extended history of ovarian cysts and endometriosis. She has been on low-dose oral contraceptives to help regulate menses for more than 10 years, adding another, albeit less significant, breast cancer risk factor. Given your concern for her significant family cancer history, as well as these

additional risk factors, you decide to make a referral to a genetic counselor to give her the best possible estimate of her cancer risk and to see if genetic testing might be appropriate. You also make a referral to a breast cancer specialist for additional workup for the palpable breast mass.

5.10.1 Mrs. C's Initial Meeting with the Genetic Counselor

During Mrs. C's first appointment with the genetic counselor, the counselor obtained a detailed medical history and four-generation pedigree, probing in more detail the history of those family members with cancer. A detailed history of environmental risk factors for cancer was also taken. Using the information, the counselor provided Mrs. C with an in-depth analysis of her cancer risk. In addition, the counselor and Mrs. C discussed the psychological impact of finding the lump in her breast, as well as the different feelings Mrs. C experienced as she pondered her family history and thought about the possibility of having breast cancer herself. She was very close to her paternal aunt, and had experienced much agony in watching her battle this disease. It became apparent during the genetic counseling session that Mrs. C is not only concerned about her own risk for breast cancer, but also her two daughters' risks as well. She had mixed emotions in talking about her daughters, hoping on one hand that they could be spared from having to face higher cancer risks themselves, but then also not wanting to have them concerned about her risks, and the possibility that she might even now have breast cancer. She had in fact not told anyone in her family (including her husband) about coming to the genetic counseling session that day.

Mrs. C had read about *BRCA* gene testing on the Internet, and hoped it could provide concrete information about her risk and her daughters' risks. The counselor explained to Mrs. C that, while the discovery of a *BRCA* mutation could indeed provide tangible information for her and her daughters, *BRCA* testing had several important limitations. The counselor explained that, based on a CancerGene analysis of the family history of breast cancer alone, there was a rather low (5–14%) likelihood that Mrs. C would have a *BRCA* mutation. If Mrs. C's paternal grandmother did indeed have ovarian cancer, however, the risk for a *BRCA* mutation would increase to 35%. HBOC is an autosomal dominant condition; if through testing Mrs. C is found to be heterozygous for a risk-increasing *BRCA* allele, she would also possess a typical-risk allele along with the risk-increasing allele. Therefore, the probability she will pass down the risk-increasing allele is 50% for each of her two daughters.

The counselor also explained that the *BRCA* tests do not always provide one with useful information. The counselor explained that approximately 10% of people who have comprehensive *BRCA1/BRCA2* testing are reported to possess a genetic variant of uncertain significance, which constitutes an uninterpretable result. The counselor also informs Mrs. C that even if *BRCA* analysis does not reveal any risk-increasing alleles in *BRCA1* or *BRCA2*, her family could conceivably possess a mutation in still undiscovered breast cancer genes.

Mrs. C and the genetic counselor also discussed the cost of testing. Some of the larger health insurance plans now provide coverage for *BRCA* testing when a participant meets certain eligibility criteria, such as having a diagnosis of breast cancer under age 45 or if there is a known mutation in the family. Because Mrs. C is unaffected, it is possible her insurance company will not pay for these tests, which cost up to $3300.00.

Because Mrs. C's risk for carrying a *BRCA* mutation is somewhat low, and the cost of testing is so high, the genetic counselor advises Mrs. C that testing should begin in a family member who has had breast or ovarian cancer, and ideally in the person with the youngest age of onset. In this case, this would be the newly diagnosed cousin, Mary. If cousin Mary gets tested and does not have a *BRCA* gene mutation, Mrs. C would still face an elevated risk for breast cancer given her significant and as yet unexplained family cancer history. This risk would be in the more moderate, 20–30% lifetime range, however, rather than the 60–85% lifetime risk that is associated with a *BRCA* gene mutation. However, if cousin Mary was tested, and was found to have a deleterious *BRCA* gene variant, then Mrs. C can have a very powerful predictive test that shows that she either does or does not have the family's *BRCA* gene mutation. This would be the most informative and cost-effective test.

Mrs. C. is given a record release form to obtain medical records on cousin Mary, who has had breast cancer. She plans to speak with Mary, and see if she is interested in meeting with a genetic counselor to discuss the potential benefits and limitations of *BRCA* analysis. Mrs. C's genetic counselor provides the contact information for a genetic counselor whose office is convenient to Mary in the consult note, a copy of which is sent to you and the patient. The consult note also requests that if Mary does decide to have the testing performed, a copy of the report is sent to you and the family. The genetic counselor will then follow up with Mrs. C once Mary's records have been received and reviewed.

Shortly thereafter, you receive a phone call from the breast cancer specialist reporting that a biopsy was performed on the left breast mass, which showed benign findings only. However, given the significant family cancer history, the plan is to continue follow-up with the breast specialist with clinical exams and breast imaging studies.

5.10.2 Interpreting the Test Results and Following Them Up

Cousin Mary decides to see the local genetic counselor and has comprehensive *BRCA* testing performed, which finds a IVS20 + 1delG *BRCA1* gene mutation. This sort of mutation is expected to disrupt the process of RNA splicing (see Sect. 1.6.3), causing the gene to produce a nonfunctional protein, and is definitely a risk-increasing allele. In fact, this allele is reported to have at least 60–85% penetrance for breast cancer; any woman who possesses this allele has a 60–85% chance of developing breast cancer in her lifetime. The IVS20 + 1delG allele also has 39–63% penetrance for

ovarian cancer; a woman who possesses this allele has a 39–63% risk of developing ovarian cancer in her lifetime as well. Mary opts to release her *BRCA1* test report to the family, and contacts Mrs. C directly.

Mrs. C makes an appointment with her genetic counselor, to provide a blood sample and informed consent to initiate genetic testing. Letters of medical necessity facilitate full insurance coverage for single-site analysis of the known familial mutation. Mrs. C. and her husband meet with the genetic counselor in person when the test results come back, and learn that Mrs. C does, indeed, have the same IVS20 + 1delG *BRCA1* allele that Mary has. The genetic counselor explains that, because the IVS20 + 1delG allele has 60–85% penetrance for breast cancer, and a 39–63% penetrance for ovarian cancer, Mrs. C has a 60–85% chance of developing breast cancer in her lifetime, and a 39–63% risk of developing ovarian cancer in her lifetime.

Although Mrs. C is not surprised by the result, she finds it difficult to fully incorporate this information into her life and that of her family. She is concerned about her high risk for cancer, but is also relieved to finally have concrete information to consider. Given the recent scare with the mass in her left breast, she feels more comfortable knowing that she will continue to be followed by a breast specialist. She also feels relieved to finally know why her aunt and other family members died from cancer, and feels empowered because she is able to make truly informed decisions regarding her cancer screening and prevention plan.

Mrs. C's husband is very supportive, and during the one-hour counseling session there was much discussion of how to move on from here, especially with respect to when the most appropriate time would be for Mrs. C's two daughters to get tested. The counselor informs Mrs. C that, because only one of her two copies of the *BRCA1* gene has the risk-increasing allele, each of her two daughters has a 50% chance of possessing the risk-increasing allele herself. Because her children are ages 19 and 21 respectively, and as the cancer risk in *BRCA* carriers begins in early adulthood, Mrs. C and her husband plan to inform their children of these risks and the availability of genetic counseling and testing. Each daughter can then decide whether she wants to know if she possesses the risk-increasing allele or not. If she does, she will meet with a genetic counselor when she is ready to have *BRCA* testing performed. If one of the daughters chooses not to have the testing performed, given the significant cancer risks associated with *BRCA* mutations, she will be advised to have annual clinical breast examinations until age 25, and thereafter initiate high-risk cancer screening, acting as she were a *BRCA* carrier unless proven otherwise. This would include clinical breast exams every 6 months and appropriate breast imaging per NCCN guidelines.

In addition, Mrs. C's test results also suggest that her father has the risk-increasing allele. Her father is designated an "obligate carrier" of the mutation, because it is so highly likely that he is the parent from whom Mrs. C inherited the risk-increasing *BRCA1* allele. *BRCA1* mutations have been associated with a slightly increased risk for prostate cancer in male carriers, and also likely increase the risk for male breast cancer, although these risks have not yet been clearly delineated.

Lastly, the genetic counsellor recommends that Mrs. C inform her more distant paternal family members of her *BRCA* test result, because several of her other family members may also possess the *BRCA* mutation, and could therefore benefit from this knowledge.

5.10.3 The Plans for Mrs. C and Her Family Members

With respect to Mrs. C's health care plan, the genetic counselor informs her that the standard of care for women who have a high genetic risk for breast cancer due to *BRCA* mutation is annual breast imaging, including mammograms and magnetic resonance imaging (MRI) studies. In addition, Mrs. C is told that she should have a clinical breast examination by a physician every 6 months, and perform monthly breast self-examinations. It is beneficial that Mrs. C. is already seeing a breast specialist who can provide this more comprehensive screening and preventive health care.

If one of Mrs. C's daughters is also found to have the IVS20+1delG *BRCA1* allele, she will also be advised to follow this preventive health care plan. As Mrs. C's father is an obligate carrier, he should have a chest examination by a physician every 6 months, and perform monthly chest self examinations. He should continue with routine prostate cancer screening, including annual digital rectal examination and prostate-specific antigen (PSA) test beginning at age 50 (or 10 years earlier than the earliest diagnosis of prostate cancer in the family).

The genetic counselor also informs Mrs. C that she might want to consider taking a preventive medication to reduce the risk of developing breast cancer. Both tamoxifen (sold as Nolvadex, Istubal, and Valodex) and raloxifene (Evista) may be effective in preventing breast cancer in women who possess risk-increasing alleles in either the *BRCA1* or *BRCA2* genes. The potential benefits of these drugs must be weighed against their side effects, however. Further discussion on the use of these preventive medications will be done through the breast specialist that Mrs. C is currently seeing.

With respect to Mrs. C's risk for ovarian cancer, the genetic counselor informs Mrs. C that her prior use of birth-control pills may have reduced her risk of ovarian cancer. With respect to the decision whether to undergo regular screening tests for ovarian cancer, however, because the current screening procedures have significant limitations, Mrs. C. is advised to consider the possibility of prophylactic bilateral salpingo-oophorectomy (PBSO). This surgery will lower her risk for developing ovarian cancer by approximately 80–90%, and will also lower her risk for developing breast cancer by approximately 50% as well.

Finally, the genetic counselor informs Mrs. C that there are several support group and Internet websites for women with risk-increasing alleles in the *BRCA* genes. For example, groups such as Facing Our Risk For Cancer Empowered (FORCE, http://www.facingourrisk.org/) and BE BRIGHT PINK (http://www.bebrightpink.org/)

provide educational materials and psychological support for women who have a high risk for breast and/or ovarian cancer.

A results note detailing the issues that were discussed in the genetic counseling session is then sent to you and the patient. The genetic counselor includes the recommendations for screening and prevention in the note.

5.10.4 Further Developments

The following year Mrs. C is diagnosed with an early-stage invasive ductal carcinoma. The tumor is 1.2 cm in size, and was found in the left breast. The pathologist's report noted that the tumor tissue did not express either the estrogen, progesterone or HER2/neu receptors, a histological feature often seen in *BRCA1*-associated breast tumors. Mrs. C is informed that she has a 40–60% risk of developing a second primary breast tumor given her mutation status. Because of the high risk for a new primary tumor, Mrs. C opts for bilateral mastectomy as treatment rather than breast conservation surgery.

5.11 Colon Cancer Gene Variants with Low Penetrance

Colorectal cancer, also called colon cancer or large bowel cancer, is the fourth most common form of cancer in the United States, and the third leading cause of cancer-related death in the Western world. It causes 655,000 deaths worldwide per year. The lifetime risk of developing colon cancer in the United States is approximately 7%. Most colorectal cancers are thought to arise from adenomatous polyps in the colon, which tend to occur more often with age. In fact, most cases of colon cancer occur in the 60s and 70s, while cases before age 50 are uncommon unless due to a strong genetic influence. Other risk factors include, but are not limited to, smoking, high-fat diet and physical inactivity. History of inflammatory bowel disease is a risk factor; the risk varies inversely with the age of onset of the colitis and directly with the extent of colonic involvement and the duration of active disease. Colorectal cancer risk is also increased in patients with Crohn's disease and ulcerative colitis.

Genome-wide association studies have identified ten colon cancer risk-increasing alleles that have low penetrance (Table 5.3). In many cases, it is uncertain whether the polymorphism in Table 5.3 is the critical functional polymorphism, or merely linked to the critical functional polymorphism(s). Interestingly, five of these SNPs are in linkage disequilibrium blocks that include or are near genes (*SMAD7*, *GREM1*, *BMP2* and *BMP4*, and *RHPN2)* that are part of the TGF-beta superfamily, previously implicated in colon tumorigenesis. Some commercial laboratories' GWA tests include these polymorphisms, but little is known of their true value in predicting colon cancer risk.

Table 5.3 Low-penetrance SNPs that affect risk for colorectal cancer

dbSNP No.	Gene symbol and name	Risk-allele frequency	Odds ratio
rs6983267[a]	*POU5F1P1*; POU class 5 homeobox 1B *DQ515897*; Gene with unknown function *MYC*; myelocytomatosis viral oncogene homolog (avian)	0.51	1.21
rs4939827	*SMAD7*; Mothers against decapentaplegic, drosophila, homolog of, 7	0.52	1.18
rs4779584[a]	*SCG5I;* Secretogranin V *GREM1*; Gremlin 1, cysteine knot superfamily, homolog *FMN1*; Formin 1	0.18	1.26
rs 3802842	*LOC120376*; Uncharacterized	0.29	1.12
rs16892766	*c8orf53*; Uncharacterized	0.07	1.25
rs10795668	*FLJ3802842*; Uncharacterized	0.67	1.12
rs4444235	*BMP4*; Bone morphogenetic protein 4	0.46	1.11
rs9929218	*CDH1*; Cadherin-1	0.71	1.10
rs10411210	*RHPN2*; Rhophilin, Rho GTPase binding protein 2	0.90	1.15
rs961253	*BMP2*; Bone morphogenetic protein 2	0.35	1.12

[a]SNP located in close proximity to a cluster of tightly linked genes

5.12 Highly Penetrant Colorectal Cancer Gene Variants

At least 5–10% of all colorectal cancers are the result of a single highly penetrant inherited cancer gene mutation. A number of syndromes have been described with a general breakdown based on polyp histology, polyp number and family cancer history (Table 5.4). The most common are Lynch syndrome (formerly known as hereditary non-polyposis colorectal cancer, HNPCC) and familial adenomatous polyposis (FAP), which together account for ~3–5% of all colon cancer cases. There are also rarer syndromes, typically involving hamartomatous polyps (e.g. juvenile polyposis, Peutz-Jeghers syndrome), for which gene testing is available. Most of these colon cancer syndromes are associated with an autosomal dominant pattern of inheritance; the exception is *MYH*-associated polyposis, which exhibits an autosomal recessive pattern of inheritance.

5.13 Lynch Syndrome

Lynch syndrome (LS) is the most common form of hereditary colorectal cancer, affecting as many as 1 in 400 individuals. Most Lynch syndrome patients have a germline mutation in one of four DNA mismatch repair genes (*MLH1, MSH2, MSH6, PMS2*); mutations in *MLH1* and *MSH2* are the most common. Penetrance is high and the phenotype is quite variable within families. Individuals with LS have a 50–100% chance of developing colorectal cancer by age 70, although early and frequent polypectomy are effective in prevention.

Table 5.4 Highly penetrant genes for colon cancer

Gene	Syndrome	Inheritance	Protein function	Penetrance
MLH1	Lynch	Autosomal dominant (rarely recessive)	Tumor suppressor	At least 80%
MSH2	Lynch	Autosomal dominant (rarely recessive)	Tumor suppressor	At least 80%
MSH6	Lynch	Autosomal dominant	Tumor suppressor	60–80%
PMS2	Lynch	Autosomal dominant and recessive	Tumor suppressor	Dependent on mode of inheritance
APC	Familial adenomatous polposis	Autosomal dominant	Tumor suppressor	80–90%
MYH	MYH-associated polyposis	Autosomal recessive	Tumor suppressor	80–90%
STK11 (LKB1)	Peutz-Jegher syndrome	Autosomal dominant	Tumor suppressor	80–90%
SMAD4	Juvenile polyposis	Autosomal dominant	Tumor suppressor	80–90%
BMPR1A	Juvenile polyposis	Autosomal dominant	Tumor suppressor	80–90%

The colorectal cancer of LS differs from sporadic colorectal cancer in several ways. Average age of onset is much earlier (average age is 45 years), and there is a paucity of adenomatous colonic polyps compared to other familial polyposis conditions. Typically one may see a few adenomas (possibly dozens over the lifetime), rather than hundreds to thousands in the adenomatous polyposis syndromes. The proximal or "right" colon is the preferred site (60–70%), and there is significant risk for synchronous and metachronous cancers. The progress from adenoma to carcinoma also occurs more rapidly in LS (2–3 years) than in sporadic tumors, in which it is believed to take approximately 5–10 years. In patients with LS, these tumors are often poorly differentiated on histology, with distinct mucoid and signet-cell features, and noted presence of infiltrating lymphocytes. Despite this more rapid growth pattern and the adverse histologic features seen on pathologic analysis, these lesions are associated with a better prognosis than sporadic colon tumors, and may be more amenable to targeted therapeutic interventions.

For both men and women who develop colorectal cancer, subtotal colectomy may be indicated, as the risk for developing a second primary colorectal cancer is 30% within 10 years after the original surgery, and 50% within 15 years. Other significant lifetime cancer risks include endometrial (40–60%), ovarian (9–12%) and stomach (13–19%), as well as small intestine, pancreas, brain, hepatobiliary tract and urinary tract cancers, for which the risk is in the range of 2–4%. It should be noted that LS-associated uterine cancers have a mean age at diagnosis of 40–60 years, while LS-associated ovarian cancers have a mean age of diagnosis of 43 years, with ~30% diagnosed before age 40. This is considerably earlier than the average age of diagnosis for non-LS-associated ovarian cancers, which is in the mid-60s.

The phenotypic variability observed in LS has given rise to several variants of LS, each with its own characteristic phenotype. Sebaceous gland tumors, multiple keratoacanthomas, basal cell carcinomas, and possibly breast cancer may be more common in the Muir-Torre variant of LS. In contrast, the Turcot variant should be considered when the family history includes glioblastoma mutiforme. The molecular basis of most cases of Turcot syndrome is either a mutation in the *APC* gene (described in the next section on familial adenomatous polyposis), or a mutation in the mismatch repair genes associated with LS. A few patients have been reported to be homozygous for mutations in *PMS2*; these patients are rare, but severely affected. Affected children have onset of colon cancer prior to the second decade of life. Hematologic cancer, brain tumors and café-au-lait macules have also been reported.

Assessment of family history is invaluable in identifying individuals with Lynch syndrome. The International Collaborative Group on Lynch Syndrome established diagnostic criteria based on personal and family cancer history patterns, known as the Amsterdam criteria, in 1991. These criteria were further modified in 1999 (Amsterdam II Criteria), in an attempt to incorporate extracolonic cancers. However, even for individuals who met criteria for LS, no more than 50% were found to have mutations in the associated mismatch repair genes. For this reason a third set of criteria, the Bethesda criteria, were developed. These criteria are more sensitive, especially in identifying LS families with pathogenic mutations. However, the specificity of the Bethesda Criteria is significantly lower than that of its counterparts; the use of the Bethesda criteria results in more false positives than the use of the Amsterdam II Criteria does.

5.14 Molecular Genomic Testing in Patients Suspected of Having Lynch Syndrome

5.14.1 Microsatellite Instability (MSI) Is a Hallmark of Lynch Syndrome

There is some disagreement as to how best to work up an individual who is suspected of having LS. Referral for genetic consultation and testing is always appropriate, as the workup and results interpretation can be complex. The vast majority of cases of LS are caused by mutations in one of four different genes: *MLH1, MSH2, MSH6* and *PMS2*. These are mismatch repair (MMR) genes, whose proteins correct DNA replication errors caused by microsatellite-induced DNA polymerase slippage. A defect in the MMR mechanism causes microsatellite instability (MSI, see Sect. 4.10). MSI is a hallmark of LS-associated tumors; approximately 90–95% of LS-associated colorectal tumors will be MSI positive, versus 15% of "sporadic" colorectal tumors (more on this matter below). Because observing MSI is easier, faster and less expensive than screening genes for mutations,

it is often wise to confirm the diagnosis of LS by observing MSI and/or performing immunohistochemistry (IHC) analysis before embarking on a search for mutations in the LS-associated MMR genes.

5.14.2 Some Labs Offer Complementary MSI and IHC Analyses

A number of pathology laboratories now use IHC analyses to screen incident colorectal tumors for MMR proteins. IHC analyses afford the extra advantage of being able to detect situations in which the expression of these proteins is suppressed by epigenetic changes, as has been seen in patients who have aberrant methylation of *MLH1*. MSI can also be used as a complementary tumor marker test, and some laboratories offer both IHC and MSI as part of their screen. For IHC, the loss of protein indicates a high likelihood of mutation in specific MMR genes, and testing should begin with the gene whose protein product is absent. MSH2 and MSH6 are complementary binding proteins, as are MLH1 and PMS2. When there is a mutation that prevents one gene from making a functional protein, both members of that binding set are usually absent on IHC (e.g. MSH2 and MSH6). Because mutations in *MLH1* and *MSH2* are much more common than mutations in *MSH6* or *PMS2*, when a pair of complementary binding proteins is absent, gene testing should begin with *MLH1* or *MSH2*, and proceed to the second gene if nothing is found.

If all four proteins are present in the tumor and the tumor is also MSI negative, it becomes highly unlikely that the patient has LS. There is probably another underlying genetic cause of the cancer in the family, and genetic testing of the MMR genes would not be indicated. Depending on the patient's age at diagnosis and family history, the physician may elect to refer the patient to a clinical genetic professional.

Going directly to MMR gene testing is also an option if there is a high suspicion for LS, or if an individual meets the Amsterdam or Bethesda criteria. Most often individuals with LS will have a mutation in either the *MLH1* or *MSH2* gene, less often in *MSH6* and *PMS2*. There are a number of laboratories that offer LS gene testing, with the cost approximately $1,200 per gene. As we discussed in Chap. 3 (see Sect. 3.2), the report will list any variants that were found, and state whether each variant is believed to be a risk-increasing allele, a benign polymorphism or a variant of uncertain significance.

Lastly, there are subsets of colon tumors where one will find MSI-High status and/or loss of MLH1 expression due to hypermethylation of the promoter region of the *MLH1* gene. In these tumors one often finds a somatic mutation in the V-raf murine sarcoma viral oncogene homolog B1 gene (*BRAF*). Some of these MSI-High, BRAF-positive tumors will be sporadic in nature, while others may be due to other hereditary cause, such as the more recently described hereditary serrated neoplasia/hyperplastic polyposis syndrome.

5.15 Cancer Screening and Prevention Measures for LS Mutation Carriers

It is recommended that individuals who carry mutations in *MLH1*, *MSH2* and *PMS2* undergo colonoscopy every 1–2 years starting at age 20–25 (age 30 years in *MSH6* families), or 10 years earlier than the youngest age of colon cancer diagnosis in a family member. If colon cancer is found, the whole colon is often removed, but annual screening should continue for rectal cancer. Female LS carriers should have annual endometrial biopsies, transvaginal ultrasound and CA-125 screening for endometrial and ovarian cancer beginning at age 30. Prophylactic hysterectomy with bilateral salpingo-oophorectomy can be considered to reduce the likelihood of developing endometrial and ovarian cancer. To screen for other LS-associated cancers, urinalysis with cytology is done every 1–2 years beginning at age 25–35 years. Lastly, examination for gastric cancer is done if this has occurred in the family, with gastroscopy every 1–2 years, starting at age 30.

5.16 Familial Adenomatous Polyposis

Familial adenomatous polyposis (FAP) is the second most common form of hereditary colon cancer, and accounts for approximately 1% of hereditary cancer cases. It is characterized by the development of hundreds to thousands of adenomatous polyps throughout the colon and rectum, with an extremely high lifetime risk of colon cancer. It is an autosomal dominant condition caused by germline mutations in the adenomatous polyposis coli (*APC*) gene; it affects about 1 in every 5,000 people. While most individuals (75–80%) with FAP will have an affected parent, quite a few are the result of a new (*de novo*) mutation, and therefore the first affected person in their family. While penetrance is almost complete, there is significant variability within a family with respect to the age of diagnosis and the number of adenomatous polyps present.

The clinical diagnosis of FAP is made if an individual has greater than 100 colorectal adenomas. Seventy-five percent of individuals with FAP will develop polyps by age 20, and nearly 90% will develop them by age 30. Although the polyps are initially benign, at least a subset will proceed to malignancy, so colonoscopic surveillance should begin at an early age. Prophylactic colectomy eventually becomes necessary, as 90% of untreated individuals with FAP will have malignancy by age 50. Historically, the presence of both colonic and extracolonic features has been referred to as Gardner syndrome, which we now know is due to mutations in the same gene as is mutated in FAP.

There are a number of extracolonic features of FAP to note. One is congenital hypertrophy of the retinal pigment epithelium (CHRPE), which is seen in 60–88% of affected individuals. Additional features include dental anomalies (unerupted

teeth, congenital absence of one or more teeth, supernumerary teeth, dentigenous cysts and odontomas), epidermoid cysts (notably on the scalp) and osteomas of the mandible, skull, fingers, toes and long bones. Soft tissue desmoid tumors typically develop in the abdominal cavity, but can occur anywhere in the body, and occur in approximately 5–10% of children and adults with FAP. Lastly, polyps of the upper gastrointestinal tract, including the gastric fundic glands, antrum and duodenum can also be present.

Individuals with FAP may develop extracolonic malignancies. Gastric adenomas and fundic gland polyps are common, affecting 10% and 50% of patients, respectively, although the associated risk for cancer is small. In contrast, adenomas invariably form in the second and third portions of the duodenum (>90%), especially in the periampullary region, and present an increased risk for malignancy. Papillary thyroid carcinoma occurs in approximately 2% of individuals with FAP, more often in female carriers. Other cancers observed in FAP include hepatoblastoma in children (risk of 1–2% to age 6 years) and pancreatic carcinoma (approximately 2% lifetime risk).

There is a milder clinical variant of FAP known as Attenuated FAP (AFAP), caused by mutations in the same gene as classic FAP, but characterised by fewer polyps (between 10 and 100, with an average of 30) and a later age of colon cancer diagnosis (50–55 years). The adenomas in AFAP tend to aggregate more often in the proximal colon, although they can be present anywhere in the colon and in the upper gastrointestinal tract. The lifetime risk for colon cancer and upper GI polyps in AFAP is similar to that seen in classic FAP

5.17 *APC* Mutation Screening in Familial Adenomatous Polyposis

Germline mutations in the adenomatous polyposis coli (*APC*) gene cause FAP and its variants. The gene contains 15 exons, with the last exon comprising more than 75% of the coding sequence. The normal protein product is a tumor suppressor protein that has a multitude of important functions in the cell, including cell cycle control, differentiation, migration and apoptosis. The role of APC in many of these processes occurs through its interactions with beta-catenin in the WNT signaling pathway. APC binds to beta-catenin and targets it for degradation by the cell. This, in turn, prevents beta-catenin from entering the nucleus and interacting with the Tcf family of transcription factors. APC also has a role in kinetochore-microtubule attachment, chromosome segregation at mitosis and stabilization of microtubule-kinetochore interactions. Thus, inactivation of *APC* may contribute to chromosomal instability and an enhanced mutation rate, promoting tumor growth in colorectal cancer.

Screening the *APC* gene can identify mutations in 50–80% of individuals with classic FAP. A number of laboratories offer *APC* gene analysis at a cost of USD$1,200, providing results in approximately 2–4 weeks. Nearly 30% of individuals with FAP have no family history of the disease, and will have *de novo APC* gene mutations. Over 500 distinct germline mutations of *APC* have been described.

Protein-truncating mutations between codons 169 and 1,393 result in the classic FAP phenotype. The two most common germline mutations, 5-bp deletions in codons 1,061 and 1,309, occur in the mutation cluster region of exon 15, and correlate with a high number of adenomas at an early age. Retinal lesions are commonly associated with alterations between codons 463 and 1,387. Desmoid tumors and mandibular osteomas, as seen in the Gardner variant, occur with mutations between codon 1,403 and 1,578.

In contrast, the AFAP phenotype results from mutations either in the 5' part of the *APC* gene (5' to codon 158), exon 9, or in the 3' part of the gene beyond codon 1,595. The mutation detection rate in patients with AFAP is much lower (~30%) than in patients with the classic form of FAP (50–80%).

For patients with either classic FAP or AFAP, a negative *APC* test result does not rule out the diagnosis. Most labs offering *APC* gene analysis also provide analysis of the *MYH* gene, as *MYH* mutations account for a subset of classic FAP and AFAP patients. *MYH*-associated polyposis typically presents with an affected sibship in the family. Mutations in *MYH* are recessive, which means that an affected individual has inherited a mutation from each of his/her parents. Because both parents carry the mutation, if neither is affected (both are heterozygous for the mutant and normal alleles), each parent has a 50% probability of passing down the mutant allele, and there is a 25% probability they will produce an affected child on any given pregnancy. If a parent is affected, that parent is homozygous for the mutation, and will pass the mutant allele down to every child he/she has. If one parent is affected, there is a 50% probability they will produce an affected child on any given pregnancy. If both parents are affected, they will both pass down a mutant allele to every child, because they do not have any normal alleles to pass down.

If a patient has a negative test for *APC* and *MYH*, he/she should be managed based on the clinical diagnosis, which should include a thorough evaluation of the polyp pathology. Some of the rarer forms of polyposis (ex. hamartomatous polyposis, Table 5.4) may also have a component of adenomatous polyps. These rarer forms of polyposis should be ruled out, because the risks for cancer and the plan for monitoring and management differ for the different forms of polyposis. In addition, it is appropriate to have genetic testing performed on any children who are suspected of having any form of polyposis, because test results could affect management during childhood.

5.18 Cancer Screening and Prevention Measures for APC Mutation Carriers

Individuals with FAP should receive annual sigmoidoscopy beginning at 10–11 years of age. Discussion of prophylactic colectomy should begin in adolescence, because the colon will need to be removed once adenomas are found (usually between ages 17 and 20). Depending on how much of the rectum is left after surgery, the patient may need proctoscopy every 6 months to check the rectum for polyps. Polyps in the

ileum are frequent, leading to a recommendation for ileoscopy every 3–5 years after colectomy. Individuals with FAP should also undergo upper endoscopy with side view by age 25, and then every 1–5 years depending on the number and size of polyps found. To screen for thyroid cancers, physical examination and ultrasound have been suggested, especially for young women with FAP. Screening for hepatoblastoma (liver cancer) in children should include liver palpation, serum alpha-fetoprotein blood test and/or liver ultrasound each year up to age six.

5.19 Personalizing Drug Therapy for Cancer Patients

5.19.1 Gene Expression Assays can Classify Cancers into Molecular Subtypes

Personalized medicine has already made a significant impact in cancer care, where individuals have been making health management decisions with their physicians based on single gene tests and molecular biomarkers for two decades now. The discovery in the mid-1980s of the role that the human epidermal growth factor receptor 2 (HER2/neu) plays in breast cancer is the classic example of how a specific biomarker could revolutionize cancer treatment. The HER2 receptor is produced in excess in approximately one-fourth of all invasive breast cancers. Because anti-HER2/neu therapy benefits only individuals with invasive breast carcinomas over-expressing *HER2/neu*, testing is used to identify individuals most likely to respond to anti-HER2/neu therapies. HER2/neu receptor status is commonly determined by testing for gene amplification (using fluorescence in situ hybridization, aka FISH) or protein over-expression (by immunohistochemistry). The receptor is targeted by Herceptin (trastuzumab), which acts to halt the growth of the tumor, and can actually reduce the risk for tumor recurrence by ~50% when used in combination with chemotherapy, as compared to chemotherapy alone.

Some breast cancers are associated with over-expression of the *ERBB2* gene, which codes for human epidermal growth factor receptor 2. Copy number variations (and other types of mutations) can cause over-expression of *ERBB2,* leading to more aggressive forms of breast cancer. Measuring the *ERBB2* copy number can provide a diagnostic tool for breast cancer and other cancers.

The Onco*type* DX® test (Genomic Health, Inc.) uses the reverse transcription-polymerase chain reaction technique to assess the level of expression of *HER2* as well as 20 other cancer genes, to determine whether women with certain types of breast cancer are likely to benefit from chemotherapy. The test measures the expression of these genes within the tumor and provides a score that places the individual into one of three categories: low, intermediate or high risk of having a tumor recurrence within 10 years. Individuals with a low risk of tumor recurrence may be treated successfully with hormone therapy alone, avoiding the expense and toxic effects of chemotherapy, while those with a high risk of recurrence may benefit from more aggressive treatment.

It should be noted that the Onco*type* DX test has been validated for newly diagnosed breast cancer patients with stage I or II lymph node negative, estrogen receptor positive disease, but not for other breast cancer patients.

Routine IHC analysis for ER, PR and HER2 in breast tumors is also being offered by another commercial enterprise, Agendia (http://www.agendia.com/), which provides single gene expression profiles such as TargetPrint® and multi-gene expression assays such as MammaPrint.® Agendia has also developed BluePrint®, an 80 gene expression profile, for the classification of breast cancer into basal, luminal and ERBB2 (HER2/neu positive) molecular subclasses.

Advances in molecular technology now allow for profiling the expression of thousands of genes simultaneously in individual tumors. This in turn is helping researchers and clinicians better understand how individual tumors form and propagate. Each step in the process is characterized by changes in distinct molecular signaling pathways and gene expression patterns. Work by Vogelstein and colleagues revealed approximately 80 separate mutations that comprise the genomic landscape in a typical breast tumor, with a similar number of mutations in a typical colon tumor. Although the number of mutant genes was similar in breast and colon tumors, the particular genes involved and the types of mutations were quite different. For example, mutations converting 5'-CG to 5'-TG were much more frequent in colorectal cancers than in breast cancers. Of particular importance was the finding that only about 15 separate genes drove the initiation, progression, or maintenance of these tumors, as the remaining 60-plus mutated genes were seen in <5% of all cancers that were studied. The next wave of testing will involve full genomic sequencing, further stratifying breast tumors into possibly hundreds of distinct tumor types with different prognoses and varied responses to specific treatments.

5.19.2 Several Gene Polymorphisms Affect Activity in Signal Transduction Pathways

The use of the drug Gleevac® (imatinib) has transformed the treatment of chronic myelogenous leukemia (CML). As discussed in Sect. 1.6.2, some individuals with CML have a translocation involving chromosomes 9 and 22, often called the Philadelphia chromosome, which creates the fusion gene *BCR/ABL*. The fusion protein bcr-abl is a tyrosine kinase; it speeds up the rate of cell division and impairs the process of DNA repair, causing excessive proliferation of neutrophils, eosinophils and basophils. Imatinib works by binding to the ATP-binding site of bcr-abl and selectively inhibiting its activity.

In addition to bcr-abl, imatinib also inhibits other tyrosine kinase receptors such as c-kit and platelet derived growth factor receptor-A (PDGFR-A). Gastrointestinal stromal tumors (GISTs) have mutations of cell surface proteins that activate cell proliferation, including mutations in the *c-kit* and *PDGFRA* genes. In December 2008, the FDA approved the use of imatinib as an adjuvant therapy for resected adult GISTs.

The epidermal growth factor receptor (EGFR) has been validated as a therapeutic target in several human tumors, including colorectal cancer. EGFR activates the RAS/RAF/MAPK, STAT, and PI3K/AKT signaling pathways, which together modulate cellular proliferation, adhesion, migration and survival, as well as angiogenesis. The anti-EGFR targeted monoclonal antibodies, cetuximab (Erbitux®) and panitumumab (Vectibix®) are used as standard treatment for colon cancer. However, approximately 30–40% of metastatic colon cancers harbor a somatic mutation of the *KRAS* oncogene, which indicates poor prognosis and is associated with lack of response to EGFR inhibitors. The KRAS protein mediates a portion of the pathway that lies downstream from the EGFR receptor. Therefore, blockade of the EGFR receptor cannot prevent a gain-of-function mutation in *KRAS* from stimulating the growth of the tumor. Current practice guidelines recommend that only individuals with the wild-type sequence of the *KRAS* gene should be treated with EGFR inhibitors along with chemotherapy.

The *KRAS* gene is also considered in the selection of treatment for certain forms of non-small cell lung cancer (NSCLC). Somatic *KRAS* mutations occur in approximately 15–30% of NSCLC variants, with mutation most often in codons 12 and 13. Lung cancers that harbor *KRAS* mutations are more aggressive and resistant to treatment with tyrosine kinase inhibitors. However, by targeting a second, more easily inhibited "codependent" gene, known as *TBK1,* one can effectively destroy *KRAS*-dependent lung cancer cells. TBK1 is also a protein kinase. Mutations in *TKB1* have not been reported to cause cancer, but in *KRAS*-driven tumors, both *TKB1* and *KRAS* must be active for the cancer cells to survive. The key role of *TBK1* was discovered by using new molecular techniques that employ RNA interference (RNAi, see Sect. 1.6.5) methods to turn off thousands of different genes in various cancer and non-cancer cell lines. By finding genes that were shut off by these short pieces of RNA strands, *TBK1* was identified as selectively causing *KRAS*-dependent cancer cells to self-destruct. This new approach to targeted cancer treatment, known as "synthetic lethality," refers to a partnership in which two genes in a cell have a combined effect that neither has by itself.

Another example of synthetic lethality that is used against *BRCA*-related breast, ovarian and prostate cancers involves the poly (ADPribose) polymerase inhibitors (*PARP1*). Cells that lack BRCA are unable to use homologous recombination to repair double stranded breaks and interstrand cross-links. However, the other major DNA repair mechanism, base-excision repair, compensates for that loss; PARP-1 activity is required for this process. By directly targeting PARP-1, olaparib disables the cell's remaining base-excision repair mechanism, leading to selective cell death. PARP1 inhibitors may be even more promising for the treatment of early-stage tumors and sporadic cancers with similar defects, and possibly for prevention.

Rituximab is often prescribed for patients with non-Hodgkin's lymphoma. Rituximab binds to the CD20 protein on B cells, and stimulates natural killer cell activity by engaging their Fc gamma 3A receptors (FCGR3As). The *FCGR3A* gene has a functional phe158val polymorphism in its coding sequence; the valine

isoform of the protein has a higher affinity for rituximab, and produces more efficient cell lysis in *in vitro* preparations. Two studies have reported that patients with follicular lymphoma who were homozygous for the valine isoform of this *FCGR3A* polymorphism achieved significantly higher response rates than patients who possessed one or more copies of the phenylalanine allele of *FCGR3A*.

5.19.3 Glioblastomas with MSH6 Mutations Resist Alkylating Agents

Alkylating agents prevent malignant cells from replicating and dividing further by cross-linking the DNA strands, causing breaks and abnormal base pairing. The cell's innate DNA repair mechanisms attempt unsuccessfully to repair the damage, eventually causing the cell to go into apoptosis. The MSH6 protein plays an integral role in DNA damage repair and induction of apoptosis, and some patients with glioblastomas have somatic mutations in the tumor cells that inactivate the *MSH6* gene. Patients with MSH6-deficient glioblastomas resist the effects of alkylating agents. One study has reported that patients with MSH6-negative glioblastomas are more likely to experience further growth of the tumors during temozolomide treatment than patients with MSH6-positive glioblastomas.

5.19.4 Pharmacokinetically Relevant Gene Polymorphisms Influence Drug Response, Especially the Risk for ADRs

As is true for so many drugs, the patient's status for *CYP450* genes (see Sect. 4.6) alters the metabolism of several anticancer agents. CYP2D6 metabolizes estrogen receptor antagonists such as tamoxifen, alkylating agents such as cyclophosphamide, and vinca alkaloids such as vincristine. For many of the drugs that are metabolized by the CYP enzymes, patients with the ultrarapid metabolizer phenotype may not respond to standard doses of these drugs, while patients with the poor metabolizer CYP450 phenotype have a greater risk for adverse drug reactions (ADRs). Tamoxifen, however, is a prodrug; it is metabolized into the active compound endoxifen. Unlike drugs that are given as active compounds, ultrarapid metabolizers achieve higher levels of tamoxifen's active compound than others do, while poor metabolizers often fail to respond to standard doses. You should consider prescribing an alternative therapy, such as aromatase inhibitors, for patients with the poor metabolizer phenotype.

The genes that encode drug transporter proteins also significantly affect the pharmacokinetics of some anticancer drugs. Irinotecan is a substrate for the organic ion transporter 1B1 (OATP1B1, gene name *SLCO1B1*), and the patient's status for

polymorphisms in *SLCO1B1* has also been reported to influence risk for irinotecan ADRs. Similarly, the OCT1 drug transporter transports imatinib into cells; patients with higher baseline levels of OCT1 expression achieve higher intracellular concentrations of imatinib, and have better rates of progression-free survival as well as overall survival.

One of the earliest pharmacogenetic tests to find widespread use in the clinic involved measuring the activity of the enzyme thiopurine methyltransferase (TPMT) in red blood cells. TPMT activity is generally considered to be trimodally distributed, but one recent study has suggested the existence of a fourth group of ultrarapid metabolizers with unusually high TPMT activity. TPMT methylates thiopurine immunosuppressants such as mercaptopurine and azathiopurine, and patients with low TPMT activity must be given low doses of these thiopurines, to avoid life-threatening myelosuppression. It is less clear whether TPMT status affects the therapeutic efficacy of these drugs. The results of studies addressing the therapeutic response as a function of TPMT status are mixed.

Several other enzymes have functional polymorphisms that affect the pharmacokinetics of anti-cancer drugs. For example, dihydropyrimidine dehydrogenase (DPD) plays an important role in the metabolism of 5-fluorouracil. DPD deficiency increases the individual's risk for ADRs such as severe GI reactions, myelosuppression and neurotoxicity. Once again, however, the effect of DPD status on drug efficacy is less clear.

The FDA has approved testing the patient's status for polymorphisms in the gene that encodes the enzyme uridine diphosphate glucuronosyltransferase 1A1 (*UGT1A1*) in order to guide the choice of the dose of irinotecan to be administered to patients with colorectal or lung cancer. *UGT1A1* has a repeated sequence length polymorphism in its promoter region. Some individuals have the sequence A(TA)6TAA (i.e. 6 repetitions of the AT dinucleotide repeat), while others have an extra AT in the repeated string, giving them the sequence A(TA)7TAA. The 7-repeat allele of the *UGT1A1* promoter polymorphism reduces the level of activity in the gene. Patients with low UGT1A1 activity are at increased risk for ADRs such as diarrhea and neutropenia after irinotecan administration.

Busulfan (BU) is an alkylating agent that is commonly used to ablate existing bone marrow cells to prepare the patient for hematopoietic cell transplant. BU is metabolized predominantly by several glutathione S-transferases, including GSTA1, GSTM1 and GSTP1. The *GSTA1* gene has a functional polymorphism in it that affects the enzyme's activity. The lower-activity allele has been associated with higher blood levels of BU and greater risk for graft versus host disease. In addition, some patients have no GSTM1 activity, due to a homozygous deletion of the *GSTM1* gene. As would be expected, these patients achieve significantly higher blood concentrations of GSTM1's target drugs than patients who do not have the deletion. One study has also reported that they have a higher risk for hepatic veno-occlusive disease. Other studies have failed to corroborate this finding, however. There are two functional polymorphisms in the *GSTP1* gene that produce isoforms with altered activity, although neither has been associated with an altered response to BU.

5.19.5 HLA Type May Influence the Risk for Hypersensitivity Reactions After Abacavir

Although it is often difficult to specify the mechanism whereby HLA type affects risk for disease or response to a drug, several HLA types appear to be associated with altered risk for several specific diseases/ADRs. For example, approximately 3–8% of patients who are prescribed abacavir experience a hypersensitivity reaction that may include fever, rash, gastrointestinal symptoms, respiratory symptoms, fatigue or myalgia. It has been reported that the HLA-B*5071 allele is strongly associated with susceptibility to abacavir hypersensitivity reactions. This has been reported to be a highly sensitive and specific test, with high predictive utility in both black and white patients. The FDA now recommends testing a patient for the presence/absence of the HLA-B*5071 allele before prescribing abacavir.

Further Readings

Bodmer W, Tomlinson I (2010) Rare genetic variants and the risk of cancer. Curr Opin Genet Dev Jun;20(3):262–7

Boss DS, Beijnen JH, Schellens JH (2010) Inducing synthetic lethality using PARP inhibitors. Curr Clin Pharmacol Aug;5(3):192–5

Easton DF, Eeles RA (2008) Genome-wide association studies in cancer. Hum Mol Genet Oct 15;17(R2):R109–15

Ioannidis JP, Castaldi P, Evangelou E (2010) A compendium of genome-wide associations for cancer: critical synopsis and reappraisal. J Natl Cancer Inst Jun 16;102(12):846–58

Jasperson KW, Tuohy TM, Neklason DW, Burt RW (2010) Hereditary and familial colon cancer. Gastroenterology Jun;138(6):2044–58

Lynch HT, de la Chapelle A (2003) Hereditary colorectal cancer. N Engl J Med Mar 6;348(10):919–32

Markowitz, S., Bertagnolli, M (2009) Molecular basis of colorectal cancer. N Engl J Med Dec 17;361(25):2449–60

Narod SA (2010) BRCA mutations in the management of breast cancer: the state of the art. Nat Rev Clin Oncol Oct 19 7(12):702–7

Walter MJ, Graubert TA, Dipersio JF, Mardis ER, Wilson RK, Ley TJ (2009) Next generation sequencing of cancer genomes: back to the future. Per Med Nov 1;6(6):653

Chapter 6
Personalizing Risk Assessments and Treatments for Complex Cardiovascular Disease

Abstract The term "cardiovascular disease" encompasses a number of different conditions that affect different aspects of the circulatory system. Family history is often very useful in predicting one's risk for cardiovascular disease, and this chapter includes a list of several Internet websites that will analyze family and personal data and estimate one's level of risk. The development of predictive genetic tests has been difficult for several reasons, including the likelihood that there is more than one molecular subtype present in any group of patients with the same clinical diagnosis. Most of the risk-increasing alleles that have been discovered have low penetrance; it may require hundreds of individual genetic tests to provide the critical genetic data to predict one's risk. The list of candidate genes is growing rapidly, however. The list includes genes whose proteins are integral to the cardiac musculature, as well as genes that influence aspects of metabolism that may predispose one to cardiovascular disease, including blood lipid and sugar levels, blood pressure, cardiac ion channel function, inflammation, oxidative stress and blood homocysteine levels. The chapter includes a discussion of these low-penetrance gene variants, as well as a discussion of two examples of cardiovascular diseases for which highly penetrant risk-increasing alleles have been reported: familial hypercholesterolemia and hypertrophic cardiomyopathy. The chapter also includes a discussion of the pharmacogenetic tests that can influence the choice of treatment for patients with cardiovascular disease, and a case report that illustrates both the benefits and limitations of genetic testing in a patient with type 2 diabetes and a family history of myocardial infarction.

6.1 Cardiovascular Diseases Are Complex, Multifactorial Diseases with Highly Variable Phenotypes

Cardiovascular disease (CVD) is the leading cause of death for both men and women in the United States. More than one in every four deaths is attributable to CVD. According to the American Heart Association, approximately 15 million people have some form of the condition, and in 2009, CVD cost more than

K.M. Sweet and R.C. Michaelis, *The Busy Physician's Guide To Genetics, Genomics and Personalized Medicine*, DOI 10.1007/978-94-007-1147-1_6,

$304.6 billion in health care services, medications and lost productivity. The cumulative risk for CVD is more significant in males than females (35% vs. 24% by age 70; 49% vs. 32% by age 90, respectively), with females typically developing the disease 10 years later than males. Although usually a disease of advanced age, approximately 15% of CVD cases are diagnosed before age 65.

The term "cardiovascular disease" encompasses a number of different conditions that affect different aspects of the circulatory system. The most common forms of CVD are:

- Coronary artery disease (CAD) is the most common form of CVD. It accounts for more than 50% of cases, and usually manifests as a narrowing or blockage of the coronary arteries due to atherosclerosis. CAD progresses over time, and is the major cause of heart attacks (myocardial infarctions, MIs).
- Cardiomyopathy is a progressive heart disease in which the heart is abnormally enlarged, thickened and/or stiffened. There are several forms of cardiomyopathy, including hypertrophic cardiomyopathy, dilated cardiomyopathy, arrhythmogenic right ventricular cardiomyopathy and restrictive cardiomyopathy.
- Heart rhythm disorders (arrhythmias) cause an irregular heartbeat and can lead to sudden cardiac arrest. Specific forms of arrhythmia include ventricular fibrillation, atrial fibrillation, cardiac channelopathies, tachycardias and bradyarrhythmias.
- Heart valve disease occurs when the heart valves do not function properly.
- Pericardial disease, including inflammation of the membrane sac surrounding the heart (pericarditis) or an accumulation of fluid surrounding the heart (pericardial effusion).
- Heart failure results from impaired cardiac function with subsequent inadequate systemic perfusion to meet the body's needs. Systolic heart failure results when there is reduced cardiac contractility. Diastolic heart failure stems from impaired cardiac relaxation and abnormal ventricular filling.
- Aortic aneurysm
- Congenital heart disease/defects

CVDs are complex, multifactorial diseases, and they all demonstrate considerable variability with respect to their pathophysiology, clinical presentation and genetics. One prime example is coronary artery disease (CAD), where the accumulation of atherosclerotic plaques in the walls of the coronary arteries begins early in life. Over time this may manifest as one of several different clinical phenotypes, including atherosclerosis, chronic stable angina, acute coronary syndrome, myocardial infarction, congestive heart failure or even sudden cardiac death.

Atypical levels of gene activity are among both the causes and effects of CAD. The transition from a stable coronary atherosclerotic plaque to an unstable inflamed plaque, platelet accumulation and activation, fibrin deposition, thrombus formation and the potential for vessel occlusion is mediated in part by the inheritance of risk-increasing alleles in one or more of several critical genes. In addition, the progression of CAD is accompanied by changes in of the level of activity in

multiple genes. Some conservatively estimate that hundreds of modifier genes influence the CVD phenotype, each with an incremental effect on risk, and each with the potential to interact with other traits that are themselves polygenic (e.g. diabetes, blood pressure).

Numerous dietary, environmental and lifestyle risk factors and their interactions also impact plaque stability and inflammation, platelet function and the coagulation cascade. Some well known risk factors include family history, obesity, lack of exercise, diet, smoking, metabolic syndrome and injury to the heart. People who are affected with other health conditions, such as diabetes, hypertension, high cholesterol and/or stroke are at increased risk.

6.2 Family History and CVD Risk

Twin and family studies have long established that CVD aggregates in families. In fact, family history contributes to risk independently of other established risk factors such as elevated LDL cholesterol, decreased HDL cholesterol and diabetes. Three-fourths of all early-onset CVD (prior to age 55 years in men, prior to age 65 years in women) patients have a family history of CVD, as do almost 50% of CVD cases of all ages. A history of early CVD in a first-degree relative (parent, sibling, or child) approximately doubles an individual's risk. Having two or more first-degree relatives with CVD is associated with a three- to sixfold increased risk. Sibling history of CVD appears to be a greater risk factor than parental history. Gender differences also exist, as the presence of CVD in multiple females within a family is more often associated with greater genetic burden compared to having multiple affected males, and elevates risk for all family members.

Systematic collection and assessment of family history information is the most appropriate approach to identify individuals at risk for CVD. Demographic and medical information should be obtained from all first- and second-degree relatives, including current age or age at death, history of CVD and related conditions such as stroke, peripheral vascular disease, thrombosis, arrhythmia, hypertension, aortic aneurysm, diabetes and lipid abnormalities. Environmental risk factors such as smoking history and dietary factors should also be assessed. Because lay people's reports of "heart disease" or "heart problems" in their family members often lack specific details, it is often necessary to document these reports by medical records or death certificates (especially on more distant relatives, as accuracy of recall of diagnosis may be limited), or additional questioning regarding specific types of CVD experienced or procedures undergone. Updating the family history on an annual basis allows for the most accurate assessment of risk.

Analysis of the family history may also reveal characteristic Mendelian patterns of inheritance suggestive of a single-gene disorder. Many of the more common Mendelian

CVD disorders, such as familial hypercholesterolemia and hypertrophic cardiomyopathy, exhibit an autosomal dominant pattern of inheritance, as discussed later in this chapter. Given the potentially serious consequences that people who possess risk-increasing alleles may face, it is important to analyze the family history, to identify family members who may be carrying potentially dangerous gene variants.

6.3 Useful Online Programs to Estimate Heart Disease Risk

The Framingham Risk Score, often used to predict heart disease risk, was developed from the Framingham Heart Study (http://www.framinghamheartstudy.org), a joint project of the National Heart, Lung and Blood Institute and Boston University. The study began in 1948 with 5,209 adult subjects from Framingham, Massachusetts and is now on its third generation of participants. The objective of this prospective study was to identify the common factors or characteristics that contribute to the development of heart disease. Over the years, the Framingham Heart Study has led to the identification of many of the significant heart disease risk factors such as high blood pressure, elevated serum LDL cholesterol and triglycerides, age, gender, smoking, obesity, diabetes and physical inactivity. Risk modeling takes into account many of these established heart disease risk factors. However, it does not include family history, and therefore will likely underestimate heart disease risk for those with a significant family history.

The online CardioSmart Risk Assessment Tool (http://www.cardiosmart.org) is also based on findings from the Framingham Heart Study. It is a patient education site that predicts risk for developing a heart attack or dying from coronary disease within the next 10 years. It is best used for people aged 20 years or older without known heart disease or diabetes. CardioSmart was designed by the American College of Cardiology.

The Reynolds Risk Score for women (http://www.reynoldsriskscore.org) was developed from assessment of 35 risk factors in 24,558 initially healthy US women who were followed for 10.2 years for incident cardiovascular events through the Women's Health Study sponsored by the National Heart, Lung, and Blood Institute (NHLBI). The Reynolds Risk Score for men was similarly developed using data from 10,724 initially healthy non-diabetic American men who were followed over a 10 year period for the development of heart attack, stroke, angioplasty, bypass surgery, or death related to heart disease. This tool is more comprehensive than CardioSmart, in that it uses serum C-Reactive Protein level and family history of heart disease prior to age 60 in the parents of the user.

Your Disease Risk (http://www.yourdiseaserisk.wustl.edu) assesses heart disease risk using medical and smoking history, diet, physical activity measures and family history of affected first-degree relatives.

Lastly, Family Health*Link* (https://familyhealthlink.osumc.edu) is an online risk triage tool that assesses risk for heart disease and for cancer based on family

history, and which was developed at Ohio State. This tool stratifies risk into average, moderate and high risk categories. Individuals estimated to have a high risk for heart disease or cancer should be offered genetic consultation.

6.4 CVD-Associated Gene Variants Have Been Particularly Difficult to Identify

The proportion of the risk that is attributable to genetic versus nongenetic factors can vary from one complex disease to another. In the case of CVD, the nongenetic and genetic components are thought to contribute approximately equally to the individual's risk. This, coupled with the significant genetic and phenotypic heterogeneity, has made it difficult to identify the gene variants that contribute to CVD.

Early linkage studies that included all patients within the disorder in question failed to identify causative genes. Later studies, which used small families with more extreme phenotypes (very early onset CAD or MI), identified associated genetic loci on a number of chromosomes, including 1, 2, 3, 12, 13, 14, 16 and X. Unfortunately, most of these findings have not been corroborated by subsequent studies, further testifying to the genetic heterogeneity in CVDs. Subsequent larger genome-wide linkage analyses, which included patients with relatively early-onset CVDs (≤59 years), identified only a single associated locus, at 2p.12–2q23.3. Unfortunately, this locus encompasses a stretch of 80 Mb (80 million bp) in chromosome 2, and contains more than 440 putative genes. The sheer number of plausible candidates makes it very difficult to identify candidate genes from this region without further studies that narrow down the region.

After linkage analyses failed to identify candidate genes for CVD, researchers turned to the newly developed whole-genome microarray technology, and used genome-wide association (GWA) studies to identify candidate genes. By screening large populations, these GWA studies can identify gene variants that contribute to the individual's risk, even if the variant only exerts a small influence on the individual's risk for the disorder (i.e. has low penetrance). The increased power afforded by these studies has enabled GWA studies to identify a number of low-penetrance variants that influence either the individual's risk for CVD or his/her risk for predisposing phenotypes such as hyperlipidemia.

Because these risk-increasing alleles all have low penetrance, there is still considerable work to be done before we can translate most of these basic research discoveries into clinically useful genetic tests. When the risk-increasing alleles have low penetrance, information about dozens, perhaps even hundreds, of these markers will have to be combined to produce a clinically useful genetic test. Current research is focused on identifying more low-penetrance variants that increase one's risk for CVDs, and determining the most effective ways to combine the results from multiple tests into a predictive algorithm.

6.5 Most CVD-Associated Gene Variants Have Low Penetrance

6.5.1 A Cluster of Linked Markers in 9p21.3 Are Associated with Several CVDs

The Wellcome Trust Cardiovascular Research Initiative (WTCRI) spearheaded efforts using GWA array analyses in coronary artery disease, and in 2007, revealed strong associations with the 9p21 locus (Table 6.1). This locus has since become the focus for the study of genetic risk factors for several CVDs, including stroke, abdominal aortic aneurysm and intracranial aneurysm. Thus many diseases previously thought to have distinct etiologies may in fact share a common molecular cause.

The marker that was initially most strongly associated with CVD was SNP rs1333049. The rs1333049 SNP does not lie within any known gene, but lies in close proximity to two cyclin dependent kinase inhibitor genes, *CDKN2A* and *CDKN2B*. In addition, the rs1333049 SNP lies close to the sequence that produces the *CDKN2B* gene's antisense transcript (*CDKN2BAS*). Antisense transcripts

Table 6.1 Low-penetrance gene variants that influence the risk for CVD

dbSNP No.	Locus/gene	Chrom band	Gene product
rs10757278[a] rs1333049[a]	CDKN2A/CDKN2B CDKN2BAS	9p21.3	Cyclin-dependent kinase inhibitors and antisense transcript
rs599839	CELSR2/PSRC1	1p13.3	cadherin, EGF LAG seven-pass G-type receptor 2
rs1746048 rs501120	CXCL12	10q11.21	Chemokine (C-X-C motif) ligand 12
rs2291834 rs17465637	MIA3	1q42.12	melanoma inhibitory activity family, member 3
rs9818870	MRAS	3q22.3	muscle RAS oncogene homolog
rs7739181	PHACTR1	6p24.1	phosphatase and actin regulator 1
rs3184504	SH2B3	12q24.13	SH2B adaptor protein 3
rs6725887	WDR12	2q33.2	WD repeat domain 12
rs6922269	*MTHFD1L*	6q25.1	methylenetetrahydrofolate dehydrogenase (NADP+dependent) 1-like
rs17228212	SMAD3	15q22.33	SMAD family member 3

[a]These two markers' association with CVD has been replicated in multiple independent studies. Other markers from this locus, including the SNPs rs10757274, rs2383206, rs10116277, rs1333040 and rs2383207, have also been associated with CVDs in large, well-controlled studies

(see Sect. 1.6.5) are thought to regulate the activity of many genes, and genetic variability in these sequences could hypothetically contribute to the variability in protein activity that underlies interindividual differences.

At this point in time, it is not clear which of the sequences that are linked to rs1333049 is the critical functional polymorphism that is responsible for rs1333049's association with CVD. The *CDKN2B* gene is induced by transforming growth factor beta (TGF-β), and this induction has previously been implicated in the development and progression of atherosclerosis. Mutation analyses have been performed on both the *CDKN2A* and *CDKN2B* genes in a significant number of patients with CVD, however, and no mutations have been found.

Whatever specific sequence is responsible, the individual's status for the rs1333049 SNP clearly influences his/her risk for CAD. The WTCRI study reported that heterozygotes who possess one copy of the rs1333049 risk-increasing allele (the C allele), have a relative risk of 1.3 for CAD, while homozygotes have a relative risk of 1.7. In addition, a number of studies have confirmed this finding. A subsequent meta-analysis, incorporating seven case-control studies involving a total of 4,645 cases of MI or coronary artery disease and 5,177 controls, reported that the C variant of rs1333049 was more prevalent in cases than controls in all seven studies, as well as in the combined analysis.

Other SNPs that lie within the sequence that encodes the *CDKN2BAS* transcript have also been associated with CVDs. For example, the Ottawa Heart Study reported that two additional SNPs that lie within the *CDKN2BAS* locus (rs10757274 and rs2383206) were also associated with severe, premature CAD. This finding has been replicated in another subgroup of the Ottawa Heart Study, the Copenhagen City Heart Study and the Dallas Heart Study. Subsequently a number of GWA analyses were performed in Icelandic subjects, by primary investigators who then went on to form one of the original direct-to-consumer genetic testing companies, deCode. Different microarray platforms were used, as was a different study population – individuals with MI before the age of 70 years in males and 75 years in women. In these studies, three new SNPs that lie within the same *CDKN2BAS* locus (rs10116277, rs1333040 and rs2383207) were found to be associated with MI. A fourth locus in this region, rs10757278 was also identified and shown to have the strongest association with MI. It also had the strongest association with coronary vascular disease when individuals with MI were excluded. Each of these SNPs' associations with MI was then replicated in additional case-control studies. Independent association between rs10757278 and abdominal aortic and intracranial aneurysms was also observed, suggesting that the biological effect of this SNP influences multiple vascular disease phenotypes.

At this point in time, there can be little doubt that SNPs on the chromosome 9p21.3 region are associated with a variety of diseases that affect the cardiovascular system. The association of several SNPs in this locus with CVD has been replicated consistently, in several different ethnic groups, including Caucasian and Asian subjects. A haplotype that includes specific alleles of two of these SNPs (rs10757274 and rs2383206) is associated with a 15–20% increase in CVD risk for heterozygotes, and a 30–40% increase in risk in homozygotes.

Several commercial laboratories offer whole-genome SNP analyses using some of these distinct but tightly linked SNPs in a 9p21 marker panel. Some marker panels also include some of the additional CVD-associated SNPs that are listed in Table 6.1 below, and more will surely follow. Most of the variants that have been associated with CVD exert only small effects on the individual's risk. Because of this, no single marker, or even small panel of markers, can provide a clinically useful prediction of the individual's risk for CVD. In order to have appreciable clinical utility, tests that are based on these markers will have to include a combination of hundreds, or even thousands, of SNPs that together influence several distinct but interacting aspects of the disorder or drug response.

Two recent studies have testified to the limited clinical utility of the 9p21 genotype. The Women's Genome Health Study demonstrated that the 9p21 genotype did not improve the ability to predict CVD risk in a healthy middle-aged Caucasian female cohort over conventional predictors such as high sensitivity C-reactive protein and family history of premature MI. In addition, the 9p21 genotype did not improve the ability to predict CVD risk in a cohort of healthy middle-aged Caucasian men when used in combination with the Framingham risk score algorithm. The current consensus is that, while a 20% increased risk for 9p21 heterozygotes may not change recommendations for prevention or treatment in individuals from low-risk populations, it may make more of a difference in the prevention or treatment plans for individuals from moderate-risk populations.

6.5.2 Other Low-Penetrance Variants That Influence Risk for CVD

Additional studies using patients with CAD or MI have identified a few SNPs in other chromosomal loci, some located within or near plausible candidate genes for CVD. For example, rs6922269 is located on chromosome 6q25, within the gene encoding methylenetetrahydrofolate dehydrogenase 1-like protein (*MTHFD1L*). In addition, four other SNPs on chromosome 1 (rs599839, rs17465637, rs501120 and rs17228212) were identified. Each of these SNPs lies within a gene whose protein is thought to influence atherosclerotic plaque stability: proline/serine-rich coiled-coil 1 (*PSRC1*), melanoma inhibitory activity family, member 3 (*MIA3)*, chemokine (C-X-C) ligand 12 (stromal cell derived factor 1) (*CXCL12*), and SMAD family member 3 (*SMAD3*), respectively. Given the low penetrance of these variants, the associated genetic tests will have limited predictive power, and these markers must be combined with other markers from genes whose proteins influence other aspects of cardiovascular function in order to develop test panels with good clinical utility.

At this point in time, there is much to be done before an algorithm that predicts the individual's risk for CVDs accurately can be developed. Fortunately, the work is proceeding at a rapid pace. Additional microarray and full genome sequencing studies using more carefully selected cohorts of patients with better defined phenotypes are identifying additional genetic risk factors for specific CVDs. There is a need to

extend these association analyses to include non-Caucasian populations, and to develop models that incorporate the most predictive genetic and nongenetic factors. Finally, as we learn more about the pathophysiology of the hereditary single gene CVDs, these studies will provide insights into the pathways that are affected in these disorders, and reveal other potentially fruitful avenues of investigation.

6.6 Low-Penetrance Gene Variants That Affect Predisposing Phenotypes

6.6.1 Low-Penetrance Gene Variants That Alter Blood Lipid Levels

GWA studies have also identified important SNP associations for predisposing CVD risk phenotypes such as Type II diabetes mellitus (T2D, see Sect. 8.2) and dyslipidemia. SNPs in the 3-hydroxy-3-methylglutaryl coenzyme A reductase (*HMGCR*) gene have been reported to influence serum low-density lipoprotein (LDL) cholesterol levels. In addition, meta-analysis has identified 11 SNPs associated with increased LDL-cholesterol concentrations, including SNPs in the HMGCR gene and additional genes known to be involved in lipid metabolism. These same SNPs were then re-examined and shown to be more prevalent in individuals with CAD than in controls in the Wellcome Trust Case Control Consortium cohort. More recently, two large meta-analyses of individuals of European ancestry identified more than 20 loci associated with dyslipidemia. Many of these loci were in, or near, genes known to be involved in lipid metabolism or transport, or in genes that are known to cause single-gene disorders that include dyslipidemia among their features.

6.6.2 Sodium/Potassium Regulation and Essential Hypertension

Essential (aka primary) hypertension affects one in four American adults, and family and twin studies have suggested that there is a strong genetic component to the etiology of essential hypertension. In the search for genetic factors, much attention has been focused on the renin-angiotensin-aldosterone system and regulation of sodium (Na^+) and potassium (K^+) levels. For example, the angiotensin converting enzyme (ACE) converts angiotensin I to angiotensin II, which is a vasoconstrictor and stimulator of aldosterone secretion. The *ACE* gene has a 287 bp insertion/deletion (I/D) polymorphism in intron 16; the I allele produces a low-activity isoform of the protein. Individuals with the DD genotype have an increased risk for left ventricular hypertrophy, MI, increased plaque instability, stent restenosis, ischemic

or idiopathic cardiomyopathy, and CAD. They also benefit less from treatment with ACE inhibitors than patients with the ID or II genotypes do.

Several genes whose proteins influence transmembrane ion currents have been identified as risk factors for essential hypertension. For example, adducin is a heterodimeric protein that promotes the binding of Na^+/K^+ ATPase to the cell membrane's cytoskeleton, and promotes assembly of the spectrin-actin network that anchors many ion channel proteins in the cell membrane. The individual's genetic status for the G460W polymorphism in the gene that encodes the alpha subunit of the adducin protein (*ADD1*) has been reported to influence the activity of the Na^+/K^+ ATPase, and thereby influences sodium retention and blood pressure. The W allele has been associated with essential hypertension in European and Japanese subjects, but one study that focused on Scandinavian subjects reported a reduced frequency of the W allele in hypertensive individuals. It is not known whether the G460W polymorphism has any functional consequences. It is possible that the associations observed are actually due to the G460W marker being tightly linked to a functional polymorphism, and that the risk-increasing allele of the true functional polymorphism is associated with different alleles of the G460W marker in the different populations.

One of the genes whose proteins influence K^+ currents is the *KCNJ1* gene, which encodes the inward rectifying K^+ channel ROMK1. While it has been known for some time that loss-of-function mutations in the *KCNJ1* gene can cause a single-gene form of essential hypertension known as Bartter syndrome, it has recently been discovered that less deleterious, and therefore less penetrant, variants in the *KCNJ1* gene are associated with essential hypertension. In addition, the lysine deficient kinases type 1 and type 2 (*WNK1* and *WNK2*, respectively) are serine-threonine proteases that are expressed in the distal nephron. WNK1 activates the kinase SGK1, thereby activating the epithelial sodium channel. *WNK1* and *WNK2* variants have been reported to influence the individual's risk for essential hypertension as well.

6.6.3 Cardiac Channelopathies

It goes without saying that all the genes that encode subunits of cardiac muscle ion channels are potential risk factors for cardiac channelopathies. The list of companies that offer testing for cardiac channelopathies is growing, and the test panels are expanding to include not only the cardiac channel subunit genes, but genes whose proteins secure the ion channels within cell membranes as well.

For example, several companies offer tests designed to help estimate the individual's risk for several forms of Long QT Syndrome (LQTS). LQTS is inherited in an autosomal dominant manner, and mutations in the LQTS-related genes have incomplete penetrance; at least one-third of individuals who are heterozygous for mutations in these genes are asymptomatic. As Table 6.2 illustrates, most of the genes that are included in these LQTS risk tests encode subunits of cardiac ion channels, but a few encode proteins that help anchor the ion channels in the membrane, such as the

Table 6.2 Genes screened to determine susceptibility to LQTS

Gene	Protein encoded
KCNQ1	Core subunit, type 1 KQT-like voltage-gated potassium channel
KCNE1	Auxiliary subunit, type 1 KQT-like voltage-gated potassium channel
KCNH2	Core subunit, subfamily H, member 2 voltage-gated potassium channel
KCNE2	Auxiliary subunit, subfamily H, member 2 voltage-gated potassium channel
SCN5A	Alpha subunit, type V voltage-gated sodium channel
CACNA1C exons 8 and 9	Alpha 1 C subunit, L-type voltage-dependent calcium channel
KCNJ2	Inwardly-rectifying potassium channel, subfamily J, member 2
SCN4B	Beta subunit,type IV voltage-gated sodium channel
CAV3	Caveolin 3
SNTA1	Syntrophin alpha 1, aka dystrophin-associated protein A1
AKAP9 exon 18	A-kinase anchor protein 9

SNTA1 gene that encodes syntrophin alpha 1 (aka dystrophin-associated protein A1) and the gene that encodes the A-kinase anchor protein 9 (*AKAP9*). Mutations in the genes listed in Table 6.2 account for over 75% of the familial cases of LQTS.

Several of these genes can result in different specific channelopathy phenotypes, depending on the specific mutation in the gene and the influence of other genetic, and perhaps nongenetic, risk factors. For example, mutations in the *SCN5A* and *CACNA1C* genes can produce LQTS in some patients, and Brugada syndrome (BrS) in others. Similarly, *KCNJ2* mutations may produce LQTS in some patients, and Andersen-Tawil syndrome in others.

BrS (aka Sudden Unexpected Nocturnal Death Syndrome) is inherited in an autosomal dominant pattern, and is more common in males than females. In addition, it has a higher prevalence in Asian populations than others. Only 20% of cases of BrS are attributable to *SCN5A* mutations. Other genes that have been reported to be mutated in patients with BrS are the *GPD1L* gene, which encodes the glycerol-3-phosphate dehydrogenase-like peptide, the *CACNB2* gene, which encodes the beta-2 subunit of the L-type voltage dependent calcium channel, the *KCNE3* gene, which encodes the potassium voltage-gated channel subfamily D member 3 protein, and *SCN1B*, which encodes the beta-1 subunit of the SCN5A sodium channel.

Another cardiac channelopathy for which genetic testing is available is catecholaminergic polymorphic ventricular tachycardia (CPVT). Approximately 50–55% of individuals affected with CPVT possess a mutation in the *RYR2* gene, which encodes the pore-forming subunit of the cardiac calcium release channel. CPVT is inherited in an autosomal dominant manner, but the mutations responsible have limited penetrance. Because patients with CPVT often have normal resting heart rhythms, the condition may not be detected until the individual suffers an exercise-related cardiac event. Most of the CPVT-related cardiac events are seen in children, adolescents and young adults. If left untreated, 30% of CPVT patients will develop symptoms by age 10, and 80% will develop symptoms by age 40. CPVT has an overall mortality rate of 30–50%.

6.6.4 Variants That Influence the Inflammation Response Influence Risk for CVDs

Excessive inflammation is being implicated in the etiology of a growing list of diseases, and several genes whose proteins participate in the inflammation response constitute risk factors for CVDs. This includes some of the genes listed in Table 6.1. For example, the CXCL12 protein is a cytokine that activates leukocytes, often in response to the release of proinflammatory molecules such as interleukin-1. In addition, the SH2B adaptor protein 3 helps mediate the signal transduction pathway that is involved in T cell activation.

The list of inflammation-related proteins that influence risk for CVDs also includes cytokines such as interleukin-6 (IL-6) and interleukin-10 (IL-10), as well as receptors such as the type 4 toll-like receptor protein (TLR4) and the chemokine (C-C motif) receptor 5 (CCR5). In addition, mutations in the cyclooxygenase type 2 (*COX-2*) and 5-lipoxygenase (5-*LOX*) genes, which encode enzymes that are involved in synthesizing pro-inflammatory molecules such as prostaglandins and leukotrienes, have also been reported in patients with CVDs. Screening for variants in the *IL-6*, *IL-10*, *TLR4*, *CCR5*, *COX-2* and 5-*LOX* genes will probably be used in the near future to help estimate the individual's risk for CVDs, especially CAD and MI.

6.6.5 Genetic Variants Influence the Level of Oxidative Stress

A high level of oxidative stress is another factor that is well known to predispose the individual to CVD. Several genetic variants have been discovered that influence the individual's level of oxidative stress. Some influence the generation of superoxide radicals, while others influence their catabolism.

NAD(P)H oxidase is a membrane-bound enzyme complex that generates considerable amounts of superoxide radicals. The *CYBA* gene encodes the p22phox subunit of NAD(P)H oxidase. Possessing the high-activity allele of a functional polymorphism in the *CYBA* gene increases the individual's level of superoxide radicals and risk for hypertension. In addition, possessing the high-activity allele of this polymorphism is also associated with vascular stiffness and thickening of the wall of the carotid artery.

The individual's level of oxidative stress is also influenced by his/her status for genes whose proteins detoxify these superoxide radicals. The mu class of glutathione S-transferase enzymes actively detoxify superoxide radicals. One large, well-designed study has reported an association between one allele of a SNP in the *GSTM* gene, which encodes the mu type glutathione S-transferase, and hypertension. Another study, which used considerably different methods, failed to corroborate this finding, however. At present, none of these findings have been translated into clinically useful tests.

6.6.6 *Plasma Homocysteine Levels Influence Risk for CVD*

One of the metabolic factors that influence the individual's risk for CVD is the level of homocysteine in the blood. Elevated blood homocysteine levels are associated with increased risk for arteriosclerotic cerebrovascular disease and stroke. The enzyme cystathionine beta synthase (CBS) converts homocysteine to cystathionine; low CBS activity will result in elevated homocysteine levels and increased risk for cardiovascular disease. The patient's status for the G919A (G307S) polymorphism in the *CBS* gene affects not only his/her plasma homocysteine level, but also his/her response to pyridoxine therapy. Patients with the A allele of this polymorphism do not respond as well to pyridoxine as patients with the G allele do. In addition, the MTHFD1L protein listed in Table 6.1 helps support methylation pathways by regenerating methionine from homocysteine, thereby influencing plasma homocysteine levels. This *CBS* gene polymorphism will no doubt be included in multigenic marker panels that are developed to predict the individual's risk for specific CVDs.

6.6.7 *Knowing They Possess Low-Penetrance Variants May Motivate Patients' Behavior*

The increase in disease risk that comes from possessing the variant alleles of these CVD-associated SNPs is relatively small, and alone not sufficient to modify management or screening strategies. This information can still have a positive impact on the patient's health, however. Knowing that they possess some of these genetic risk factors, but also knowing that dietary, environmental and lifestyle factors have an equally important influence over CVD risk, may motivate some patients to assume more responsibility for health-promoting behaviors. Understanding that the increase in risk that is due to these genetic factors may be offset by reducing the risk from nongenetic factors may lead some patients to take more control over the controllable nongenetic factors, by doing things like quitting smoking, improving their diet or getting more exercise.

6.7 Case Report – Genetic Testing in a Patient with Type 2 Diabetes (T2D) and a Family History of Myocardial Infarction (MI)

Jameer is a 47-year-old African-American male who presents at your internal medicine practice for periodic health assessment. His medical history is significant for a diagnosis of T2D at age 36. He is a former smoker, but he quit 10 years ago. He is also a social drinker, but has no history of illegal drug use. Although Jameer is

physically fit, exercises aerobically four or five times a week, and adheres to a low-fat diet, he has recently developed borderline hypertension. He is worried and anxious about his risk for heart disease and other medical conditions.

6.7.1 Jameer's Initial Visit

Jameer's physical exam was remarkable for a BP of 135/88 and a BMI of 23.8 kg/m^2. His cardiac exam revealed nothing remarkable, and his levels of serum cholesterol, HDL, LDL, triglycerides and glucose were in the normal range. Review of his family history (Fig. 6.1) reveals that his father had an MI at age 49, with coronary revascularization surgery at age 52, and died at age 54. Jameer's paternal grandfather died of MI at age 58.

Jameer tells you today that he used the Family HealthLink online risk assessment tool, which classified him as having a high risk for coronary heart disease.

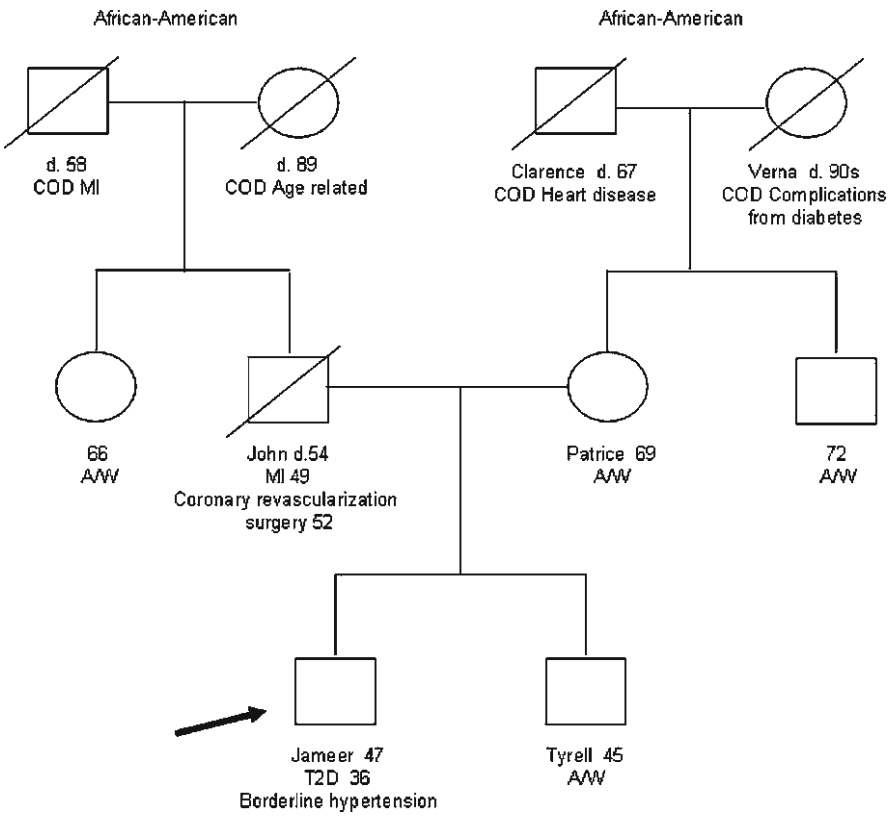

Fig. 6.1 Pedigree illustrating members of Jameer's family. *MI* myocardial infarction, *A/W* alive and well, *DM II* type 2 diabetes

He also tells you that he recently had a genome-wide analysis performed by one of the companies that offers testing directly to consumers, and he has the test report to share with you. He states that he understands some of the findings from the report, but not all of them, and that the sheer volume of information that is contained in the report is overwhelming. For example, given his family history of MI, Jameer is not surprised when the report states that he is heterozygous for an allele of the rs10757278 SNP in 9p21 that is a risk-increasing allele for MI. He is confused, however, when the report states that Jameer has "normal" alleles for the SNP whose variant allele is thought to increase one's risk for T2D.

The testing company's website provides some information about these gene variants and the associated risks. Unfortunately, however, Jameer's primary reason for having the testing performed was to get an accurate estimate of his risks for developing MI, and the information provided does not make him feel confident that he knows his level of risk. He isn't certain what action, if any, he should take in the face of this information, and his uncertainty is compounded by the fact that the report includes a disclaimer that states that the information in the report is not intended to provide medical advice.

The lengthy report provides information on SNPs[1] that are thought to influence one's risk for more than 100 different traits and diseases. For example, the report states that Jameer possesses gene variants that reduce one's risk for asthma, celiac disease, dyslexia and systemic lupus erythematosus, but he also possesses a gene variant that increases one's risk for adult related macular degeneration. The report also states that Jameer is homozygous for a gene allele that increases the rate at which we metabolize caffeine. Finally, the report states that Jameer is heterozygous for a low-activity allele of the *CYP2C19* gene. The CYP2C19 protein is an important drug-metabolizing enzyme that catalyzes the biotransformation of many clinically useful drugs including antidepressants, barbiturates, proton pump inhibitors, antimalarial drugs and antitumor drugs. You decide to refer Jameer to a cardiology clinic that also has the services of a genetic counselor available.

6.7.2 Meeting with the Cardiologist and Genetic Counselor

The genetic counselor and cardiologist meet with Jameer. They examine his medical history and obtain a four-generation pedigree, including a detailed history of nongenetic risk factors for each family member. They do not detect any significant nongenetic factors in the family history. The early age of onset of MI in Jameer's father and grandfather suggests the presence of inherited genetic risk factor(s).

[1] Recall from Chap. 1 that some testing laboratories have adopted the term "single nucleotide variant," abbreviated as SNV, in place of the traditional term "single nucleotide polymorphism," abbreviated as SNP.

The cardiologist and genetic counselor first explain the significance of the 9p21 rs10757278 allele to Jameer. The odds ratio associated with possessing one risk-increasing allele of the rs10757278 SNP is 1.26, so Jameer has a 26% greater risk for MI than the typical person in his age and ethnic group. Given the significant family history of MI, and the fact that Jameer is over age 40 and a male, he is considered to have a higher risk. The cardiologist orders a baseline EKG and additional blood studies, including C-reactive protein, which is reported to be within the normal range.

The cardiologist and genetic counselor next explain why Jameer can have two normal alleles for the T2D marker, despite the fact that he has developed borderline T2D. They explain that, while a genome-wide test may provide information about 1,000,000 genetic markers, this only represents 0.3% of your DNA sequence. There may be dozens or even hundreds of genes that influence one's risk for T2D. Jameer does not have the risk-increasing allele for the gene that was tested, but he may possess one of many other gene variants that can increase the risk for T2D. In addition, although the cardiologist and genetic counselor did not identify any significant nongenetic factors, they explain that nongenetic factors may have contributed to Jameer's father's and grandfather's MIs.

The cardiologist and genetic counselor then discuss the findings that have implications for Jameer's responses to drugs. Jameer is heterozygous for a low-activity allele of the *CYP2C19* gene, which is expected to reduce the rate at which Jameer metabolizes the antiplatelet drug Plavix (Clopidogrel). Plavix is given to reduce the risk of heart attack, unstable angina, stroke, and cardiovascular death in patients with cardiovascular disease. Researchers have found that patients with variant alleles in *CYP2C19* have lower levels of the active metabolite of clopidogrel, less inhibition of platelets, and a greater risk for major adverse cardiovascular events such as death, heart attack and stroke. For this reason, in March 2010 the FDA issued a warning that CYP2C19 poor metabolizers, are at high risk of treatment failure. Because Jameer is heterozygous for the low-activity allele of *CYP2C19*, the cardiologist might consider prescribing a higher dose of Plavix than he/she prescribes for the typical patient, or prescribing a drug that is not metabolized by the CYP2C19 protein.

Another finding from the report that has implications for Jameer's risk for MI involves the report that Jameer is homozygous for a gene variant that increases the rate at which one metabolizes caffeine. Excessive caffeine intake increases one's risk for MI, but because Jameer is a "fast caffeine metabolizer," drinking a moderate amount of coffee should not significantly increase his risk.

Finally, the genetic counselor discusses the finding that Jameer has a risk-increasing allele for AMD. The genetic counselor explains that the risk for macular degeneration is age-related and rarely affects those under age 50. The counselor also explains, however, that this risk-increasing allele contributes to almost half of all AMD cases, with an odds ratio ranging from 2.45 for heterozygotes to 7.4 in homozygotes. In addition, Jameer's history of smoking increases his risk for AMD two-fold.

The genetic counselor goes on to explain that there are a number of preventive measures Jameer can take to help reduce his risk of developing AMD. The counselor

cites the landmark Age-Related Eye Disease Study (AREDS), which found that supplementing the diet with antioxidants plus zinc decreased the likelihood of developing AMD. Jameer was advised to take an eye-specific antioxidant vitamin supplement containing the established dosages of vitamins A, C and E, as well as zinc, selenium and copper. In addition, the antioxidants lutein and zeaxanthine (nutrients that are found in green leafy vegetables such as spinach, kale and collard greens) are under clinical investigation and often contained in these supplements. Jameer was also advised that taking a daily low-dose aspirin may have a preventive effect that is believed to associated with their antiinflamatory activity. Further, in Jameer's case, taking a daily aspirin may also help prevent heart attack and stroke. Finally, Jameer was advised to see an ophthalmologist for routine eye exams, and to wear eye protection that blocks UV rays.

The report also noted that Jameer possesses several other gene variants that are thought to influence an individual's phenotype in ways that are not usually medically significant. For example, Jameer possesses one gene variant that predisposes the individual to accumulate excessive ear wax. In addition, the counselor and Jameer discussed several variants that provided information about Jameer's ethnic heritage.

6.7.3 The Plan for Follow-Up

The plan for Jameer's care is to see Jameer every year in the cardiology clinic for follow up EKG and to monitor other heart disease risk factors. The review of the medical records from Jameer's father and grandfather only noted myocardial infarction, and therefore testing for mutations that cause some of the single-gene disorders that include heart disease among their features was not indicated at the present time. Jameer was advised that each of his family members may also have an elevated risk for MI, T2D and AMD, and should be monitored for these conditions through their healthcare team. A results note detailing the issues that were discussed in the genetic counseling and cardiology session was then sent to you and Jameer. The appropriate recommendations for screening and prevention were included in the note.

6.8 CVD Gene Variants with High Penetrance: Familial Hypercholesterolemia (FH)

Familial hypercholesterolemia and hypertrophic cardiomyopathy are two well-characterized CVDs that are commonly seen in clinical practice, but thought to be under-diagnosed. Early studies of familial hypercholesterolemia, especially the more extreme homozygous forms, led to the development of the HMG-CoA reductase inhibitor (statin) class of drugs so commonly utilized today. These two

disorders are generally thought of as single-gene disorders, and the mutations that occur in the associated genes are often assumed to have 100% penetrance. It has recently been discovered, however, that some of the risk-increasing alleles of these genes have less than 100% penetrance, and that the phenotypes of many of the patients who have these disorders are influenced by the patient's status for other gene variants, as well as nongenetic factors.

6.8.1 FH Is Often Underdiagnosed

Autosomal dominant familial hypercholesterolemia (FH) is one of the most common inherited disorders of lipoprotein metabolism. Inherited forms of FH account for ~5–10% of CVD in patients less than 55 years of age. In the general population, FH affects 1 in 500 individuals. In certain ethnic populations, it is even more common; for example, there is a high frequency of FH in Christian Lebanese (1/170), African (1/70–100) and French Canadian (1/200) populations.

In the heterozygous form of FH, elevated plasma levels of LDL-C and total cholesterol lead to excessive deposition of cholesterol in the arterial walls, accelerated atherosclerosis and premature CVD. The penetrances of these gene variants are high, and without appropriate intervention, ~50% of males and 12% of females will develop CAD by age 50. Lifetime penetrance is close to 100% in males, but only approximately 70% in female carriers. Additional key features may include the presence of achilles tendon xanthomas, xanthelasmata, or an arcus cornealis. There is wide variation in the age of onset and the severity of disease symptoms, illustrating the importance of additional modifying genetic and environmental influences.

Although much less common (approx. 1/1,000,000), a more extreme phenotype is seen in the homozygous form of FH. Plasma LDL-C levels can reach 1,000 mg/dL, with massive atherosclerosis and xanthomas (including planar xanthomas) presenting even in childhood. This form of FH can be lethal at an early age without specific interventions such as LDL apheresis and liver transplantation.

A clinical diagnosis of FH is made by using personal and family history of early onset CVD, abnormal plasma lipids, the presence of tendon xanthomas, or the existence of a known gene mutation. Those meeting any of the diagnostic criteria categories in Table 6.3 should be offered referral to a cardiologist or combined cardiology/genetics clinic. Even with the use of these established diagnostic criteria, FH remains under-diagnosed.

6.8.2 Highly Penetrant FH Gene Variants

All of the genes that are known to be defective in patients with familial FH are involved in either the synthesis of lipoproteins or the receptor-mediated uptake of LDLs by the hepatocytes (Table 6.4). The most common cause of FH is germline mutations in the LDL receptor gene, *LDLR*. More than 1,000 distinct mutations of the

Table 6.3 Dutch lipid clinic network diagnostic criteria for FH[a]

Criteria	Points
Family history	
First-degree relative with known premature (men: <55 years; women: <60 years) coronary and vascular disease, or	
First-degree relative with known LDLC[b] above the 95th percentile	1
First-degree relative with tendinous xanthomata and/or arcus cornealis, or	
Children aged less than 18 years with LDLC above the 95th percentile	2
Clinical history	
Patient with premature (men: <55 years; women: <60 years) coronary artery disease	2
Patient with premature (men: <55 years; women: <60 years) cerebral or peripheral vascular disease	1
Physical examination	
Tendinous xanthomata	6
Arcus cornealis prior to age 45 years	4
Cholesterol levels (mmol/l)	
LDLC, ≥ 8.5	8
LDLC, 6.5–8.4	5
LDLC, 5.0–6.4	3
LDLC, 4.0–4.9	1
DNA analysis	
Functional mutation in the *LDLR* gene	8
Diagnosis (diagnosis is based on the total number of points obtained)	
A "definite" FH[b] diagnosis requires more than 8 points	
A "probable" FH diagnosis requires 6–8 points	
A "possible" FH diagnosis requires 3–5 points	

[a] World Health Organization. Familial hypercholesterolemia – report of a second WHO consultation. Geneva, Switzerland: World Health Organization, 1999. (WHO publication no. WHO/HGN/FH/CONS/99.2)

[b] *LDLC* low density lipoprotein cholesterol, *FH* familial hypercholesterolemia

Table 6.4 Genes with highly penetrant variants that increase risk for FH

Gene	OMIM number[a]	Inheritance	Gene product
LDLR	606945	Autosomal dominant	low density lipoprotein receptor
APOB	144010	Autosomal dominant	ligand-defective apolipoprotein B-100
PCSK9	607786	Autosomal dominant	proprotein convertase subtilisin/kexin type 9
ABCA1	600046	Autosomal recessive	ATP-binding cassette A1
APOA1	107680	Autosomal recessive	apolipoprotein A
LCAT	606967	Autosomal recessive	lecithin-cholesterol acyltransferase

[a] *OMIM* online mendelian inheritance in man (http://ncbi.nlm.nih.gov/omim)

LDLR gene have been described; among patients who meet clinical diagnostic criteria for FH, 50–75% will have *LDLR* mutations. Loss of function mutations in these genes lead to either a deficiency or a complete absence of LDL-receptor function, resulting in two to threefold elevations in plasma levels of LDL-C in heterozygotes, and greater than fivefold increases in homozygotes and compound heterozygotes than controls.

Additional genetic and nongenetic influences are capable of impacting the phenotypes of patients with mutations in these genes. This is particularly true for patients with a specific *LDLR* splice site mutation, IVS14+1 G>A, which by itself results in a truncated protein product, and a resultant 50% reduction in functional LDLR protein. Several SNPs in additional genes further alter the phenotype in patients with this mutation. For example, patients who are simultaneously homozygous for the IVS14+1 G-A mutation and have the −265 C allele of the −265 C>T polymorphism in the promoter region of the apolipoprotein A-II (*APOA2)* gene have significantly lower total cholesterol and LDL-C values. A modifier effect is also seen with the IVS14+1 G>A mutation in patients who have the leu526-to-ile (L526I) substitution of the *GHR* (growth hormone receptor) gene. The lowest levels of plasma HDL are observed among leu/leu homozygotes, highest levels among ile/ile homozygotes, and intermediate levels among leu/ile heterozygotes. A third SNP, the arg287-to-gln polymorphism in the *EPHX2* gene, seems to have the opposite effect when in concert with the IVS14+G>A splice site mutation; the variant allele correlates with elevations in plasma cholesterol and triglycerides.

A second highly penetrant form of autosomal dominant hypercholesterolemia is seen in patients with normal LDLR activity and defective apolipoprotein B-100 (APOB). Familial ligand-defective apolipoprotein B-100, or type B familial hypercholesterolemia (FDB) is due to germline mutations in the ligand-binding domain of the apolipoprotein B100 gene, *APOB*. Approximately 5–10% of all hypercholesterolemic patients have germline mutations in *APOB*.

There is considerable overlap in the phenotypes of patients with FDB versus classic FH due to *LDLR* mutations, although FDB patients have lower plasma LDL-C levels, and a lower risk of CAD. They may also have a better response to statins (discussed below). These differences could be due to the fact that FDB patients maintain normal clearance of very-low-density lipoprotein remnants through apolipoprotein E-mediated uptake. In classic FH patients with *LDLR* mutations, clearance of both LDL and VLDL remnants is affected.

A third autosomal dominant form of hypercholesterolemia is due to mutations in the proprotein convertase subtilisin/kexin type 9 (*PCSK9)* gene. PCSK9 is a serine protease that degrades hepatic LDLR in the endosomes. Three missense mutations (S127R, F216L and D374Y), all of which result in a gain of function, were identified in families with a clinical phenotype resembling FH and FDB. Additional gain-of-function mutations have since been identified, but *PCSK9* mutations account for a much smaller percentage of dominant hypercholesterolemia than do mutations in *LDLR* and *APOB*. Interestingly, loss-of-function *PCSK9* variants (which may be more prevalent in African-Americans) are associated with lower levels of plasma LDL-C, due to increased hepatic LDLR protein levels and accelerated LDL clearance. These mutations therefore reduce the individual's risk of CVD.

Although they are rare, several autosomal recessive forms of FH exist. These forms of FH are caused by mutations in several other genes that play important roles in lipoprotein metabolism (Table 6.4). One such gene encodes the ATP-binding cassette A1 (ABCA1) protein, also known as the cholesterol efflux regulatory protein (CERP). Mutations in *ABCA1* have been found in patients with Tangier disease.

ABCA1 helps provide the cholesterol and phospholipids that are used in the synthesis of HDLs. Another such gene is the lecithin-cholesterol acyltransferase gene (*LCAT*). The LCAT protein, also called phosphatidylcholine-sterol O-acyltransferase, is an enzyme that converts free cholesterol into more hydrophobic cholesteryl esters, allowing them to be packaged into HDLs. Finally, mutations in the gene encoding apolipoprotein A1 (*APOA1*) also produce recessive forms of FH.

Up to one-third of patients with a clinical diagnosis of FH do not harbor any disease-causing variants in the known loci. This suggests either the possibility of additional hypercholesterolemia loci remain to be identified, or the lack of robustness in mutational screening currently employed by various laboratories.

6.9 Molecular Genomic Testing in Patients Suspected of Having FH

There are a number of laboratories that offer FH gene testing, with the cost approximately $1,200 per gene. As is true for all disorders, the NIH's GeneTests website (http://www.ncbi.nlm.nih.gov/sites/GeneTests/lab/clinical_disease_id/26358?db=genetests) provides the most comprehensive and current listing of laboratories that offer genomic testing for FH. A search of the website for FH tests produces four listings: tests for FH, FH Type B, the autosomal dominant form of FH and the autosomal recessive form of FH. As described in Chap. 3 (see Sect. 3.2), the report will detail any variants that are discovered that are known to have, or may possibly have, clinical significance. If the company reports that no deleterious variants have been found, however, it is important to know whether there were no actual variants found, or whether there were variants found, but those variants are believed to have no clinical significance. There are several situations in which a variant that is not believed to have clinical significance may actually have clinical significance, and future research may bring this to light.

6.10 CVD Gene Variants with High Penetrance: Hypertrophic Cardiomyopathy (HCM)

6.10.1 The HCM Phenotype Is Highly Variable

Hypertrophic cardiomyopathy is a complex multifactorial disease that is caused by a genetic defect of the cardiac sarcomeric apparatus. It is the most common inherited CVD, with prevalence in young adults of 1 in 500. The natural history and phenotypic expressivity of this disorder is highly variable, attesting to the influence of other genetic factors and nongenetic factors on the phenotype. Because it is a frequent cause of sudden death, particularly in young individuals and competitive athletes, it is essential to identify patients with this condition.

HCM is characterized by myocardial hypertrophy, usually of the left ventricle, typically in the absence of other loading conditions such as hypertension or aortic stenosis. In contrast to the universal presence of left ventricle hypertrophy, there is considerable variability in other aspects of the HCM phenotype. Some individuals remain asymptomatic throughout life. Others may present with symptoms ranging from palpitations and dizziness to syncope and sudden death. Notably, sudden death often occurs during exercise, but also demonstrates a circadian distribution, with clustering of deaths in the morning and early evening. This clinical heterogeneity is a reflection of the underlying and complex pathophysiology, which includes diastolic dysfunction, ventricular arrhythmias, small-vessel disease leading to subendocardial ischemia, and left ventricular outflow tract obstruction. Several of these latter factors are highly variable among patients with mutations in the HCM genes. For example, only 25% of HCM patients have left ventricular outflow tract obstruction (LVOTO), while only 70–90% of HCM patients have an abnormal ECG.

Transthoracic echocardiology is used to identify the hallmark feature of HCM, asymmetric hypertrophy of the interventricular septum, with or without left ventricular outflow tract obstruction and systolic anterior motion of the mitral valve. The extent of left ventricular hypertrophy can vary, even between individuals of the same family who have the same gene mutation, suggesting that there are a number of important additional genetic and nongenetic modifiers, such as blood pressure, exercise, diet and body mass.

There are a number of other genetic disorders that present with increased left ventricular wall thickness, such as Noonan syndrome, Friedreich ataxia, Swyer syndrome and metabolic disorders such as Fabry disease and glycogen storage disorders. Thus, referral to a combined cardiology/genetics clinic is appropriate for patients with HCM and any of the features that are seen in these disorders.

6.10.2 Highly Penetrant Gene Variants That Increase Risk for HCM

Mutations in the genes that are listed in Table 6.5 account for 70–95% of all HCM cases. HCM is inherited in an autosomal dominant manner, although there are reports of individuals who have mutations in two of the relevant HCM genes. As can be seen from Table 6.5, most of the genes that influence risk for HCM encode protein components of the cardiac sarcomeric apparatus. Four genes encode components of the thick filament (*MYH7*, *MYBPC3*, *MYL2* and *MYL3*), five genes encode thin filament proteins (*TNNT2*, *alpha-TM*, *cTn1, ACTC*, *cTnC*).

Patients with mutations in two additional non-sarcomeric genes may present with increased wall thickness mimicking HCM. These genes, PRKAG2 and LAMP2, are involved with glycogen accumulation in cardiac myocytes; mutations alter myocarcial metabolism, resulting in increased wall thickness, cardiac storage abnormalities and conduction irregularities. Patients with these mutations are classified as storage disorders, or metabolic HCM.

Table 6.5 Genes with highly penetrant variants that increase risk for HCM

Gene	Protein	Protein function	Disease %[a]
MYH7	Myosin heavy chain 7	Thick filament protein	30–35
MYH6	Myosin heavy chain 6	Thick filament protein	<0.5
MYBPC3	Myosin binding protein C (thick filament protein)	Thick filament protein	20–30
MYL2	Regulatory light chain (thick filament protein)	Thick filament protein	<1
MYL3	Essential light chain (thick filament protein)	Thick filament protein	<1
TNNT2	Cardiac troponin T (thin filament protein)	Thin filament protein	10–15
Alpha-TM/TPM1	Alpha-tropomyosin (thin filament protein)	Thin filament protein	5–15
cTnl	Cardiac troponin I (thin filament protein)	Thin filament protein	<5
cTnC	Cardiac troponin C (thin filament protein)	Thin filament protein	<0.5
ACTCI	Cardiac alpha actin 1	Thin filament protein	<0.5
CSRP3	Cardiac muscle LIM protein	Stimulates myogenesis; possible role in stretch sensing	<5
TCAP	Telethonin	Scaffold for sarcomere assembly	<2
TTN	Titin	Scaffold for sarcomere assembly	<0.5
VCL	Vinculin	Cytoskeletal protein; anchors F-actin to membrane	<0.5
ACTN2	Alpha-2 actinin	Helps anchor actin filaments	<0.5
GLA	Alpha galactosidase	Lysosomal enzyme	<0.5
LAMP2	Lysosomal-associated membrane protein 2	Assembly, maintenance and function of the lysosome	<0.5
PRKAG2	Protein kinase, AMP-activated, gamma 2 non-catalytic subunit	Regulates synthesis of fatty acids and cholesterol	<0.5
PLN	Phospholamban	Inhibitor of cardiac muscle sarcoplasmic reticulum Ca^{++}-ATPase	<0.5
CAV3	Caveolin 3	Membrane protein involved in endocytosis	<0.5

[a] Percent of patients with HCM with detectable causes who possess mutations in these genes

As evident in Table 6.5, approximately 70–90% of all HCM mutations are in three genes: *MYH7, TNNT2* and *MYBPC3*. As noted above, the *MYH7* gene product is the B-myosin heavy chain protein involved in cardiac muscle contractility. The majority of the mutations occur in critical functional domains of the myosin head or head-rod junction of the protein, and many of these mutations are associated with early onset and poor clinical prognosis including sudden cardiac death. Certain mutations in the *TNNT2* gene (e.g. R92Q), have been associated with high incidence

of sudden death, even with mild hypertrophy. In contrast, patients with mutations of the *MYBPC3* gene may have a later onset and a better prognosis.

Generally, individuals with HCM-associated mutations are heterozygous. However, recent genetic studies in HCM have suggested that up to 5% of families may carry two distinct disease-causing HCM mutations. Individuals with two distinct highly penetrant HCM mutations can be homozygotes or compound heterozygotes (if the mutations are in the same gene) or double heterozygotes (mutations in different genes). Compared with individuals with only one mutation, the presence of two disease causing HCM mutations typically causes a more severe phenotype, with more severe left ventricular hypertrophy and a higher incidence of sudden cardiac death. They are usually younger at diagnosis, sometimes even presenting with childhood-onset hypertrophy. This suggests a dosage effect with greater disruption of normal sarcomere function resulting in a more severe clinical course.

Because the *MYH7, MYBPC3* and *TNNT2* genes account for such a large percentage of HCM mutations, testing is usually tiered. Depending on cost considerations, the first test may include all three of these genes, or the analyst may test them individually (*MYH7* first, *MYBPC3* second and *TNNT2* third), stopping when a causative mutation is identified.

6.11 Therapy and Prevention Measures for HCM Mutation Carriers

There are various medical and surgical therapy approaches for HCM carriers. Paramount to any therapy is reduction in the risk of sudden death by the early identification of high-risk patients and effective medical intervention and/or surgical implantation of an automatic defibrillator for the prevention of sudden arrhythmic death. Additional therapies may include left ventricular myomectomy, mitral valve replacement, pacemaker implantation, and transvenous catheter ablation of the septal region. Appropriate referrals include consultations with cardiologists, cardiothoracic surgeons, cardiac electrophysiologists and geneticists.

Avoidance of strenuous exercise and competitive level sports is critical, given the history of sudden death in patients (and other at-risk family members) with HCM. This includes abstaining from highly strenuous physical exertion, such as shoveling snow or lifting heavy objects. Participation in noncompetitive recreational sports activities is not believed to be contraindicated.

6.12 Genes Influencing the Risk for Other Cardiomyopathies

Several genes, including some that are listed in Table 6.4 (*ACTC1*, MYH7, *MYBPC3*, *TNNT2*, *TNNI3*, *TNNC1* and *TPM1)* also influence the individual's risk for dilated cardiomyopathy (DCM). In addition, the genes encoding the LIM

domain binding 3 (*LDB3)*, lamin A/C (*LMNA*), phospholamban (*PLN*) and the alpha subunit of the type V voltage-gated sodium channel (*SCN5A*) are all risk factors for DCM. Mutations in these genes account for only 25% of cases of DCM, however, attesting to the large number of DCM risk genes that exist.

Lone atrial fibrillation (AF) aggregates in families; nearly 30% of lone AF patients have an affected first-degree relative. One of the SNPs in the *KCNN3* gene, which encodes a voltage-independent calcium-activated potassium channel protein, has been reported to influence the risk for lone AF.

A completely different set of genes has been implicated as risk factors for arrhythmogenic right ventricular cardiomyopathy (ARVC). Mutations in the genes that encode the desmocollin 2 (*DSC2*), desmoglein 2 (*DSG2*), desmoplakin (*DSP*), plakophilin (*PKP2*) and the transmembrane protein 43 (*TMEM43*) proteins account for 40–50% of cases of ARVC. The functions of the TMEM43 protein are not well known. All these other proteins, however, are integral parts of the desmosomes that hold epithelial cells tightly together. These findings are providing insights into the molecular pathology of these specific CVDs, but considerable research is needed before these basic findings are translated into clinically useful tests.

6.13 Cardiovascular Pharmacogenomics

At this point in time, there are a few commercially available pharmacogenomic tests that can help guide treatment decisions and predict the patient's response to cardiac drugs. Unfortunately, this first generation of tests is not as comprehensive as these test panels need to be. As discussed in Chap. 4, it has been easier to identify genetic markers that influence the pharmacokinetics of the drugs of interest, but it is also critical to identify markers that influence the pharmacodynamics of the drugs, as well as the pathways and processes that maintain the integrity of the tissues and pathways with which the drugs interact. It has long been known that the level of plasma triglycerides and lipoproteins, as well as several aspects of lipid metabolism, influence the individual's risk for CVD. In addition, animal models have identified several other pathways that influence the individual's susceptibility to CVD, including sodium and potassium regulation and oxidative stress.

6.13.1 The CYP450 Enzymes Metabolize Several Drugs That Are Prescribed for CVDs

As was illustrated in Chap. 4 (see Sect. 4.6), the CYP450 enzymes metabolize approximately 40% of the drugs that are prescribed today. Because the adverse drug reactions (ADRs) that may result from cardiac drugs can be life-threatening, it is critical to predict those poor metabolizers who are at risk for ADRs.

Antiarrhythmic drugs generally have low therapeutic indices, so genotype-dependent differences in metabolism may have great clinical relevance. Among the antiarrhythmics, CYP2D6 is most significantly involved in the metabolism of propafenone, flecainide and mexiletine. Some (but not all) of the evidence suggests that reduced CYP2D6 activity results in greater efficacy of propafenone, but also a higher frequency of ADRs involving the central nervous system. In addition, one small study suggested that patients who have the "poor metabolizer (PM)" CYP2D6 phenotype experience more nausea and lightheadedness after mexiletine than patients who have the "extensive metabolizer (EM)" phenotype do.

CYP2D6 is not involved in the metabolism of several commonly prescribed beta-blockers, but CYP2D6 does play a significant role in the metabolism of carvedilol, metoprolol, propranolol and timolol. Not only do PMs achieve higher blood levels of timolol (after oral, inranasal or intraocular administration), but the beta-blocking effects are of greater magnitude and duration than in EMs. These studies are complicated by the fact that the different enantiomers of these drugs often have different biological activities. In addition, the clinical significance of beta-blocker metabolism includes the fact that some drugs have alpha-blocking activity as well. Such is the case with the R-enantiomer of carvedilol. One study has reported that CYP2D6 PMs (volunteers) had significantly reduced systolic blood pressures during 1 week of carvedilol administration.

CYP2D6 status clearly influences the clearance of metoprolol; clearance is not only reduced in PMs after a single dose, but also after repeated dosing. The studies that have assessed clinical outcomes are suggestive, but not unanimous. For example, studies using healthy volunteers have reported that the beta-blocking effects of metoprolol are prolonged in PMs versus EMs. In addition, one study has reported an excessively high frequency of the CYP2D6 PM phenotype in a group of patients who had experienced ADRs after metoprolol. In contrast, one study reported that there was no difference in the frequency of ADRs after sparteine administration (sparteine is also metabolized by CYP2D6) in patients who tolerated metoprolol versus patients who suffered ADRs after metoprolol. In addition, three studies reported no relationship between CYP2D6 status and the frequency of ADRs after metoprolol. Interestingly, in one study, the blood pressure changes did not correlate either with CYP2D6 status or metoprolol concentration.

The antianginal drug perhexiline is extensively metabolized by CYP2D6, so it is no surprise that CYP2D6 status significantly influences metabolism of the drug. Perhexiline has two serious, and concentration-dependent, side effects: hepatotoxicity and peripheral neuropathy. Both these ADRs occur more frequently in CYP2D6 PMs than in EMs. At least one study has suggested that the dose of perhexiline should differ depending on the patient's CYP2D6 status.

6.13.2 The Pharmacogenomics of Cholesterol-Lowering Drugs

All patients with FH, regardless of the underlying cause, should maintain a healthy diet, get regular exercise and control their weight. The 2004 National Cholesterol

Education Program (NCEP) Adult Treatment Panel III (ATPIII) defines target LDL-C levels and levels based on risk, and is a useful resource for patients to consult (http://www.nhlbi.nih.gov/guidelines/cholesterol). Improving the diet alone can result in up-regulation of LDL receptors, although the effect is limited, especially in the more severe phenotype of the homozygous FH state, as there is little to no receptor activity to up-regulate. In these patients, interventions such as LDL apheresis or portacaval anastomosis are necessary.

It is well known that lowering the level of LDLs and triglycerides and raising the level of HDLs in the blood reduces the individual's risk for CVD. Statin drugs have been shown to achieve these effects for many patients, and they are among the most commonly prescribed drugs. A number of studies have attempted to identify gene variants that affect blood triglyceride, cholesterol and lipoprotein levels, as well as the individual's response to the drugs that are prescribed for patients with hypercholesterolemia.

As is true for so many drug classes, the individual's status for polymorphisms in the *CYP450* genes and *ABC* drug transporter genes influence the pharmacokinetics of some of the statin drugs. Several of the statins are substrates for some of the CYP450 enzymes, including CYP3A4, CYP3A5, CYP2C9, CYP2D6 and CYP2C8, and the CYP450 tests discussed in Chap. 4 (see Sect. 4.6) can help predict the patient's response to some of these drugs. For example, the CYP3A5*3 allele contains a single nucleotide substitution in intron 3 that produces a truncated protein with no function. Testing the individual's status for the *CYP3A5* gene can predict the response to lovastatin, symvastatin and atorvastatin, which are metabolized by the CYP3A5 enzyme. In addition, another study has reported that the CYP3A5*3 allele led to increased elevation of creatine kinase in patients who experienced myalgia after atorvastatin. Variation in CYP3A5 activity may be a significant contributor to the difference in statin response seen in Caucasian versus African-American patients; the active *CYP3A5*1* allele is the major allele in African-Americans, but present in only 10% of Caucasians. In addition, the CYP7A1 protein catalyzes the rate-limiting step in bile acid synthesis, thereby indirectly regulating cholesterol synthesis and plasma LDL levels. The individual's status for a polymorphism in the promoter region of *CYP7A1* influences plasma LDL levels as well as the efficacy of atorvastatin.

Several types of transport proteins, including ABCB1, ABCG5, ABCG8 and SLCO1B1 (OATP1B1) transport several cholesterol-lowering drugs, and the individual's status for polymorphisms in these genes has been reported to influence his/her response to some statin drugs as well. Further, the individual's status for the *CYP7A1* gene interacts with his/her status for the *ABCG8* gene to influence the response to atorvastatin. Patients who possessed both the H allele of the D19H polymorphism in the *ABCG8* gene and the C allele of the A-204 C polymorphism in the *CYP7A1* gene promoter achieved the greatest reductions of LDL cholesterol with statin therapy.

In addition to these pharmacokinetically relevant polymorphisms, there have been a number of pharmacodynamically relevant polymorphisms discovered by recent research. Statin drugs inhibit the activity of 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGCR), which catalyzes the rate-limiting step in hepatic cholesterol

synthesis. The *HMGCR* gene, therefore, is a good candidate in which to find pharmacodynamically relevant polymorphisms. The Pravastatin Inflammation/CRP Evaluation (PRINCE) study analyzed 148 SNPs in ten genes, including *HMGCR*, and reported that the individual's status for two linked SNPs in *HMGCR* influenced his/her response to pravastatin. Patients who possessed the minor allele of one of these two *HMGCR* SNPs achieved approximately 20% smaller reductions in both total cholesterol and LDL cholesterol than patients who possessed the more common allele.

Another enzyme that plays an important role in cholesterol synthesis is squalene synthase, which is encoded by the *FDFT1* gene. The PRINCE study also reported that the individual's status for a SNP in the squalene synthase gene influences the individual's baseline lipid levels, as well as the response to pravastatin.

Lipoprotein metabolism represents another aspect of lipid metabolism that promises to provide important sources of genetic variability. The most extensively studied gene in the field of cardiovascular genetics is the apolipoprotein E (*APOE*) gene. The APOE protein binds the hepatic cell receptors, enabling the hepatic cells to take the lipoproteins out of the circulation. The E4 allele of the APOE gene is a well-known risk-increasing allele for CVD (and Alzheimer disease, see Sect. 7.5). Further, the individual's status for the *APOE* gene influences the efficacy of some of the common treatments for hyperlipidemia. Patients who possess the E4 allele of the *APOE* gene have been reported to respond better to dietary interventions than patients with other *APOE* alleles. Patients who possess the E2 allele, on the other hand, are more likely to respond well to statin drugs than patients with the other *APOE* alleles. Some research has suggested that this latter effect may be more pronounced in male patients than in female patients.

Another focus of these studies has been the cholesteryl ester transfer protein (CETP), which transfers triglycerides and cholesteryl esters between different classes of lipoproteins. A meta-analysis of seven large, population-based studies and three pravastatin clinical trials confirmed that the individual's status for the TaqIB polymorphism in the *CETP* gene influences his/her level of CETP, as well as his/her plasma HDL level. In addition, the *CETP* allele (the B1 allele) that is associated with lower HDL levels is also associated with a higher risk for CAD and faster progression of coronary arteriosclerosis. In contrast, however, this study was unable to confirm the association between CETP status and response to pravastatin that had been reported in a few previously published studies.

As we learn more about the regulation of the genes that influence lipid metabolism, we discover more sources of genetic variability. For example, the sterol-regulatory element binding proteins (SREBPs) and the peroxisome proliferator-activated receptors (PPARs) are transcription factors that regulate the activity of several genes whose proteins are involved in lipid homeostasis. In one study, individuals who possessed the insertion allele of a single nucleotide insertion/deletion polymorphism in the *SREBF1a* gene achieved a greater increase in apoA-I levels after fluvastatin than individuals who possessed the deletion allele. In addition, one study has reported that certain PPAR haplotypes influence the degree of increase in HDL levels and artery lumen diameter after fluvastatin.

As discussed above, genes whose proteins participate in the inflammation response also influence the risk for CVDs and some of these variants also influence the response to cardiovascular drugs. For example, patients who are homozygous for the C allele of the G174C polymorphism in the *IL-6* gene achieve significantly greater reductions in risk for coronary heart disease after statin drug therapy than patients with the CG or GG genotypes.

In FH patients who are homozygous for mutations in the *LDLR* gene, plasma LDL-C levels are so high that it is often necessary to combine high doses of a statin drug with LDL apheresis. One of the side effects of this therapy, however, is a transient reduction in HDL levels in the blood during the first 4 weeks of statin therapy. The magnitude of the transient HDL reduction is higher in patients with *LDLR* mutations than in those who do not have *LDLR* mutations, with receptor-negative patients showing greater reductions than receptor-defective patients. In addition, after exended therapy, HDL cholesterol levels remain higher in receptor-defective patients than in receptor-negative patients.

Fibrates are sometimes used in conjunction with statins, or in patients for whom statin therapy is not appropriate. The patient's status for polymorphisms in the *APOE* and *APOA* genes influence the effectiveness of fibrates. In patients with primary hypertriglyceridemia or mixed hyperlipidemia, those who possess the E4 allele of the *APOE* gene achieve less of a reduction in apolipoprotein B, apolipoprotein E and triglyceride levels than do patients with the E3 allele, who in turn achieve lesser reductions in these three measures than patients with the E2 allele do. In addition, three SNPs in the *APOA5* gene have been reported to influence the response to fenofibric acid.

6.13.3 Pharmacogenomic Testing Is Particularly Important for Patients Taking Warfarin

It is particularly important to optimize each patient's dose of warfarin. If the dose is too low, the patient will not respond to the drug. If the dose is too high, the patient is at risk for excessive bleeding. The patient's international normalized prothrombin ratio (INR) is carefully monitored, to insure proper dosing.

The currently available genetic tests that predict warfarin response focus on variants in the *CYP2C9* and *VKORC1* genes. CYP2C9 plays an important role in the metabolism of several coumarin anticoagulants that have low therapeutic indices, including warfarin, acenocoumarol and phenprocoumon. The homozygous *3/*3 genotype results in a 90% decrease in clearance of the active S-enantiomer of warfarin. Patients with the CYP2C9*2 and *3 alleles require considerably lower doses of warfarin, and several studies have reported higher frequencies of bleeding episodes and longer times to establish a stable regimen in these patients. Patients with the *3 allele also require significantly lower doses of acenocoumarol, but patients with the *2 allele require doses similar to those required by patients with the more active CYP2C9 alleles.

In addition, warfarin inhibits activity in the VKOR complex, which regenerates reduced vitamin K from oxidized vitamin K. The individual's status for a functional polymorphism in the promoter region of the *VKORC1* gene (-1639A/G), which encodes the major subunit of the vitamin K oxidoreductase (VKOR) complex, plays an important role in determining the proper dose of warfarin for that individual. As discussed in Sect. 4.4, recent research indicates that the optimal dose of warfarin for each individual patient can be predicted most accurately by considering not only *VKORC1* and *CYP2C9* genotypes, but also nongenetic factors such as age, height, body weight and the coadministration of certain drugs that interact with warfarin.

Another potentially important side effect of warfarin is tissue necrosis. A deficiency in the innate anticoagulant protein C or its cofactor, protein S, increases the patient's risk for tissue necrosis following warfarin administration. Protein C and/or S deficiencies are rare enough, however, that testing will probably be restricted to patients who have experienced tissue necrosis.

6.13.4 The Pharmacogenomics of Antihypertensive Drugs

Several gene variants are known to influence either the pharmacokinetics or pharmacodynamics of antihypertensive drugs. Table 4.1 lists the antihypertensive drugs whose pharmacokinetic profiles are influenced by variants in the CYP2D6, CYP2C19 and CYP2C9 enzymes. CYP2C9 plays an important role in the metabolism of several of the angiotensin II receptor blocker drugs that are used to treat high blood pressure, diabetic nephropathy and congestive heart failure. With losartan, possessing the CYP2C9*3 allele clearly reduces the accumulation of the active moiety (the metabolite E-3174), but possessing the *2 allele does not seem to have the same effect. Consistent with the pharmacokinetic data, one study has reported that individuals (healthy volunteers) with the *1/*3 genotype had a reduced response to losartan.

Angiotensin converting enzyme (ACE) inhibitors are also prescribed for many patients with essential hypertension, and the 287 bp insertion/deletion (I/D) polymorphism in intron 16 of the *ACE* gene is a functional polymorphism. Not only do individuals with the DD genotype of this polymorphism have an increased risk for several CVDs, they also benefit less from treatment with ACE inhibitors than patients with the ID or II genotypes do.

As discussed in Chap. 4 (see Sect. 4.7.1), the enzyme N-acetyltransferase (*NAT*) was the subject of some of the first pharmacogenomic studies. The antihypertensive drug hydralazine is a substrate for NAT, and variants in *NAT* will influence the efficacy of hydralazine, as well as the patient's risk for ADRs. *NAT* variants also influence metabolism of isosorbide mononitrate, which is used to prevent angina attacks.

6.13.5 Plasma Homocysteine Levels Influence Pyridoxine Response

As discussed above, low activity in the CBS protein will result in elevated blood homocysteine levels and increased risk for cardiovascular disease. In addition, the patient's status for the G919A (G307S) polymorphism in the *CBS* gene affects his/her response to pyridoxine therapy. Patients with the A allele of this polymorphism do not respond as well to pyridoxine as patients with the G allele do.

6.13.6 CYPC19 Testing for Patients Prescribed Clopidogrel

Because clopidogrel is metabolized to its primary active compound, patients with the poor-metabolizer CYP2C19 phenotype have decreased active metabolite levels, and may demonstrate a reduced antiplatelet response. It may be more advisable to prescribe alternative drugs, such as Prasugrel, for these patients. Prasugrel does not appear to be a substrate for CYP2B6, CYP2C9, CYP2C19 or CYP3A5.

It is particularly beneficial to assess the CYP2C19 status in patients who are prescribed proton pump inhibitors (PPIs) such as esomeprazole, omeprazole and lansoprazole to treat the peptic ulcers that are sometimes seen in patients on clopidogrel. These PPIs inhibit the activity of CYP2C19; it has been hypothesized that giving them to a patient who has low CYP2C19 activity may reduce the conversion of clopidogrel sufficiently to prevent the patient from achieving any therapeutic response to the drug. Interestingly, however, one recent study has reported that PPIs were more effective at inhibiting the effects of clopidogrel in patients who do not have any null alleles for *CYP2C19*. It may be that, in patients with low CYP2C19 activity, other CYP450 enzymes take over the metabolism of clopidogrel, thereby reducing the interaction between PPIs and clopidogrel.

Further Readings

Ackerman MJ, Mohler PJ (2010) Defining a new paradigm for human arrhythmia syndromes: phenotypic manifestations of gene mutations in ion channel- and transporter-associated proteins. Circulation Research 107(4):457–65

Campuzano O, Beltrán-Alvarez P, Iglesias A, Scornik F, Pérez G, Brugada R (2010) Genetics and cardiac channelopathies. Genetics In Medicine 12(5):260–7

Cowan J, Morales A, Dagua J, Hershberger RE (2008) Genetic testing and genetic counseling in cardiovascular genetic medicine: overview and preliminary recommendations. Congestive Heart Failure14(2):97–105

Darbar D (2010) Genomics, heart failure and sudden cardiac death. Heart Failure Review 15(3):229–38

Dorn GW, Cresci S (2009) Genome-Wide Association Studies of Coronary Artery Disease and Heart Failure: Where Are We Going? Pharmacogenomics 10(2):213–223

Kolovou GD, Kostakou PM, Anagnostopoulou KK (2011) Familial hypercholesterolemia and triglyceride metabolism. International Journal of Cardiology 147(3):349–358

McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Tybjaerg-Hansen A, Folsom AR, Boerwinkle E, Hobbs HH, Cohen JC (2007) A common allele on chromosome 9 associated with coronary heart disease. Science 316(5830):1488–91

Roden DM (2008) Clinical practice. Long-QT syndrome. New England Journal Medicine 358(2):169–76

Scheuner MT (2004) Clinical application of genetic risk assessment strategies for coronary artery disease: genotypes, phenotypes, and family history. Primary Care 31(3):711–37

Chapter 7
Other Multifactorial Disorders for Which Genetic/Genomic Testing Is Providing Insights

Abstract This chapter describes the progress that has been made in personalizing medicine in several fields other than cancer and cardiovascular disease. A brief review is given of the gene variants that help predict, diagnose and treat age-related macular degeneration, type 2 diabetes mellitus, psychiatric disorders, substance abuse, Alzheimer disease and age-related cognitive decline, asthma and other respiratory disorders. The chapter discusses the clinically useful tests that have already been developed, several promising early discoveries that must now be reevaluated, and a number of basic research discoveries that will provide the basis for developing clinically useful tests in the near future.

7.1 Age-Related Macular Degeneration

Age-related macular degeneration (AMD) is a complex degenerative disease of the macula. There are two major clinical presentations: dry (non-neovascular, atrophic) AMD, characterized by the degeneration of choriocapillaries, retinal pigment epithelium and neurosensory retina; and wet (neovascular) AMD, characterized by the development of serous retinal pigment epithelium detachments and/or new choroidal vessels that can lead to bleeding, exudation and eventual scar formation. Although only 10% of patients have the wet, neovascular form, it is responsible for >80% of severe vision loss and legal blindness related to AMD.

Risk is age-related, as AMD rarely affects those under age 50. Current and former smokers have a two-fold risk for developing AMD when compared to non-smokers. Female gender and obesity also increase one's risk. Individuals with a single relative with AMD have a two to threefold increase in their risk, while those with two or more relatives are nearly four times as likely to be diagnosed. The risk is even higher if the affected family members were diagnosed before the age of 65.

Early investigations into the genetic risk factors for AMD benefitted from the use of well-defined AMD phenotypes (ex. only patients with a very large drusen were included in the initial study), which reduced phenotypic heterogeneity. A common intronic variant in the complement factor H (*CFH*) gene was the first

K.M. Sweet and R.C. Michaelis, *The Busy Physician's Guide To Genetics, Genomics and Personalized Medicine*, DOI 10.1007/978-94-007-1147-1_7,

genetic risk factor associated with the dry (non-neovascular) form of AMD. Further sequencing and fine mapping revealed that this intronic variant was tightly linked to a tyr402his SNP in the *CFH* coding sequence (rs1061170). The risk-increasing allele of this SNP contributes to almost half of all AMD cases, with an odds ratio ranging from 2.45 for heterozygotes to 7.4 in homozygotes.

Nongenetic factors clearly interact with *CFH* status to determine risk for AMD. For example, obesity increases the risk for individuals who are homozygous for the risk-increasing *CFH* allele from 7.4 times to 12 times that of the general population. In addition, *CFH* homozygous who smoked one pack of cigarettes per day for 10 years or more had an approximately 144-fold increase in disease risk compared with individuals who had smoked fewer than 10 pack-years and had at least one copy of the other *CFH* allele. It is important to make your patients aware of the potential for interactions between genetic factors and nongenetic factors such as these, because knowing that these controllable nongenetic factors have a strong influence over risk may motivate patients to reduce their risk-increasing behaviors.

A second putative AMD susceptibility locus (*LOC387715*) has been identified, but it is uncertain what protein is encoded by this locus. The T allele of an ala69ser SNP (rs10490924) in *LOC387715* confers an OR of 2.5 in heterozygotes, and 7.3 in homozygotes. This polymorphism also interacts with the tyr406his SNP in *CFH*. Patients who are homozygous for the risk-increasing alleles of these two SNPs have an odds ratio for development of AMD of 57.6, and they might be at greater risk for earlier onset of neovascular AMD. Smoking seems to further elevate this risk, as one study estimated that *CFH*, *LOC387715* and smoking might explain >60% of the attributable risk for AMD.

7.2 Type 2 Diabetes

According to the American Diabetes Association, there are approximately 16–17 million children and adults in the U.S. with Type 2 diabetes mellitus (T2D). T2D is characterized by peripheral insulin resistance, with an insulin secretory defect that varies in severity. T2D was once known as adult-onset diabetes, but because of the growing number of obese children in the U.S., the average age at which T2D is diagnosed is decreasing. It has equal incidence in women and men in most populations. Major risk factors include age, obesity, hypertension or dyslipidemia, history of gestational diabetes, or history of polycystic ovarian syndrome (which results in insulin resistance). T2D is more prevalent among Hispanics, Native Americans, African Americans and Asians/Pacific Islanders than non-Hispanic whites. Having one first-degree relative affected with T2D increases the individual's risk to twice that of the general population, while having more than two close relatives affected increases the individual's risk to four times that of the general population.

7.2.1 GWA Studies Have Identified Several Genetic Markers That Increase Risk for T2D

The hunt for genetic variants that predict the risk for T2D began a few decades ago, but after hundreds of candidate gene studies and more than 30 genome-wide linkage scans, all the variants that have been identified have very low penetrance. The gene that has been most convincingly associated with risk for T2D is the transcription factor 7-like 2 (*TCF7L2*) gene. The TCF7L2 protein is a high mobility group box-containing transcription factor that stimulates proliferation of pancreatic beta cells. A large linkage study involving Icelandic subjects identified the T allele of the *TCF7L2* SNP rs7903146 as a risk-increasing allele for T2D. This association has been replicated in more than 20 studies across different populations with diverse ancestral backgrounds. About 10% of the European and African population are homozygous for the T allele of rs7903146. The frequency of the risk-increasing allele is lower in Asian and Hispanic populations, but the relative risk associated with the risk-increasing allele is the same across ethnicities.

Table 7.1 lists a number of other genes that either contain or reside close to polymorphisms that have been reported to influence the individual's risk for T2D. All these associations involve relatively small per-allele odds ratios (between 1.1 and 1.3), and all these variants together account for only ~3% of the heritability of T2D.

Besides *TCF7L2*, only four of these additional T2D-associated variants (*PPARG*, *KCNJ11*, *WFS1* and *HNF1B*) were originally identified through candidate gene studies, which themselves were based on the involvement of rare mutations in monogenic forms of diabetes. The PPARG protein is the target for thiazolidinedione drugs. Interestingly, in addition to several risk-increasing loss-of-function alleles, a gain-of-function variant (Pro12Ala) has been described in *PPARG* that decreases the risk of insulin resistance This allele has a frequency as high as 0.12 in some populations.

Several of these findings illustrate the ability of GWA studies to identify genes that make small contributions to the disorder in question and provide new insights into the pathophysiology of disease. Without the use of GWA studies, some of these genes would never have been implicated in the etiology of T2D. Examples include the fat mass and obesity-associated (*FTO*) gene, which is now associated with BMI and the risk of being overweight or obese, and the melatonin receptor 1B (*MTNR1B*) gene, which helps regulate the circadian rhythm of fasting glucose levels.

Some of the T2D variants are also associated with other disorders. For example, *CDKAl1* is also associated with Crohn's disease and psoriasis. *TCF7L2* is also associated with colon cancer, while *HNF1B* and *JAZF1* are also associated with prostate cancer. This speaks to the functional diversity of the *T2D*-associated genes and the multitude of pathways in which they participate.

Table 7.1 Genes that contain or lie close to T2D-associated SNPs

Gene name and abbreviation	Protein function
TCF7L2; transcription factor 7-like 2	Stimulates pancreatic beta cell proliferation
SLC20A8; cation efflux transporter ZnT8	Zinc transporter expressed only in pancreatic beta cells
PPARG; peroxisome proliferator-activated receptor gamma	Transcription factor, regulates genes involved in carbohydrate and lipid metabolism
KCNJ11; potassium inwardly-rectifying channel, subfamily J, member 11	Potassium ion channel associated with the sulfonylurea receptor
WFS1; Wolfram syndrome 1	Endoplasmic reticulum protein highly expressed in pancreas
HNF1B; hepatocyte nuclear factor 1 homeobox B	Transcription factor expressed in liver
HHEX/IDE; hematopoietically expressed homeobox	Possible role in hematopoietic cell differentiation
CDKAL1; CDK5 regulatory subunit associated protein 1-like 1	Regulates cell cycle progression and apoptosis
IGF2BP2; insulin-like growth factor 2 mRNA binding protein 2	Regulates translation of insulin-like growth factor 2
CDKN2B; cyclin-dependent kinase 4 inhibitor B	Regulates cell cycle progression and apoptosis
CDKN2A; cyclin-dependent kinase inhibitor 2A	Regulates cell cycle progression and apoptosis
FTO; fat mass and obesity associated	Expressed in hypothalamus and pancreas; upregulated after food deprivation
KCNQ1; potassium voltage-gated channel, KQT-like subfamily, member 1	Regulates potassium ion currents on heart and kidney
NOTCH2; neurogenic locus notch homolog protein 2	Cell differentiation
CDC123/CAMK1D; cell division cycle 123 homolog	Regulates cell cycle progression and apoptosis
ADAMTS9; a disintegrin-like and metalloprotease (reprolysin type)with thrombospondin type 1 motif, 9	Proteoglycan cleavage and inhibition of angiogenesis
THADA; thyroid adenoma associated	Regulates cell cycle progression and apoptosis
TSPAN8/LGR5; tetraspanin 8	Signal transduction, cell development, activation, growth and motility
JAZF1; juxtaposed with another zinc finger protein 1	Transcription repressor
MTNR1B; melatonin receptor 1B	G protein-coupled receptor in retina and brain

7.2.2 The Currently Available Predictive Tests Do Not Include Genetic Markers

Because all the risk-increasing alleles that have been identified have such low penetrance, tests based on only one or a few of these variants will have limited clinical utility. Some statistical models predict that several hundred markers with allele frequencies and allelic odds ratios similar to those associated with these markers would be needed to produce a test with an acceptable level of clinical utility. Two recent studies, the EPIC-Potsdam study from the University of Munchen, Freisberg, Germany and a study by the Tethys Bioscience company, reported that testing for these gene variants did not improve the ability to predict the individual's risk for T2D above that afforded by standard biomarkers. The EPIC-Potsdam study reported that the clinical utility of a test that included measures such as plasma levels of glucose, hemoglobin A1C HDL cholesterol and triglycerides, and the activities of enzymes such as gamma-glutamyltransferase and alanine aminotransferase was not improved by the addition of up to 20 T2D-associated SNPs. Similarly, the Tethys Bioscience study found that genetic tests did not improve the predictive power of a test that was based on six biomarkers: adiponectin, C-reactive protein, ferritin, interleukin-2 receptor A, glucose and insulin.

7.2.3 The Pharmacogenomics of Type 2 Diabetes

The sulfonylureas and metformin are commonly prescribed for patients with T2D. Sulfonureas bind to a hyperpolarizing K+ channel on the beta cells of the pancreas; the resulting disinhibition of the beta cells' activity increases the release of insulin. Metformin, on the other hand, reduces gluconeogenesis by inhibiting the uptake of lactate by the liver. Metformin is often more effective than sulfonureas in patients who are significantly overweight or obese.

The level of activity in the CYP450 enzymes and drug transporter proteins influence the blood levels achieved by patients with type 2 diabetes (T2D) who have been prescribed antihyperglycemic drugs. For example, patients who possess the null *CYP2C9*3* allele achieve significantly higher blood levels of sulfonureas than patients with other *CYP2C9* alleles, and have an increased risk for hypoglycemia when given sulfonureas. In addition, patients who have low-activity isoforms of OCT1 have a limited ability to take metformin up into the liver, and because of this achieve a smaller reduction in plasma glucose levels after metformin than do patients with typical-activity isoforms of OCT1. In contrast, patients who possess low-activity alleles of *OCT*2, which is primarily responsible for the renal clearance of metformin, achieve higher blood metformin levels than patients with other *OCT2* alleles do.

The patient's status for *OCT* gene polymorphisms also influences metformin's interactions with other drugs. Metformin interacts with several other prescription drugs, including cimetidine, the histamine H2 receptor antagonist that is commonly

used in the treatment of heartburn and peptic ulcers. Cimetidine reduces renal clearance of metformin, and one study has shown that the reduction of metformin clearance by cimetidine is significantly lower in patients with the TT genotype for the G808T polymorphism in the *OCT2* gene.

Many drugs act as enzyme inhibitors, and different isoforms of the enzyme may be more or less easily inhibited by certain drugs. For example, angiotensin converting enzyme (ACE) inhibitors are used to reduce proteinuria in patients with diabetic and non-diabetic nephropathies. The ACE gene has a 287 bp insertion/deletion (I/D) polymorphism in intron 16; the insertion allele produces a low-activity isoform of the protein that appears to be more sensitive to ACE inhibitors than the isoform that the deletion allele produces is. Patients who have the DD genotype for the ACE I/D polymorphism have an increased risk for diabetes and diabetic nephropathy. In addition, they are less sensitive to the antiproteinuric effects of ACE inhibitors (caproptil, enalapril or atenolol) than patients who have the ID genotype, who are less sensitive than patients who have the II genotype. This is true in both patients with diabetic and non-diabetic nephropathy.

Thiazolidinedione (TZD) drugs and incretin enhancers/mimetics lower blood levels of glucose and HbA1c effectively, but the incretin drugs' efficacies are highly variable, and many patients experience ADRs such as weight changes, edema, severe vomiting, acute pancreatitis and potentially fatal heart failure after taking TZDs such as exenatide, sitagliptin, pioglitazone or rosiglitazone. The FDA has required a black box warning on the labels of pioglitazone and rosiglitazone since August 2007, because of the risk for potentially fatal heart failure. There are no tests yet available that are designed to help predict the patient's response to TZD drugs, but at least one company is developing a test that is intended to predict the efficacy of TZD drugs, as well as the patient's risk for ADRs.

7.3 Personalizing the Approach to Psychiatric Disorders

7.3.1 Many Original Findings Must Be Confirmed

A considerable amount of effort has been expended in an effort to develop tests that can personalize the approach to psychiatric disorders, but the psychiatric genetics literature is marked by a particularly high level of discrepancy between the results of different studies. Early studies reported that there were several gene polymorphisms that influenced one's risk for one or more psychiatric disorders, or one's response to one or more psychiatric drugs. Several studies also suggested that there is an interaction between genetic status and early life events in determining one's risk for depression or antisocial behavior. Unfortunately, however, most of these initially promising findings have not withstood the test of time. Very few of the promising initial research findings are likely to be translated into clinically useful tests anytime soon.

In addition to the limitations to which all personalized medicine tests are subject (see Sects. 3.3 and 3.6), the field of psychiatric genetics is often further complicated by the fact that patients with disorders such as depression or anxiety

are a heterogeneous group. Within any given disease population, different subsets of patients are likely to have different molecular pathologies, which reflect different genetic causes. Researchers must study each subgroup of patients independently if they are to find the gene variants that influence one's risk for that disease. There is little agreement, however, over the best way to classify patients with a given disorder in their relevant subgroups, or *endophenotypes*.

Another interesting finding that must be examined further is the suggestion that the influence of genetic factors on one's personality, risk for drug abuse, and risk for psychiatric disorders may change as the individual ages. Studies have reported age-related changes in associations between specific gene variants and cognitive behavior, personality factors, alcohol consumption, fears and phobias, and anxiety and depression.

In order to more confidently establish the field's foundation, meta-analyses that review multiple studies, and therefore include large populations of patients, are being conducted to determine which of the early findings are reliable enough to form the basis for further research. Large multicenter studies are also being conducted, which will avoid the problems previous studies have encountered due to their small sample sizes. At this point in time, most authorities are skeptical about the results of any study in this field, except for those relatively few findings that have been subsequently confirmed by a large meta-analysis or multicenter study.

7.3.2 Testing Genetic Variants That Influence Pharmacokinetics Is Helpful in the Treatment of Some Psychiatric Disorders

There are a few tests, such as the CYP450 test described in Chap. 4 (see Sect. 4.6), that can identify individuals who have extremely low or extremely high rates of metabolism of a number of drugs prescribed for psychiatric disorders. Tests such as these, that focus on the pharmacokinetics of the drug, can help reduce the incidence of adverse drug reactions, and reduce the number of patients who fail to respond to their drug because they were given too low a dose. The CYP450 test is clinically useful for guiding dosing strategies in patients who have either the poor metabolizer or ultrarapid metabolizer phenotypes. Unfortunately, however, it is not very useful for other patients.

7.3.3 Some Associations with Attention Deficit-Hyperactivity Disorder (ADHD) Have Been Confirmed

Attention deficit-hyperactivity disorder (ADHD) is one of the most common psychiatric disorders seen in children and adolescents, affecting approximately 5% of school-aged children throughout the world. Polymorphisms in several neurotransmitter-related genes have been associated with ADHD, and a few of the original findings have been upheld by subsequent meta-analyses. In addition, some research has suggested that

paternal transmission of alleles from several of these genes may convey greater risk than maternal transmission. The reason for this bias in parental origin is not known at present, but it may reflect the role that imprinting or other epigenetic factors (see Sect. 1.7) play in regulating the activity of these genes.

Recent meta-analyses suggest that there is a reliable association between ADHD and polymorphisms in several genes whose proteins influence neurotransmission. There is strongest support for the association between ADHD and polymorphisms in the genes that encode the type 4 DA receptor (*DRD4*), the type 5 DA receptor (*DRD5*), the DA transporter (*DAT1*) and the 5HT transporter (*5HTT*). In addition, a reasonable body of evidence suggests that polymorphisms in the genes that encode dopamine beta-hydroxylase (*DBH)*, the type 1B 5HT receptor (HTR1B) and the synaptosomal-associated protein 25 (*SNAP-25*) influence one's risk for ADHD.

Serotonergic and catecholaminergic neural pathways [pathways that use serotonin (5HT) or the catecholamines dopamine (DA) and norepinephrine (NE) as their neurotransmitters] are known to be involved in the regulation of mood and cognition. In addition, the SNAP-25 protein regulates the fusion of synaptic vesicles with the nerve cell membrane, so polymorphisms in any of these genes could plausibly influence the risk for ADHD. In addition, patients with ADHD are often prescribed the DAT1 inhibitor methylphenidate (MPH). MPH increases the concentration of DA at both its postsynaptic receptors and its presynaptic autoreceptors. The findings from these meta-analyses should provide the foundation for translational research studies that will result in clinically useful tests.

7.3.4 *Polymorphisms in Genes That Influence Serotonergic and Dopaminergic Function May Influence Personality*

There have been a number of gene polymorphisms reported to influence personality type, many in genes whose proteins influence neurotransmission. Several of these findings have been confirmed by recent meta-analyses. For example, it has been suggested that variants in the *5HTT* gene, as well as the gene that encodes the brain-derived neurotrophic factor (*BDNF*) influence one's scores on the Neuroticism subscale of the NEO-Personality Inventory, which reflects the tendency to feel negative emotions and interpret situations as threatening. In addition, recent meta-analyses have confirmed an association between two functional polymorphisms in the *DRD4* gene and novelty seeking and impulsiveness.

7.3.5 *Copy Number Variation May Influence the Risk for Schizophrenia*

It is estimated that genetic factors account for at least 80% of the heritability of schizophrenia, but there are as yet no well-established associations between gene sequence variants and schizophrenia. GWA studies have suggested that the genetic

risk factors for schizophrenia probably include thousands of potential susceptibility genes, with thousands of low-penetrance gene variants capable of contributing to the individual's risk.

Another interesting finding that has emerged from recent GWA studies is that copy number variations (CNVs, see Sect. 3.12) may be more important contributors to the risk for schizophrenia than previously thought. If this is confirmed, microarray or other assays might someday be used to detect CNVs that increase the risk for schizophrenia. Recent studies have reported the presence of CNVs up to 2 Mb in length in a surprising number of patients with schizophrenia. Unfortunately, however, the CNVs seem to be present in different locations in different patients, and there does not seem to be an association between schizophrenia and CNV at any particular locus.

7.3.6 Several Polymorphisms May Influence the Risk for Schizophrenia and the Response to Antipsychotic Drugs

Dopaminergic neurotransmission strongly influences prefrontal lobe-mediated behaviors, as does activity in the ascending 5HT pathway that makes extensive connections throughout the forebrain. In addition, all antipsychotic drugs block the D2 DA receptor (DRD2), and the newer antipsychotic drugs also block the DRD3 and DRD4 receptors, as well as the 5HT2A receptor. It is therefore no surprise that several studies have reported that variants in DA-related and 5HT-related genes influence one's susceptibility to schizophrenia and response to antipsychotic drugs.

In addition, several potentially serious ADRs may emerge during treatment with antipsychotic drugs. The two most important are the metabolic syndrome (weight gain, hypertriglyceridemia and hypercholesterolemia) and clozapine-induced agranulocytosis. Because these ADRs can have such serious consequences, it is particularly important to confirm the initial findings regarding gene variants that can influence the risk for these ADRs. Table 7.2 lists several promising initial findings still awaiting confirmation by meta-analyses.

7.3.7 Several Polymorphisms Have Been Associated with Depression and Bipolar Disorder (BPD)

There have been a number of gene polymorphisms reported to influence the risk for depression or suicide and the response to antidepressant treatments, including both drugs and sleep deprivation (Table 7.3). All these findings must be confirmed, however, before these findings can be translated into clinically useful tests.

It is no surprise to see that genes whose proteins influence serotonergic or catecholaminergic neurotransmission influence the risk for depression/suicide. Serotonergic and catecholaminergic systems are well known to regulate mood and

Table 7.2 Gene polymorphisms that have been reported to influence risk for schizophrenia and response to antipsychotic drugs

Gene	Function of protein	Associated with
DRD2, type 2 DA receptor	DA receptor	Response to antipsychotics; susceptibility to post-traumatic stress disorder
DRD3, type 3 DA receptor	DA receptor	Risk for tardive dyskinesia after neuroleptic drugs
DRD4, type 4 DA receptor	DA receptor	Risk for psychotic symptoms
COMT, catechol-O-methyl transferase	Catabolizes catecholamine neurotransmitters	Risk for schizophrenia; response to antipsychotic drugs, incl. activation of the prefrontal lobe
5HTT, 5HT transporter	5HT reuptake protein	Risk for schizophrenia
TPH, tryptophan hydroxylase	Catalyzes the rate-limiting step in 5HT synthesis	Response to neuroleptic drugs
HTR2A, 5HT receptor 2A	5HT receptor	Response to neuroleptic drugs, incl. risk for metabolic syndrome
HTR2C, 5HT receptor 2 C	5HT receptor	Response to neuroleptic drugs, incl. metabolic syndrome
GNβ3, G beta 3 subunit of the G-protein complex	Critical for many neural signal transduction pathways	Response to neuroleptic drugs, incl. risk for metabolic syndrome
MAOA, monoamine oxidase A	Catabolizes 5HT and CA neurotransmitters	Risk for schizophrenia in men
GRM3, glutamate receptor	Glutamate receptor	Risk for schizophrenia; verbal list learning and verbal fluency; level of excitation in the dorsolateral prefrontal lobe; level of N-acetylaspartate (NAA) and glial glutamate transporter EEAT2 in the prefrontal lobe
DAAO, D-amino acid decarboxylase and *G27*, DAAO activator	Oxidizes D-serine, reduces activation of the NMDA glutamate receptor	Risk for schizophrenia
DRP-2, aka *DPYSL-2* or *CRMP-2,* dihydropyrimidinase-related protein 2, aka collapsin response mediator protein 2	Axon growth and connections	Risk for paranoid schizophrenia
ZNF804A, zinc finger protein 804A	Transcription factor, targets unknown	Coordination between the hippocampus and dorsolateral prefrontal cortex
RELN, reelin	Cerebral cortex growth and organization	Risk for schizophrenia in females
NRG1, neuregulin	Growth and differentiation of neurons and glia	

(continued)

Table 7.2 (continued)

Gene	Function of protein	Associated with
DTNBP1, dystrobrevin-binding protein 1, aka dysbindin	Integral protein in neural membranes	Risk for schizophrenia
CACNA1C, L-type voltage gated calcium channel	Neuronal calcium channel	Risk for schizophrenia
ADRB3, beta-3 type adrenergic receptor	NE receptor	Olanzapine-induced metabolic syndrome
ACCα, acetyl-CoA carboxylase	Catalyzes the rate-limiting step in fatty acid synthesis	Hypertriglyceridemia after olanzepine, quetiapine or chlorpromazine
TGF β1, transforming growth factor β1	Transcription factor	Hypertriglyceridemia after olanzepine, quetiapine or chlorpromazine
PECAM-1, platelet/endothelial cell adhesion molecule 1	Blood clotting	Hypertriglyceridemia after olanzepine, quetiapine or chlorpromazine
NPY, neuropeptide Y	Neurotransmitter	Hypertriglyceridemia after olanzepine, quetiapine or chlorpromazine
API, angiotensinogen proteinase inhibitor	Inhibits angiotensin formation	Hypertriglyceridemia after olanzepine, quetiapine or chlorpromazine
ABCB1	Multidrug transporter	Risk for clozapine-induced agranulocytosis
HLA-B38, DR4 and DQw3	Human leukocyte antigens	Risk for clozapine-induced agranulocytosis in Ashkenazi Jews
HLA-Cw-B, HLA-DRB5-DRB4 and HLA-Cw-B-DRB5	Human leukocyte antigens	Risk for clozapine-induced agranulocytosis in non-Jewish caucasians

cognition. In addition, patients with depression are often prescribed selective 5HT reuptake inhibitors (SSRIs) or selective NE reuptake inhibitors (SNRIs), which inhibit the proteins that take 5HT and NE back up into the presynaptic nerve terminals that released them. Almost all the genes listed in Table 7.2 encode proteins that regulate neurotransmission. If these findings are confirmed, they will allow researchers to add several pharmacodynamically relevant gene polymorphisms to the pharmacokinetically relevant ones, to construct tests with true clinical utility.

7.3.8 Polymorphisms in the Serotonin Transporter (5HTT) Gene May Not Influence One's Risk for Depression After All

One of the important mechanisms by which SSRIs exert their antidepressant effects involves a reduction in the density and/or responsiveness of the presynaptic 5HT autoreceptors that occurs in response to the excessive 5HT concentration present at

Table 7.3 Polymorphisms that have been reported to influence risk for depression/suicide and response to antidepressant treatments

Gene	Function of protein	Associated with
5HTT, serotonin transporter	Takes 5HT back up into presynaptic neuron	Altered response to fluvoxamine, paroxetine, citalopram, nortryptiline and sleep deprivation; risk for suicide
MAOA, monoamine oxidase A	Catabolizes 5HT and CA neurotransmitters	Severity of symptoms; response to fluvoxamine
TPH, tryptophan hydroxylase	Catalyzes the rate-limiting step in 5HT synthesis	Risk for depression; response to paroxetine, citalopram, fluoxetine and fluvoxamine
HTR1A, 5HT receptor 1A	5HT receptor	Response to antidepressant drugs
HTR2A, 5HT receptor 2A	5HT receptor	Risk for suicide
HTR3A, 5HT receptor 3A	5HT receptor	Response to antidepressant drugs
HTR3B, 5HT receptor 3B	5HT receptor	Response to antidepressant drugs
DAT1, dopamine transporter	Takes DA back up into presynaptic neuron	Response to SSRIs, tricyclic antidepressants, mirtazapine and venlafaxine; risk for suicide
COMT, catecholamine-O-methyltransferase	Catabolizes both DA and NE	Risk for depression; response to several antidepressant drugs
NET1, norepinephrine transporter	Takes NE back up into presynaptic neuron	Risk for depression; severity of symptoms
GNβ3, G beta 3 subunit of the G-protein complex	Critical for many neural signal transduction pathways	Response to nortryptiline and fluoxetine
KCNK2, aka *TREK1*, potassium channel, subfamily K, member 2	Potassium ion channel	Regulates activity in the basal ganglia, dorsal anterior cingulate cortex, orbitofrontal cortex and mesial prefrontal cortex; Influences response to several antidepressant drugs
ANK3, ankyrin 3	Regulates the activity of sodium channels at the node of Ranvier	Risk for BPD
CACNA1C, alpha 1 C subunit of the L type voltage-dependent calcium channel	Regulates dendritic calcium influx	Risk for BPD
DGKH, diacylglycerol kinase eta	Participates in several neural signal transduction pathways	Risk for BPD
MYO5B, myosin5B	Regulates vesicular trafficking	Risk for BPD
BDNF, brain-derived neurotrophic factor	Regulates neuronal differentiation and development, Synaptic vesicle trafficking	Risk for BPD

(continued)

Table 7.3 (continued)

Gene	Function of protein	Associated with
DAAO, D-amino acid decarboxylase and *G27*, DAAO activator	Oxidizes D-serine, reduces activation of the NMDA glutamate receptor	Risk for BPD
CREB1, cAMP response element binding protein 1	Participates in several neural signal transduction pathways	Risk for suicide
GRIK2, glutamate kainate receptor subunit Glu6R	Receptor for the excitatory neurotransmitter glutamate	Risk for suicide
GRIA3, glutamate AMPA receptor subunit AMPA3	Receptor for the excitatory neurotransmitter glutamate	Risk for suicide
IL28RA, interleukin receptor	Class II cytokine receptor	Risk for suicide
PAI-1, aka *SERPINE1,* plasminogen activator inhibitor type 1	Enhances blood clotting	Risk for depression; response to fluoxetine and citalopram

the receptors. One of the very exciting early research findings in this field involved reports that a polymorphism in the *5HTT* gene influenced the individual's risk for depression, and in particular, influenced the likelihood that the individual would develop depression after experiencing adverse early life events. A recent meta-analysis has suggested, however, that there is significant heterogeneity in the results of the studies that have addressed this issue, and that there is no reliable association between *5HTTLPR* status and depression risk, nor does *5HTTLPR* status interact with the presence/absence of adverse early life events to influence the individual's risk for depression.

This area of research has been complicated by the recent discovery that the *5HTT* gene's sequence is more variable than previously thought. As discussed in Chap. 1 (see Sect. 1.6.1), the *5HTT* gene has a 44 bp insertion/deletion polymorphism in its promoter region (*5HTTLPR*) that influences the rate of transcription of the gene and the amount of the 5HTT protein the gene produces. The *5HTTLPR* polymorphism was long thought to have two alleles: the long (insertion) allele and the short (deletion) allele. However, the insertion allele has recently been shown to have a SNP in it that has functional consequences. The insertion allele with a G in this position [the L(G) allele] conveys greater activity on the gene than the deletion allele (commonly called the short, or S allele) or the insertion allele with an A in this position [the L(A) allele] does. As this is a recent discovery, most studies include all subjects who possess the L(A) or L(G) alleles in the L allele group. Future studies must reanalyze the *5HTTLPR* sequences in patients from the old studies, or recruit new subjects whose *5HTTLPR* status can be definitively ascertained.

7.3.9 Polymorphisms in the Monoamine Oxidase a (MAOA) Gene May Not Influence the Propensity toward Antisocial Behavior after All

Another exciting early finding that received widespread attention but has not been supported by subsequent studies is the report that a polymorphism in the *MAOA* gene may interact with the individual's history of childhood abuse to influence a male's risk for antisocial behavior. Monoamine oxidase catabolizes several monoamine neurotransmitters, including 5HT and NE, and severe mutations in the *MAOA* gene have been associated with violent and antisocial behavior in boys,[1] so it is quite plausible that *MAOA* status might influence the risk for antisocial behavior. Unfortunately, however, these early findings have not been supported by subsequent studies, and further studies are needed to definitively determine whether there is an association between *MAOA* status and behavior.

7.3.10 Anxiety Disorders

At this point in time, none of the initial findings in the study of the genetics of anxiety have been confirmed. Promising early findings suggested that variants in the *5HTT* gene influence the response of the amygdala to fearful faces, as well as the anxiolytic effects of SSRIs. In addition, variants in the genes that encode the 5HT type 2A receptor and the brain-specific isoform of TPH known as TPH2 have been reported to influence the risk for panic disorder. It is also interesting to note that early studies suggested that variants in the *5HTT* and *TPH2* genes influence the anxiolytic effects of placebos.

7.4 Personalizing the Treatment for Substance Abuse

7.4.1 Most Studies Have Focused on Catecholamine and Opiate Pathways

It is well known that different individuals achieve different levels of euphoria after the same dose of a drug, as well as different intensities of withdrawal after they discontinue drug use. The genetic variants that influence these factors influence the individual's susceptibility to drug abuse. The mesocorticolimbic catecholamine

[1] *MAOA* resides on the X chromosome. Because males only have one copy of their X chromosome genes, mutations in *MAOA* are expected to affect males more often and more severely than females.

systems (which use the neurotransmitters dopamine, DA and norepinephrine, NE) are important mediators of the euphoric and reinforcing effects of many drugs of abuse. Drugs such as cocaine, amphetamine and methamphetamine directly increase DAergic and NEergic transmission by increasing release of these neurotransmitters and/or inhibiting their reuptake by the DA and NE transporter proteins (*DAT1*, aka *SLC6A3*; *NET1*, aka *SLC6A2*, respectively). In addition, opiatergic systems interact with the mesocorticolimbic DAergic and NEergic systems that mediate reinforcement. It is logical to assume, therefore, that genetic variants that affect DAergic, NEergic and opiatergic transmission will affect the individual's susceptibility to drug abuse, as well as his/her response to drugs intended to combat drug abuse. Several studies have suggested that polymorphisms in several DAergic genes, several NEergic genes and the mu opioid receptor gene (*OPRM1*) are risk factors for drug abuse in general and/or polydrug abuse.

DAergic systems may also mediate some of the personality factors that are believed to predispose the individual to drug abuse. Several studies have reported associations between DAergic gene polymorphisms and personality traits. Most relevant to drug abuse is the fact that the same allele of the *DRD4* exon 3 VNTR polymorphism that increases the risk for ADHD and personality traits such as novelty seeking is also believed to increase the individual's risk for drug abuse.

7.4.2 Stimulants

Not surprisingly, polymorphisms in both the *NET1* and *DAT1* genes have been found to influence the individual's response to stimulants. Several studies using healthy volunteers have reported that the individual's status for several polymorphisms in the *NET1* gene influences the euphoria one achieves after an acute dose of amphetamine or methamphetamine. Similarly, individuals who have the homozygous 9/9 (low-activity) genotype for the 3' UTR VNTR polymorphism in the *DAT1* gene have been reported to have a diminished euphoric response to amphetamine. Another study has also reported that possessing the 9-repeat (or fewer) allele for this polymorphism is associated with a longer duration of psychosis after cessation of the drug.

It is also not surprising to find that DA receptor gene polymorphisms have been reported to influence the individual's response to stimulants. For example, the 7-repeat allele of the *DRD4* VNTR polymorphism has been associated with an increased risk for methamphetamine abuse, while two *DRD2* alleles (A1 and B1) have been associated with an increased risk for cocaine dependence.

Several lines of evidence suggest that excessive bombardment of DA receptors over a prolonged period of time leads to the anxiety, paranoia and psychosis that are often seen in chronic drug abusers, and the pharmacogenetic literature supports this claim. There is a polymorphism (C-1021 T) in the promoter region of the dopamine beta-hydroxylase (*DBH)* gene, which encodes the enzyme that converts dopamine to norepinephrine. The T allele produces an isoform of the enzyme with

considerably less activity than the C allele's isoform, thereby increasing the concentration of DA at its receptors. Several studies have reported that cocaine abusers who possess the TT genotype for this polymorphism, or a *DBH* haplotype that included the T allele of this polymorphism, were more likely to develop paranoia than those who possess the other genotypes and haplotypes. Because disulfiram inhibits the activity of DBH, the individual's status for this DBH polymorphism may also affect his/her risk for developing paranoia after disulfiram treatment.

The slower-acting M isoform of the enzyme catechol-O-methyl transferase (COMT) is believed to increase the concentration of DA at its receptors, and would therefore be expected to increase the risk for both acute ADRs and the ADRs that develop over chronic use, such as paranoia and psychosis. Possessing the M allele of this *COMT* polymorphism has indeed been reported to increase the individual's risk for developing psychosis after acute and chronic cocaine use. In addition, possessing the M allele for this *COMT* polymorphism increased the individual's risk for spontaneous relapse after cessation of drug use.

Given the critical role of DA receptors in stimulant abuse, it is no surprise to see that DA receptor polymorphisms have been reported to influence the course of psychosis in chronic stimulant abusers. Several *DRD2* polymorphisms, including a -141 C ins/del polymorphism in the promoter region, have been shown to influence the latency of stimulant-induced psychosis, as well as the risk for spontaneous relapse after remission.

Genomic researchers have just recently begun to discover some of the genetic polymorphisms that influence transmission downstream of the DA receptors, and thereby influence the individual's experience with drugs of abuse. For example, the protein encoded by the casein kinase 1 epsilon gene (*CSNK1E*) participates in the metabotropic signal transduction pathway that lies downstream from the DA receptor. There is a C>G SNP in the 3'UTR of the *CSNK1E* gene; the C allele is associated with a greater sensitivity to low doses of amphetamine.

Genes whose proteins affect 5HTergic transmission have not been studied as extensively as DA- and NE-related genes. Given the ubiquity of the S allele of the *5HTT* promoter polymorphism in neurobehavioral disorders, however, it is no surprise to see that the individual's status for the *5HTT* promoter polymorphism, and the *5HTT* intron 2 VNTR polymorphism, have also been reported to influence the subjective response to amphetamine.

Opiatergic systems interact with DAergic systems to mediate the reinforcing effects of drugs of abuse. Opiatergic transmission inhibits the activity of γ-aminobutyric acid (GABAergic) neurons, which inhibit activity in mesolimbic-mesocortical DA pathways.

Opiatergic neural pathways are also known to influence the individual's response to stimulants independent of their interactions with DAergic systems. The mu opioid receptor gene (*OPRM1*) has a putative functional polymorphism in it (A118G, Asp40Asn). Possessing the A allele for this polymorphism is associated with an increased risk for abuse of several drugs. It is not surprising, then, that several studies have reported that the individual's status for this polymorphism influences the risk for or latency of methamphetamine-induced psychosis.

Commonly abused stimulants such as amphetamine increase the release of brain-derived neurotrophic factor (BDNF) from neurons, and the MM genotype of the *BDNF* gene's V66M polymorphism has been reported to decrease BDNF secretion by active neurons. Consistent with this, healthy Caucasian volunteer subjects who had the MM genotype for the V66M *BDNF* polymorphism reported less arousal after an acute dose of amphetamine compared to individuals with the VV genotype. In a study using Taiwanese subjects, however, the M allele was associated with a greater risk for methamphetamine abuse, and heroin abuse as well. In addition, the MM genotype was associated with an earlier onset of substance abuse than the VV genotype was. It is possible that the different ethnicities of the subjects used in these studies accounts for the discrepancies in their results. The V66M polymorphism may be linked to the actual functional polymorphism, and the M and V alleles may be associated with different alleles of the functional polymorphism in Asian subjects versus Caucasians. Alternatively, the reduced release of BDNF may require the individual to use higher doses of the drug, contributing to an increased risk for withdrawal and addiction.

7.4.3 Alcohol

As would be expected, several studies have suggested that polymorphisms in DA-related genes influence alcohol's effects, including acute effects as well as addiction and withdrawal. Several studies have suggested that the A1 and B1 alleles of the *DRD2* gene increase the risk for alcoholism, especially severe alcoholism. In addition, the individual's status for a polymorphism in the gene that encodes the tetratricopeptide repeat domain 12 (TTC12) protein has also been associated with an increased risk for alcoholism. TTC12 interacts with DA in the WNT signaling pathway, and this interaction influences the development of DAergic neurons in the ventral midbrain.

DAergic systems may mediate several of the personality factors that are believed to predispose the individual to alcoholism. For example, one study has reported an association between the individual's status for both *DRD4* and *DRD2* polymorphisms and his/her score on the Harm Avoidance scale (which loads heavily on worry, pessimism and shyness) of the Tridimensional Personality Questionnaire. The same authors then corroborated this initial finding using measures of negative affect, stress reaction, well-being and alienation from the Multidimensional Personality Questionnaire.

Recent research into the factors that influence alcoholism has expanded to include other neurotransmitters and neuromodulators that interact with DA systems. For example, the endogenous opiate neurotransmitter systems interact with the catecholamines, especially DA, to mediate the reinforcing effects of abused drugs. Opiatergic systems have long been implicated in the reinforcing effects of alcohol, as well as the mechanism for alcohol addiction and withdrawal.

One important mechanism whereby the subjective effects of alcohol are mediated involves the release of beta-endorphins in the brain. Naltrexone (NTX) reduces the

euphoric and reinforcing effects of alcohol, by causing the release of beta-endorphins, and blocking further release by alcohol. The G allele of the A118G SNP in *OPRM1*, which conveys lower risk for stimulant abuse than the A allele, has also been associated with an increased ability of NTX to reduce the positive subjective effects of alcohol, as well as a reduced relapse rate after NTX treatment. Interestingly, another study has reported that American Indians possessing minor alleles of several SNPs in the mu opioid receptor gene reported increased intensity of alcohol's effects, especially adverse effects, which may have caused them to be less prone to alcoholism than those with the more common alleles. This may indicate an interaction between the variants of the alcohol-metabolizing enzymes and variants in the neurotransmitter systems that mediate the effects of alcohol.

Polymorphisms in the *BDNF* gene appear to influence many aspects of brain function. The *BDNF* gene has a G196A polymorphism in it. A significantly increased frequency of the A allele has been reported in alcoholics with violent tendencies, as well as those who had a history of delirium tremens. In addition, the A allele was associated with earlier onset of alcoholism.

7.4.4 Opiates

The individual's status for the CYP450 genes (see Sect. 4.6) influences the individual's risk for opiate dependence, as well as his/her response to drugs designed to treat opiate addiction. Although having a poor metabolizer (PM) status for the CYP450 genes is often associated with an increased risk for drug ADRs, CYP2D6 PMs have a reduced risk for opiate dependence. This is probably due to the fact that many of the metabolites of opiate drugs are active compounds, and reduced metabolism of the parent compound reduces the accumulation of active compounds at their receptors. In addition, CYP2D6 ultrarapid metabolizers (UM) show a reduced response to methadone treatment, and experience more frequent withdrawal symptoms, than patients with the other CYP2D6 phenotypes. Buprenorphine is a better choice for treating patients with the CYP2D6 UM phenotype, because buprenorphine is not extensively metabolized by CYP2D6.

Because many of the opiates' effects are mediated through the drugs' interactions with the mu opioid receptor (OPRM1), it is no surprise to see that the individual's status for the asp40asn functional polymorphism in the *OPRM1* gene has been reported to influence the response to several opiates, including morphine, morphine-6-glucuronide, alfentanil and levomethdone. In several studies, which together included healthy volunteers, cancer patients and postoperative patients, subjects with the homozygous GG genotype of the A118G polymorphism in the *OPRM1* gene showed reduced sensitivity to the analgesic effects of morphine.

DA receptor gene polymorphisms have also been reported to influence the response to opiates. For example, the A1 allele of the DRD2 TaqA polymorphism has been associated with an increased risk for heroin addiction, greater dose consumption and poorer response to methadone treatment. In addition, the 7-repeat

allele of the VNTR polymorphism in the *DRD4* gene, which increases the risk for ADHD and personality traits such as novelty seeking, has been associated with stronger cue-elicited heroin craving, which is believed to predispose the addict to relapse when he/she returns to the environment in which he/she previously abused drugs.

7.4.5 Nicotine

Nicotine is one of the most commonly abused drugs, and also one of the most difficult addictions to treat. As DAergic systems mediate the reinforcing effects of many abused drugs, it is no surprise to see that polymorphisms in DA-related genes have been reported to influence smoking behavior, addiction and relapse. Among the usual candidates for study are the genes that encode the DA receptors and the DA transporter DAT1. Not surprisingly, subjects who possessed certain alleles of polymorphisms in the *DRD2*, *DAT1* and *DRD4* genes have been reported to smoke more due to negative mood than smokers with the other genotypes. In addition, the glycine allele of the S9G polymorphism in the *DRD3* DA receptor gene has been associated with heavier smoking.

The gene that encodes the type 1 ankyrin repeats and kinase domain containing protein (ANKK1) lies very close to the *DRD2* gene. In addition, it has a functional SNP whose minor allele causes the substitution of histidine for arginine in the protein's C-terminal ankyrin repeat domain. This variant reduces the expression of genes that are activated by the transcription factor NF-kappaB (NFKB). Because *DRD2* is activated by NFKB, this *ANKK1* gene variant may indirectly affect DRD2 density in the brain.

It is also not unexpected to see that polymorphisms that influence DAergic transmission influence the individual's ability to stop smoking. Subjects with the 7-repeat or longer alleles for the *DRD4* VNTR polymorphism have been reported to suffer more intense cravings, anger/irritability and sleep disturbances after cessation than individuals with the shorter alleles. In addition, they exhibited greater craving, less positive affect and more arousal and attention to environmental cues that are associated with smoking than individuals with the shorter alleles. In addition, one study has reported that the alleles the individual possesses for the *DRD2*, *DRD3* and *DRD4* polymorphisms interact to influence other withdrawal symptoms, including nervousness, sleep disturbances and trouble concentrating. Like so many other interesting findings in this field, these findings must be confirmed by meta-analyses or multi-center studies with populations that are large enough to produce reliable results before they can be translated into clinically useful tests.

The individual's status for these DAergic polymorphisms may determine the optimal strategy for treating the individual's addiction. For example, smokers who possess the Ins/Ins genotype of the *DRD2* promoter polymorphism have been reported to achieve better success quitting with buproprion treatment than

smokers with the other genotypes. In contrast, smokers who possessed the Del allele achieved better responses with nicotine replacement therapy than did smokers who possessed the Ins/Ins genotype. One study has suggested that the individual's status for polymorphisms in the *DRD2* gene interacts with his/her status for polymorphisms in the *DAT1* gene to influence the individual's ability to stop smoking and the best choice of treatment. If subsequent studies confirm this finding, these tests may help guide the choice of treatment for people who are trying to quit smoking.

Because opiate neurotransmitters interact with forebrain DA systems to mediate the reinforcing effects of drugs, opiate system-related genes are good candidates for pharmacogenomic investigations into nicotine abuse and addiction as well. Once again, the individual's status for the functional asn40asp functional polymorphism in *OPRM1* has been reported to influence the reinforcing effects of smoking, the individual's ability to quit smoking and also the degree of mood disturbance and weight gain experienced after cessation of smoking. In addition, it was reported that smokers who possess the asp allele of this polymorphism are more likely than others to benefit from transdermal nicotine therapy.

7.5 Alzheimer Disease and Cognitive Decline in Aging

Alzheimer disease (AD) is the most common cause of dementia, and results from a complex interaction between genetic and nongenetic factors. Both early onset (before age 65) and late onset forms have been recognized. Early onset AD, also referred to as familial AD, accounts for ~10% of cases. Over the last few decades, rare, highly penetrant mutations in the genes that encode the amyloid precursor protein (*APP*), presenilin 1 (*PSEN1*) or presenilin 2 (*PSEN2)* genes have been reported to cause early-onset AD. Inheritance for early onset AD is autosomal dominant, so recurrence risk is 50% for future offspring. In contrast, most cases of AD have a later onset (≥65 years of age), are non-familial, and likely result from the interaction between highly prevalent genetic variants with low penetrance and nongenetic factors. Age is the greatest risk factor for both sexes, although women are more likely to develop late onset AD, primarily because they live longer.

As AD is common, and the general lifetime population risk for dementia is approximately 10–12%, we utilize empiric data to estimate risk for unaffected individuals from families that do not exhibit an autosomal dominant pattern of inheritance. In these families, first-degree relatives (parents, siblings, and children) of a person with AD have a cumulative lifetime risk of developing AD of 20–25%, or 2.5 times that of the typical person. Having several additional affected family members probably increases the risk to close relatives, but the magnitude of this risk remains unclear, unless of course the pattern in the family history is characteristic of autosomal dominant inheritance. This again illustrates the importance of taking, and updating, the patient's family history.

7.5.1 The ApoE Gene Is a Known Risk Factor

The heritability of late-onset AD has been suggested to be as high as 80%. The search for its genetic underpinnings began in the 1980s, and in 1991 significant linkage was demonstrated on chromosome 19q11–13. Unfortunately, this yielded such a vast genomic region (several million nucleotides, hundreds of genes) that it did not provide much focus for subsequent candidate gene studies. Subsequent studies using ultra-high-density SNP genotyping identified the apolipoprotein E (*ApoE*) gene as a major risk factor.

To date, the ApoE4 allele of the *ApoE* gene is the only genetic variant known to increase the individual's risk for late-onset AD. ApoE status is technically defined by the individual's status for two SNPs, rs429358 and rs7412. ApoE has three well characterized isoforms: e3 is by far the most common (78% in Caucasians), with e2 (6%) and e4 (16%) the variant alleles. Of the three isoforms, e2 has the lowest affinity for the apoE receptor, and e4 the highest. The e4, or ApoE4, allele is the risk-increasing allele; it contains the C allele of the rs429358 SNP and the C allele of the rs7412 SNP.

The specific level of risk associated with possessing the ApoE4 allele depends on the individual's age, sex, ethnicity and other nongenetic factors, but possessing the ApoE4 allele obviously increases risk. Individuals who inherit one copy of the e4 allele have a two to threefold greater risk of the disease, whereas ApoE4 homozygotes have an approximately 15-fold greater risk than those who do not possess an ApoE4 allele. In addition, some research suggests that these risks increase markedly (12x for late-onset Alzheimer's and 61x for early-onset disease) if the individual is homozygous for the C allele of the rs429358 SNP.

7.5.2 Other Gene Variants That Influence the Risk for AD and Cognitive Decline in the Elderly

One of the most noticeable aspects of age-related cognitive decline involves an impairment in working memory and executive function, which are mediated in part by DAergic systems in the prefrontal lobe. Because COMT plays a major role in catabolizing DA in the prefrontal cortex, the *COMT* gene is a frequent subject of study when investigating the genetics of prefrontal lobe-mediated behaviors.

Several studies suggest that the relationship between the strength of DAergic transmission in the prefrontal cortex and working memory performance is described by an "inverted U" shaped function. The V isoform (highest COMT activity, weakest DAergic transmission) has been reported to be associated with decreased performance on prefrontal lobe-dependent behaviors (ex. letter-number sequencing task, attention and target detection task) in several patients groups, and in healthy controls (including children) as well. In addition, several studies

have reported a graded effect, whereby individuals with the VV genotype had the lowest scores on a working memory/executive function task, individuals with the VM genotype had intermediate scores and individuals with the MM genotype had the highest scores. It appears as if having the homozygous genotype for the low-activity, DA-enhancing allele puts the individual (in an unmedicated state) at the top of the inverted U.

Increasing DA transmission pharmacologically interacts with *COMT* genotype, and increasing DA transmission above the level seen in MM homozygotes appears to place the individual on the descending limb of the inverted U. One study has reported that amphetamine, which increases DAergic transmission, increased performance on a working memory/executive function task (Wisconsin Card Sorting test) in individuals with the VV genotype, but either did not improve performance or produced a decrease in performance in individuals with the MM genotype, depending on the degree of difficulty of the task. Another study reported that a similar pattern of results was seen after administration of the CNS-penetrant COMT inhibitor tolcapone. It appears from these studies that possessing the MM genotype for this *COMT* polymorphism endows the individual with a near-optimal level of DAergic transmission in the prefrontal lobe, and increasing DAergic transmission pharmacologically can remedy a natural deficiency, but lead to a decrease in performance in those who already have the optimal MM genotype.

Both fMRI and event related potential (ERP) data suggest that the subjects' performances on these working memory/executive function tasks reflect the level of efficiency in prefrontal lobe DAergic systems. As mentioned above, the combined effects of *COMT* genotype and amphetamine on task performance and fMRI data suggest that the strength of DAergic transmission in the prefrontal cortex influences performance on these tasks and efficiency of transmission in the prefrontal lobe. This suggestion has been corroborated by one study in which event related P300 potentials were measured in an "oddball" detection task. Individuals possessing the M allele of the V158M *COMT* polymorphism demonstrated a lower amplitude P300 potential, suggesting that there was less "noise" (increased efficiency of processing) in the prefrontal lobe in these individuals.

Although the V158M polymorphism has received most of the attention, a few studies have examined other putative functional polymorphisms in the *COMT* gene. For example, possessing the minor allele of one of the other SNPs in the 5' region of the COMT gene was reported to reduce performance on a verbal memory task (California Verbal Learning Test) in both healthy controls and patients with BPD.

One study has reported both cognitive testing data and fMRI data that demonstrated that patients with schizophrenia and schizophreniform disorder who possessed the MM genotype for the COMT V158M polymorphism demonstrated a greater improvement in cognitive skills and neural processing efficiency after 8 weeks of treatment with olanzepine. In addition, the CNS-penetrant COMT inhibitor tolcapone is being investigated as a potential cognitive enhancer in both healthy elderly individuals and several patient groups. Determining the individual's status

for functional polymorphisms in the COMT gene may help clarify an individual's risk for cognitive decline due to aging or one of several diseases, as well as his/her expected response to cognitive enhancers such as tolcapone.

As mentioned above, BDNF is secreted by active neurons, and plays an important role in neuronal differentiation and the trafficking of synaptic vesicles. BDNF also helps mediate the synaptic plasticity that underlies learning and memory. The BDNF protein is first synthesized as the inactive proBDNF, which is then enzymatically cleaved to create BDNF. BDNF level appears to an indicator of neuronal integrity and activity in general, and the proBDNF protein has been reported to be present in reduced concentration in the parietal cortex of patients with AD.

In healthy control subjects, *BDNF* status has been reported to influence cognition and neural function. In two studies, the low-activity MM genotype of the V66M *BDNF* polymorphism has been reported to be associated with fMRI evidence suggesting abnormal function in the hippocampus, as well as impaired episodic memory or declarative memory in healthy elderly subjects. The M allele has also been associated with reduced levels of N-acetylaspartate in the hippocampus, which is believed to indicate reduced integrity and synaptic activity of hippocampal neurons.

In contrast, however, in studies involving patients with AD, the V allele of the V66M *BDNF* polymorphism was present at a significantly higher frequency in patients with AD than in controls. At present there is no explanation for the conflicting results in healthy controls and patients with AD. Two other studies, however, have also reported that the T allele of the *BDNF* C270T polymorphism was present at a greater frequency in patients with AD than controls.

Some of the associations that have been reported are specific to diseases, and may provide especially valuable insights into the physiology of that disease. For example, the C13R polymorphism in the lymphotoxin (*LT*) gene has been reported to influence cognitive function in schizophrenic patients, but not in healthy controls or patients with bipolar disorder (BPD).

Genes whose proteins participate in immune responses may also influence the risk for AD and cognitive decline in general. Possession of the human leukocyte antigen DRB1 (HLA-DRB1) increases the risk for cognitive decline in both AD patients and healthy elderly individuals. Similarly, HLA-DR5 has been associated with reduced performance on delayed recall tasks. In contrast, HLA-DR1 has been associated with superior performance on recall tasks.

The functions of the cathepsin D (CTSD) protein include processing antigens for presentation by HLAs. The T allele of a functional SNP in exon 2 of the *CTSD* gene has been associated with reduced performance on tests assessing information processing speed, spatial recall and fluid intelligence. In addition, individuals who possess the T allele of this *CTSD* polymorphism and the ApoE4 allele scored lowest of all on tests of fluid intelligence, information processing speed and immediate recall. It is known that HLA-DR2 binds more readily than other HLAs to myelin basic protein (MBP), and by doing so prevents CTSD from cleaving MBP. Accordingly, individuals who possess HLA-DR2 and the T allele of *CSTD* performed especially poorly in memory tests.

7.6 Asthma and Other Respiratory Disorders

Asthma affects an estimated 17 million Americans (6.4% of the U.S. population), and 300 million people across the world. It is responsible for more lost time in school and childhood hospitalizations than any other disease. The societal cost to the US alone in 2007 was estimated to be $20 billion, with medication costs the single largest expenditure. One of the difficulties inherent in treating asthma is the fact that different patients with asthma will have different ages of onset, sensitivity to environmental triggers, course of the disease and response to medication.

Beta-2 adrenergic receptor agonists are vasodilators, and are frequently prescribed for patients with asthma. The beta-2 adrenergic receptor (*ADRB2*) gene has several functional polymorphisms that affect the pharmacodynamics of beta-2 adrenergic receptors agonists. There are over a dozen different SNPs in the *ADRB2* sequence, but two are of particular interest, because each influences a different aspect of the response to beta-2 adrenergic agonists. A glutamine/glutamic acid polymorphism in codon 27 influences the degree of vasodilation seen after beta-2 agonist (isoproterenol) administration, while an arginine/glycine polymorphism in codon 16 influences the degree to which the individual's system adapts to repeated beta-2 agonist (albuterol) exposure.

There is considerable disagreement between studies that have assessed the effects of the codon 16 polymorphism on response to beta-agonist drugs. The duration of action of the drug may be an important variable, as may the race of the patients being examined. For example, in one study, the morning (pre-medication) forced expiratory volume (FEV) declined gradually in those patients who were homozygous for the codon 16 arginine allele after repeated administrations of inhaled short-acting beta-agonists, while it remained steady in patients who were homozygous for the glycine allele. Further, the morning FEV deteriorated significantly after cessation of treatment in patients who were homozygous for the arginine allele, but not in patients who were homozygous for the glycine allele. In contrast, one large (~15,000 patients) study found that African-American patients who were taking beta-agonists (the study did not report which) and had the arg/arg genotype for *ADRB2* had better lung function, possibly because they had less severe respiratory disease. This *ADRB2* polymorphism is an intriguing candidate for a personalized medicine test, but further studies will be needed before we know exactly how it influences the response to beta-agonists, and in whom.

Another thing that makes these particular functional polymorphisms so interesting is the fact that the variant alleles are relatively common. The arginine and glycine alleles of the codon 16 polymorphism have frequencies of 0.6 and 0.4, respectively, while the glutamine and glutamic acid alleles of the codon 27 polymorphism have frequencies of 0.4 and 0.6, respectively. With these allele frequencies, these functional polymorphisms can influence many patients' responses to beta-adrenergic agonists. Recent research suggests, however, that the effects of

these polymorphisms are more apparent when the patient is given short-acting beta-agonists. The effects of longer-acting beta agonists do not seem to be influenced as much by the *ADRB2* polymorphisms.

Recent research has suggested that a combined assessment of the individual's status for polymorphisms in the genes that encode the ADRB2 receptor, the enzyme glutathione-dependent S-nitrosoglutathione reductase (GSNOR) and the enzyme carbamoyl phosphate synthetase-1 enable one to identify approximately 70% of the patients who will not respond to albuterol. These studies have focused primarily on African-American children, and need to be repeated in other populations. Future studies should also determine whether these polymorphisms affect the response to other, longer-acting anti-asthmatic agents.

Leukotrienes (LT) are pro-inflammatory molecules that are synthesized from the fatty acid arachidonic acid. They have been implicated in a number of respiratory disorders, including asthma, chronic obstructive pulmonary disorder (COPD) and allergic rhinitis (AR), and in atopic dermatitis as well. LT receptor antagonists are commonly prescribed to patients with asthma and other respiratory diseases. They are generally well tolerated, but the patient response is highly variable, with a disappointingly high percentage of patients failing to respond.

Variants in the genes encoding the two cysteinyl leukotriene receptors (CYSLTR1 and CYSLTR2) influence the sensitivity of the receptor to LTs, as well as the individual's susceptibility to asthma and aspirin-exacerbated respiratory disease (AERD). One variant in the promoter region of the CYSLTR1 gene is associated with an increased response of the CYSLTR1 receptor to LTD_4, as well as an increased susceptibility to atopy in both men and women, and increased susceptibility to asthma in women. In addition, several SNPs in the CYSLTR2 gene have been shown to influence either the severity of asthma or the susceptibility to AERD.

In addition to assessing the individual's status for proteins that directly interact with the drug or its metabolites (receptor proteins, drug transporters, catabolic enzymes), a complete assessment of the critical functional polymorphisms must include an assessment of genes whose proteins participate in the synthesis of the hormones, neurotransmitters and other biomolecules that populate the pathways with which the drug interacts. Several of the genes that encode enzymes that synthesize LTs have potentially functional polymorphisms in them. For example, the anti-asthmatic drug zileuton inhibits the 5-lipoxygenase (ALOX5) enzyme, which catalyzes one of the steps in LT synthesis. There are two variants in the ALOX5 gene that have been reported to influence asthma severity, susceptibility to AERD, and the patient's response to both ALOX5 inhibitors and LT receptor antagonists. In addition, a polymorphism in the promoter region of the leuokotriene C_4 synthase (LTC4S) gene, whose protein also participates in the 5-lipoxygenase pathway that synthesizes LTs, influences the severity of the patient's asthma, as well as the patient's susceptibility to AERD. The leukotriene A_4 hydrolase (LTA4H) gene is another potentially important source of variability in this pathway. One study has reported a strong association of a haplotype of LTA4H variants with asthma susceptibility.

Further Readings

Bobb AJ, Castellanos FX, Addington AM, Rapoport JL (2005) Molecular genetic studies of ADHD: 1991 to 2004. Am J Med Genet B Neuropsychiatr Genet Jan 5;132B(1):109–25

Faraone SV, Perlis RH, Doyle AE, Smoller JW, Goralnick JJ, Holmgren MA, Sklar P (2005) Molecular genetics of attention-deficit/hyperactivity disorder. Biol Psychiatry Jun 1;57(11):1313–23

Lunshof JE, Bobe J, Aach J, Angrist M, Thakuria JV, Vorhaus DB, Hoehe MR, Church GM (2010) Personal genomes in progress: from the human genome project to the personal genome project. Dialogues Clin Neurosci 12(1):47–60

Munafò MR, Freimer NB, Ng W, Ophoff R, Veijola J, Miettunen J, Järvelin MR, Taanila A, Flint J (2009) 5-HTTLPR genotype and anxiety-related personality traits: a meta-analysis and new data. Am J Med Genet B Neuropsychiatr Genet Mar 5;150B(2):271–81

Rubinsztein DC, Easton DF (1999) Apolipoprotein E genetic variation and Alzheimer's disease. A meta-analysis. Dement Geriatr Cogn Disord May-Jun;10(3):199–209

Epilogue and List of Resources

Personalized Genomic-Based Medicine Is Here to Stay, But Challenges Remain

Despite the fact that significant work must be done before genetic and genomic testing can personalize all fields of medicine, it is clear that these tests will someday help improve risk estimates, diagnoses and treatments for a great many patients. At some point physicians will use the patient's genetic, epigenetic, protein, metabolic and other biological marker profiles, along with dietary, environmental and lifestyle factors, to identify predispositions for rare and common diseases, to improve diagnosis and disease stratification and to tailor screening and prevention strategies. Physicians will be able to prescribe safer and more effective treatments, with the hope of improving health outcomes and decreasing cost. Continued decreases in the costs of microarray analyses, whole genome sequencing and other analytical techniques that provide huge amounts of data in one assay will enable ever more ambitious research projects, even at the single cell level, and greater understanding of systems biology. We will continue to discover that many genes, and even pathways, that were previously considered unrelated to a particular disorder influence the risk for the disorder, or the effectiveness of specific treatments.

This research will not only enhance our understanding of the human disease process, but will also provide insights into normal physiology, as it has for T2D, where recent genomic research has identified several novel genes that affect pancreatic B-cell formation and function, as well as genes involved in other pathways that influence fasting glucose levels and obesity. We will also come to better understand that many diseases previously thought to have distinct etiologies may in fact share common molecular causes. This knowledge will aid our ability to predict the individual's risk for, and more effectively treat, many of the common multifactorial diseases that affect so many people.

K.M. Sweet and R.C. Michaelis, *The Busy Physician's Guide To Genetics, Genomics and Personalized Medicine*, DOI 10.1007/978-94-007-1147-1,

In the future, medicine will be predictive, preventive, personalized and participatory.[1] Both the physician's and the patient's roles, and their relationship, will change as medicine becomes more personalized. As more information about genes and health emerges, doctors may see an increase in the frequency of patients who are educated about the causes of disease. In addition, some patients who know they have a greater than average genetic risk for certain diseases may play a more proactive role in their own healthcare, by adjusting their dietary, environmental and lifestyle factors to minimize their overall risk of developing the disorder. At present, however, the field of personalized medicine is in its infancy, and like any infant, its arrival has been attended by a certain amount of chaos. It is true that there is considerable disagreement and discrepancy among the findings of research studies in many areas of personalized medicine. This merely reflects the fact that our genes' influence over our health is more extensive and complex than we originally thought. The next generation of research will further refine which of the initial findings can be translated into clinically useful tests, as well as identify the critical epigenetic and nongenetic factors that influence our risks for specific disorders and our responses to specific treatments.

It is clear that there are no insurmountable barriers to the progress of personalized medicine. One of the historical impediments to the advance of personalized medicine has been the fact that it will require the testing of hundreds, perhaps thousands, of gene polymorphisms to accurately estimate risk for most diseases or accurately predict the response to any given drug. Array-based analyses and whole-genome sequencing can easily accommodate the need for large quantities of data. Another historical impediment has been the fact that, in order to identify critical low-penetrance alleles, association studies should include thousands, even hundreds of thousands, of subjects. The increasing number of large, multicenter studies that are being designed and conducted will help overcome this impediment in time.

This impending explosion of detailed genetic and genomic information will present new challenges, not only for the researchers who try to integrate this information into an understanding of the mechanisms for disease, but also for lay people, who will have much greater access to their personal medical genetic information than before. Some people will eagerly embrace the "participatory" portion of P4 medicine, want to know their genetic information, and adjust their dietary, environmental and lifestyle factors in accordance with their genetic risks. Others, however, will be less eager to know their genetic information, and less apt to modify their dietary, environmental and lifestyle factors.

It is clear that further research is needed before we can develop personalized medicine tests that have appreciable clinical utility and will be useful to everyday people. It is also clear, however, that the experience we have already had with personalized medicine has proven the basic principle: that once we know what the critical genetic, epigenetic and nongenetic factors are that influence the individual's

[1]Biologist Leroy Hood has coined and trademarked the term "P4™ medicine" to reflect the breadth of the changes that are anticipated as the field of personalized medicine develops.

susceptibility to a disease or response to a drug, we will be able to develop tests that estimate risks accurately enough and prevent enough drug failures and ADRs to be worth incorporating into routine practice. Genotyping is already preventing ADRs and drug failures in some cases, and it is clear that, with more research, tests can be developed that will allow us to use genotyping even more effectively to improve outcomes and save lives.

We also predict that the public will contribute to the development of this field to a much greater degree than it has contributed to the development of other medical fields. The public has been more willing to seek and share genetic information than previous generations have been to seek and share information about other aspects of their health. Social networking sites have enabled people to place their genetic data in the public domain. Many people have also chosen to participate in focused projects such as Harvard's Personal Genomes Project, which aims to recruit 100,000 volunteers to contribute their genomic sequence data, extensive information about nongenetic factors and other information to a publicly accessible and identifiable research database. Projects such as these will allow us to understand the full spectrum of variability in the human genome, and especially enable us to understand the way that rare gene variants, or gene variants with low penetrance, influence our risks for diseases or responses to treatments.

Biomedical advances such as these often give rise to social, ethical and legal concerns. The federal Genetic Information Nondiscrimination Act of 2008 (GINA) sought to allay the public's worries that genetic data might be used by insurance companies or employers to discriminate against those who have high genetic risks for specific diseases. There is clear potential for one's genetic information to be misused in this manner, because we all possess gene variants that constitute risk-increasing alleles for one disease or another. In addition to the possibility that institutions might misuse people's genetic data, however, some worry that individuals may misuse the information, or misinterpret the information in ways that harm the individual. The way we perceive ourselves and others may change; some people may become fatalistic when they learn they have a highly penetrant risk-increasing allele, or discriminate against others who they see as genetically inferior. In addition, people may push the limits of what is socially acceptable by seeking to use genetic information to choose the gender or other characteristics of their offspring. Currently there is much debate as to whether new genomic technologies should be regulated, with commercial enterprises arguing that the consumer should have free access to genetic testing, while others, especially those in the medical and scientific community, argue that genetic information should be handled like other kinds of medical information, and only disseminated to individuals in a setting in which a physician or genetic counselor can insure the proper interpretation of the data.

As you can see, there remain a number of hurdles to overcome. Real success will be achieved not only when we are able to offer accessible and affordable genetic and genomic data to all individuals, but when we can also provide the necessary tools and information to allow individuals to take an active role in their own health care. We need to educate as many lay people as possible about the way personalized medicine will help improve healthcare for them and their family members, and

provide physicians the training they need to be able to incorporate personalized medicine into their practices. Only then will we realize the full promise of predictive, preventive, personalized and participatory (P4™) medicine.

Resources to Help You Keep up to Date with the Status of the Field and Further Your Education

With this book we have tried to refresh your memory of the foundational material, describe the tests that are being developed and the issues one must be aware of to interpret their findings, and give you information about the clinically useful tests that are currently available in a number of medical fields. We hope this has not only helped you better understand the present state of the field, but also become better aware of what kind of genomic tests are available for your patients, and what issues you must consider to decide if a particular genomic test will truly be beneficial to your patient. No book can provide you with anything more than a beginning, however. The field is growing rapidly, and the Internet is the only medium that moves quickly enough to keep you abreast of the changes that are to come very soon in this field.

Chapter 4 provides several Internet resources that will help you keep abreast of recent developments in pharmacogenetics and pharmacogenomics, including instructions on signing up for FDA email updates. In addition, the NIH's Gene Clinics website (http://www.geneclinics.org) includes the international GeneTests laboratory directory listing not only clinical testing laboratories from the USA and other countries, but also research laboratories that perform genetic tests pertinent to particular genes or diseases. Most research laboratories are not certified to produce formal medical reports, but virtually all research laboratories that perform genomic medicine testing as part of their ongoing study of that gene or disease produce reliable information that Sometimes can be confirmed by a CLIA-certified laboratory. The individual practitioner will need to consider the relative risks and benefits of acquiring this information before it has been formally approved for clinical use.

If a physician wants to keep abreast of the current trends in the translation from basic research discoveries to clinical practice, the website for the CDC's National Office of Public Health Genomics (NOPHG) contains reports from the CDC's Evaluations of Genomic Applications in Practice and Prevention (EGAPP) project (http://www.egappreviews.org/). The EGAPP project's mission is to review the pharmacogenomics literature and clinical trial results, and assess how ready any given effort is to be translated into routine clinical practice or public health policy. This website provides a summary of the current opinion that often shapes the FDA's decisions to approve products and services for use in the clinic, and often allows one to assess the current state of a specific effort and anticipate developments that will arise in the near future.

To keep abreast of the new guidelines that emerge from the many regulatory and advisory bodies, the CDC maintains the Agency for Healthcare Research and Quality (AHRQ, http://www.ahrq.gov/). The AHRQ website also contains the National Guideline Clearinghouse™, which catalogs and collates the guidelines that have been issued by various regulatory and advisory bodies. In addition to these federal resources, many professional societies and organizations include in their websites a link to resources that summarize recent developments in research and clinical applications as well as guidelines for putting new products and services into routine practice.

For those who want to review the basic research data from genetic epidemiology studies, the NOPHG website also includes the site for the Human Genome Epidemiology Network (HuGENet) and the HuGE Navigator (http://www.cdc.gov/genomics/hugenet/default.htm). The HuGENet is a collection of research groups that voluntarily share the data from their genomic epidemiology studies. The HuGE Navigator allows one to search and collate information from genetic association studies that focus on one's disease of interest.

For those wishing to continue their education in genomic medicine, the National Coalition for Health Professional Education in Genetics (NCHPEG) website (http://www.nchpeg.org/) includes some continuing education materials on genetics and health that were developed in association with the American Academy of Family Physicians. In addition, the AMA and the FDA have developed an online CME course entitled *Pharmacogenomics and Personalized Medicine*, which can be accessed at the AMA website (http://www.ama-assn.org/ama/home/index.shtml). The AHRQ's Centers for Education and Research on Therapeutics (CERTs) provide information intended to educate the physician on the use of therapeutic drugs, including new developments in pharmacogenomics (http://www.ahrq.gov/clinic/outcomix.htm#CERTs).

Further Readings

Incorporating Personalized Medicine into Practice

Evaluation of Genomic Applications in Practice and Prevention (EGAPP): This multidisciplinary panel evaluates genetic and genomic tests that have clinical application. Reports and recommendations are published. www.egappreviews.org/

Ioannidis JPA, Boffetta P, Little J, et al. (2008) Assessment of cumulative evidence on genetic associations: interim guidelines. Int J of Epidemiology, 37:120–132

Khoury MJ, Feero WG, Reyes M, et al. (2009) The genomic applications in practice and prevention network. Genet Med, 11(7):488–494

Khoury MJ, McBride CM, Schully SD, et al. (2009) The scientific foundation for personal genomics: recommendations from a National Institutes of Health – Centers for Disease Control and Prevention multidisciplinary workshop. Genet Med,11(8):559–567

Scheuner MT, Sieverding P, and Shekelle PG (2008) Delivery of genomic medicine for common chronic adult diseases. JAMA, 299(11):1320–1334

Interpreting Genome-Wide Association Studies

How to use an article about genetic association (three part JAMA series)

Attia J, Ioannidis JPA, Thakkinstian A, et al. (2009) Part A: Background concepts. JAMA, 301(1):74–81

Attia J, Ioannidis JPA, Thakkinstian A, et al. (2009) Part B: Are the results of the study valid? JAMA, 301(2):191–197

Attia J, Ioannidis JPA, Thakkinstian A, et al. (2009) Part C: What are the results and will they help me in caring for my patients? JAMA, 301(3):304–308

Manolio TA, Collins FS, Cox NJ, et al. (2009) Finding the missing heritability of complex diseases. Nature, 461(8):747–753

Pearson TA, and Manolio TA (2008) How to interpret a GWAS Study. JAMA, 299(11): 1335–1344

Databases Containing Information About Variants Associated with Diseases

dbSNP: The NCBI database of SNPs. Good for finding SNP locations, basic info and frequencies in various populations. http://www.ncbi.nlm.nih.gov/projects/SNP/

SNPedia: A user-curated database of SNPs and their associations. It can be a useful starting place to look up a disease and related SNPs or vice versa. There are helpful links to the primary literature and other databases for search results. http://www.snpedia.com

PharmGKB: Database of curated data on gene variants and gene-drug-disease relationships. http://www.pharmgkb.org/

PolyPhen: An online tool that predicts the functional effect of SNPs. http://genetics.bwh.harvard.edu/pph/

HuGE Navigator: CDC-developed multifunctional tool to aid genetic association and human genome epidemiology searches. http://hugenavigator.net/

NHGRI Catalog of Published GWAS: A searchable list of published GWA studies. http://www.genome.gov/26525384

Preventive Health Guidelines

USPSTF (US Preventive Services Task Force): Current recommendations for clinical preventive services based on evidence of clinical effectiveness. http://www.ahrq.gov/clinic/uspstfix.htm

NCCN (National Comprehensive Cancer Network): Contains guidelines for management of increased cancer risk. Requires registration (free). http://www.nccn.org/index.asp

Educational Web Resources

Genetics Science Learning Center at the University of Utah. Contains interactive learning tools on topics ranging from basic genetics to pharmacogenomics. http://learn.genetics.utah.edu/

NHGRI fact sheet on Genome Wide Association Studies. http://www.genome.gov/20019523
National Institute of General Medical Sciences Pharmacogenomics Research Network Educational materials for those interested in pharmacogenomics. http://www.nigms.nih.gov/Initiatives/PGRN/Education.htm
Department of Energy Human Genome Project Information. Includes information on basic genetics, gene therapy, genetic testing, pharmacogenomics, and ELSI issues. http://www.ornl.gov/sci/techresources/Human_Genome/project/about.shtml
CDC's Public Health Genomics site. Links to multiple programs and publications including the weekly email bulletin of genomics in the news and literature. http://www.cdc.gov/genomics/

Ethical, Legal and Social Issues

http://www.genomicslawreport.com
http://scienceblogs.com/geneticfuture/

Index

A
ACCE Project, 56
Adenine, 2, 13
Adverse drug reactions, 85, 88, 137, 165, 179
Age-related macular degeneration (AMD), 156, 157, 173–174
Aggregate load, 21
Alcohol, 179, 189–190
Allele, 1, 8, 9, 12, 18, 20–23, 25–31, 38, 44, 51, 53–60, 62, 63, 65, 67–72, 76, 78, 79, 87–89, 93, 97, 99–101, 103, 104, 108–111, 113, 114, 122–125, 127, 130, 133, 137–139, 147, 149, 150, 152, 153, 155, 156, 160, 167–171, 174, 175, 177, 178, 185, 187–196
Allelic imbalance, 26
Alternative splicing, 25
Alzheimer disease, 31, 168, 173, 192–195
AmpliChip CYP450 microarray, 75
Analytic validity, 56
Anticipation, 30, 62
Anticodon, 15, 16
Antihypertensive drugs, 170
Antiparallel, 4
Anxiety disorders, 186
Arrhythmogenic right ventricular cardiomyopathy (ARVC), 165
Association, 47, 53, 67–74, 79, 89, 96, 97, 109, 110, 126, 141, 145–147, 149, 152, 168, 174, 175, 180, 181, 185, 186, 189, 197
Association for Molecular Pathology Test Directory, 53
Asthma, 52, 96, 155, 173, 196–197
ATP binding cassette proteins multidrug resistance proteins, 103
Attention deficit-hyperactivity disorder (ADHD), 179–180
Attributable risk, 70, 174
Autoinhibition, 95
Autophenocopying, 95
Autosomal, 35, 37–41, 43–45, 52, 113, 121, 122, 127, 128, 131, 144, 150, 151, 158–162, 192
Autosomal dominant, 35, 39–40, 43, 113, 121, 122, 127, 128, 131, 144, 150, 151, 158–162, 192
Autosomal recessive, 35, 37, 40–41, 44, 45, 127, 128, 159–161
Autosomes, 6, 8, 38, 40, 41

B
Basepairs, 4, 5
Beta-blockers, 91, 104–105, 166
Bipolar Disorder (BPD), 181–183, 195
BRACAnalysis, 118
BRACAnalysis Rearrangement Test (BART), 118
Butyrylcholinesterase, 102

C
CancerGene 5.1 model, 112
Cancer Genome Atlas project, 108
Cardiac channelopathies, 142, 150–151
CardioSmart risk assessment tool, 144
Cardiovascular disease (CVD), 141
Carrier frequency, 45, 46
Case-control studies, 68, 147
Centromere, 6, 28
Chromatin, 5, 6, 27, 28
Chromosome rearrangements, 24, 54, 64, 66, 82
Claus model, 112
Clinical utility, 2, 51, 56, 57, 59, 67, 94, 113, 115, 119, 148, 177, 183
Clinical validity, 56–58

Clopidogrel, 92, 156, 171
Coding sequence, 5, 8, 12, 13, 16, 18, 20, 24, 30, 53, 54, 60, 61, 96, 132, 136, 174
Coding strand, 13
Codon, 15, 16, 20, 60, 61, 133, 196
Cognitive decline in aging, 192–195
Comparative genomic hybridization (CGH), 78
Complementary basepairs, 4
Complementary DNAs (cDNAs), 80
Complex disorders, 18, 64
Comprehensive BRACAnalysis, 117, 118
Conjugation reactions, 87
Copy number variations (CNVs), 8, 31, 64, 180–181
Coronary artery disease (CAD), 142
Cowden syndrome (CS), 114
CYP450 enzymes, 21, 88, 98, 165–167, 171, 177
Cytosine, 2, 13

D
DailyMed project, 105
De novo mutation, 39, 131
Deoxyribose, 2, 3
Depression, 178, 179, 181–185
Dideoxysequencing, 76
Dilated cardiomyopathy (DCM), 164
Diploid, 6
DNA mismatch repair genes, 127
Dominant, 35, 38–40, 42–44, 63, 68, 74, 113, 121, 122, 127, 128, 131, 144, 150, 151, 158–162, 192
Drug-induced liver injury, 88
Dyslipidemia, 149, 175

E
Enantiomers, 94, 100, 166
3' end, 3
5' end, 3, 30
Endophenotypes, 179
Epigenetic factors, 1, 2, 27–29, 34, 48, 65, 82, 97, 113, 130, 180
Epistasis, 93, 113
Essential hypertension, 149–150, 170
Ethical, legal and social issues, 56, 57
Euchromatin, 28
Evaluation of Genomic Applications in Practice and Prevention (EGAPP), 57
Exons, 13, 14, 25, 26, 132
Extensive metabolizers, 76, 98

F
Familial adenomatous polyposis (FAP), 127, 131–132
Familial hypercholesterolemia, 141, 144, 157–159
Family HealthLink, 144, 154
Family history, 35–48, 66, 90, 93, 107, 112–118, 120–122, 129, 132, 141–144, 148, 153–155, 158, 159, 192
Fluorescence in situ hybridization (FISH), 79
Founder mutations, 115, 118
Framingham heart study, 144
Framingham risk score, 144, 147
Fusion gene, 24, 135
Fusion protein, 24, 135

G
Gail model, 112
G-banding, 82–83
Gene, 1, 35, 51, 85, 106, 141, 173
GeneTests website (http://www.ncbi.nlm.nih.gov/sites/GeneTests/), 38, 52, 105, 161
Genetically linked, 71, 110
Genetic code, 16, 19, 29, 46
Genetic consultation, 46–48, 129, 145
Genetic load, 21, 115
Genetics home reference website, 38
Genome-wide association studies, 109, 126
Genotype, 9, 22, 23, 26, 30, 33, 39, 41, 44, 45, 69, 70, 74, 78, 81, 95, 96, 99, 100, 103, 104, 110, 113, 148, 149, 166, 169, 170, 178, 187–192, 194–196
Glioblastomas, 137
Golgi apparatus, 17
Guanine, 2, 13, 97

H
Haploid genome, 6
Haplotype, 72–73, 147, 188, 197
Haplotype blocks, 72–73
Hemizygous, 9, 31, 38
Hereditary breast-ovarian cancer (HBOC) syndrome, 52, 97, 107, 114–116, 121
Heterochromatin, 28
Heterozygous, 9, 21, 38, 40, 43, 44, 63, 78, 113, 118, 122, 133, 150, 155, 156, 158, 164
High-activity alleles, 18, 93
Histone proteins, 2, 5, 6, 27, 28, 65
Homocysteine, 141, 153, 170

Homozygous, 9, 23, 29, 38, 40, 63, 104, 128, 133, 137, 138, 155–158, 167, 169, 174, 175, 187, 190, 193, 194, 196
Hypertrophic cardiomyopathy, 141, 157, 161–163

I
Immune system hypersensitivity reactions, 89
Immunohistochemical (IHC) analyses, 80–81, 129
Imprinted genes, 27, 28
Inflammation, 101, 141–143, 152, 168, 169
Insertion/deletion (ins/del or indel) polymorphism, 29, 60, 168, 170, 178, 185
Interfering RNAs, 1, 26–27
Intermediate metabolizers, 76, 98
Introns, 13, 14, 25, 26, 54
Isoform, 9, 10, 18, 25, 53, 54, 60, 87, 100, 104, 136, 149, 178, 186–188, 193

L
Li-Fraumeni syndrome (LFS), 114
Linkage disequilibrium, 71, 126
Loci, 8, 9, 31, 65, 145, 148, 149, 161
Loss of heterozygosity (LOH), 63
Low-activity alleles, 18, 87, 96, 100, 177
Lynch syndrome, 81, 107, 113, 127–130

M
Markers, 70–73, 96, 109–111, 113, 145–148, 156, 165, 175–177
Meiosis, 10–12, 24, 30, 62, 63, 71–73
Mendelian disorders, 37
Messenger RNA (mRNA), 37
Microarray, 22, 64, 65, 73, 75–77, 79, 81, 82, 92, 93, 145, 147, 148, 181
MicroRNAs (miRNAs), 26
Microsatellite analysis, 61–63
Microsatellite instability (MSI), 62–63
Missense mutation, 60
Missense variant, 60
Mitochondrial DNA variants, 65
Monoamine oxidase A (MAOA) gene, 186
Multifactorial disorders, 18, 30–32, 46, 58, 59, 62, 64, 68, 173–197
Multisite 3 BRACAnalysis, 118
Mutation, 17, 18, 22, 23, 35, 37–45, 52–54, 60, 63, 66–68, 73, 74, 108, 112–127, 129–134, 136, 147, 151, 158–160, 162, 164

N
N-acetyltransferase 2, 101–102
National Comprehensive Cancer Network (NCCN), 114, 115, 124
National Society of Genetic Counselors, 46
Negative predictive validity (NPV), 58
Nicotine, 191–192
Nitrogen-containing base, 2, 3
Nongenetic load, 21
Nonpenetrance, 21, 55, 58
Nonrecombinant chromosome, 12
Nucleosome, 5
Nucleotides, 2–5, 8, 10, 13–15, 19, 20, 24, 25, 27, 29, 54, 55, 59–63, 65, 73, 75–79, 81, 82, 97, 117, 155, 167, 168, 193
Nucleus, 5, 6, 15, 132
Null alleles, 17, 76, 98–101, 171

O
Odds ratio (OR), 69, 70, 120, 127, 156, 174, 175, 177
Oncogenes, 108
Oncotype DX test, 80, 134, 135
Online Mendelian inheritance in man (OMIM), 37, 159
Oogenesis, 10, 11
Oogonia, 10
Opiates, 190–191
Organic anion transporter, 89, 103
Organic anion transporter statin-induced myopathy, 89, 103
Organic cation transporters, 103, 104
5’ to 3’ Orientation, 3, 4
Oxidative stress, 152, 165

P
9p21.3, 146–148
p arm, 6
Penetrance, 21–22, 33, 48, 51, 52, 54–58, 63, 67–70, 79, 109–110, 113, 114, 116, 119, 123, 124, 126–128, 131, 145–153, 157–164, 175, 177, 181, 192
Peptide bonds, 17
Pharmacogenetics, 85–106

Pharmacogenomics, 85, 90, 98, 99, 105, 106, 165–171, 177–178
Phase I, 86, 87, 89, 98
Phase II, 86, 87
Phenocopy, 95, 100
Phenotype, 32, 39, 43, 63, 74, 89, 95, 96, 98–101, 108, 127, 129, 133, 137, 141–143, 145, 147–153, 157, 158, 160–162, 164, 166, 167, 171, 174, 179, 190
Philadelphia chromosome, 24, 135
Phosphate group, 2, 3, 17
Polymerase chain reaction (PCR), 65, 66, 74–78, 80–82, 97, 134
Polymorphic, 8–11, 26, 33, 54, 63, 68, 70, 71, 87, 98, 151
Polymorphism, 8–10, 12, 17, 19, 20, 22–27, 29, 30, 32, 34, 51, 55, 58–63, 65, 66, 68–72, 74–76, 78, 79, 86–90, 92–106, 110, 113, 117, 126, 130, 135–138, 147, 149–150, 152, 153, 155, 160, 167–171, 174, 175, 177–192, 194–197
Polypeptide, 15–17, 20, 61
Poor metabolizers, 76, 98, 137, 156, 165
Positive predictive validity (PPV), 58
Pre-messenger RNA, 13, 14, 19, 25
Primary transcript, 13
Primer extension assay, 77
Promoter region, 12, 13, 22–27, 30, 34, 54, 61, 65, 82, 95–97, 130, 138, 160, 167, 170, 185, 187, 188, 197
Proton pump inhibitors (PPIs), 92, 155, 171
Proto-oncogenes, 108
PTEN hamartoma tumor syndrome, 113, 114
(P4) medicine, 48
Pyridoxine, 153, 171

Q
q arm, 6

R
Recessive, 37–42, 44–46, 63, 68, 127, 128, 133, 159–161
Recombinant chromosome, 12
Recombination, 10–12, 55, 63, 71–73
Recombination frequency (Rf), 71
Regulatory sequences, 1, 12, 34, 61
Relative risk, 70, 106, 109–111, 147, 175
Repeated sequence length polymorphisms, 61–63
Repetitive sequences, 30
Restriction endonucleases (REs), 66, 76, 77
Reverse transcription (RT), 80, 134
Reverse transcription-Polymerase chain reaction (RT-PCR), 80, 81
Reynolds risk score for women, 144
Ribonucleotides, 13
Ribosome, 15–17, 20, 25, 26, 61, 81
Risk-decreasing alleles, 18, 33
Risk-increasing alleles, 18, 21–22, 25, 31, 33, 51, 52, 54–59, 67–72, 79, 109, 110, 113, 114, 122–126, 130, 142, 145, 147, 150, 155, 156, 158, 168, 174, 175, 177, 193
RNA, 2, 6, 13–15, 19, 22, 25–27, 33, 54, 55, 76, 81, 123, 136, 137
RNA polymerase, 13, 22
Rough endoplasmic reticulum, 17

S
Schizophrenia, 59, 180–183, 194
Serotonin transporter (5HTT), 23, 180, 182–186, 188
Sex chromosome, 6
Silent base substitution, 54
Single-gene disorders, 21, 31, 32, 37–38, 45, 58, 67, 149, 157, 158
Single nucleotide polymorphism (SNPs), 10, 29, 59–60, 155
Sister chromatids, 10, 11
Small interfering RNAs (siRNAs), 27
Sodium/potassium regulation, 149–150, 165
Somatic mutations, 63, 66, 107, 108, 137
Spermatogenesis, 10, 11
Spermatogonia, 10
Splice site mutations, 14, 160
START codon, 16
Stimulants, 91, 187–190
STOP codons, 16
Substance abuse, 186–192
Synonymous base substitution, 19, 54
Synonymous substitutions, 19, 54
Synthetic lethality, 136

T
Table of Valid Genomic Biomarkers in the Context of Approved Drug Labels, 105
Tagging SNPs, 73
Telomere, 6
Template strand, 13
Tetrad, 10, 11
Thymine, 2

Transcription, 13, 15, 22–24, 27, 29, 55, 61, 88, 132, 134, 168, 175, 176, 182, 183, 185, 191
Transcription factor, 13, 22, 23, 29, 61, 132, 168, 175, 176, 182, 183, 191
Transfer RNAs, 15
Translation initiation codon, 16
Translation termination codons, 16
Trinucleotide repeats, 30, 62
Tumor suppressor genes, 26, 108, 114
Type 2 diabetes mellitus (T2D), 104, 141, 149, 153–157, 174–178
Typical-activity allele, 18
Typical-risk alleles, 18

U
Ultrarapid metabolizers, 137, 138, 190
Unaffected carrier, 39–42
Uracil, 13

V
Variant alleles, 9, 153, 156, 193, 196
Variant isoforms, 9
VKOR complex, 95, 170

W
Warfarin, 92, 95, 169–170
Wild-type, 9, 20, 38, 77, 78, 136

X
X-linked, 23, 38, 39, 41–42, 44
X-linked dominant, 42–43
X-linked recessive, 41–42, 44

Y
Y-linked, 38, 39, 43
Your Disease Risk, 144