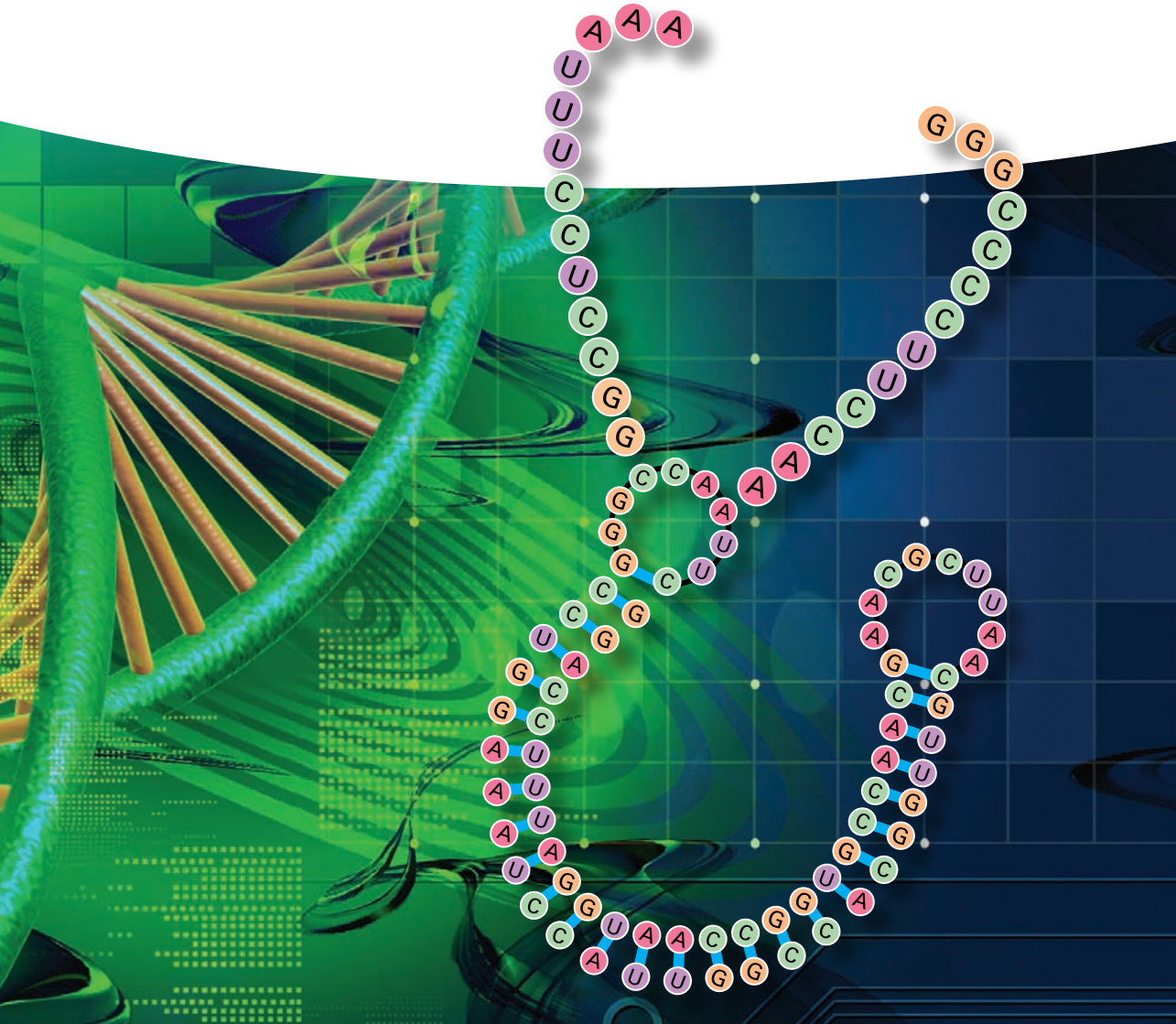
WILEY-VCH

Edited by
Andreas Keller and Eckart Meese

Nucleic Acids as Molecular Diagnostics



Edited by
Andreas Keller
Eckart Meese

Nucleic Acids as Molecular Diagnostics

Related Titles

Popp, J. (ed.)

Ex-vivo and In-vivo Optical Molecular Pathology

2014
Print ISBN: 978-3-527-33513-8,
also available in digital formats

Popp, J., Bauer, M. (eds.)

Modern Techniques for Pathogen Detection

2014
Print ISBN: 978-3-527-33516-9

Whitehouse, D., Rapley, R.

Molecular and Cellular Therapeutics

2012
Print ISBN: 978-0-470-74814-5,
also available in digital formats

Rapley, R, Harbron, S. (eds)

Molecular Analysis and Genome Discovery

2nd Edition
2011
Print ISBN: 978-0-470-75877-9,
also available in digital formats

Edited by Andreas Keller and Eckart Meese

Nucleic Acids as Molecular Diagnostics

WILEY-VCH
Verlag GmbH & Co. KGaA

Editors

Prof. Dr. Andreas Keller
Chair for Clinical Bioinformatics
Saarland University, University Hospital
Building E2.1
Saarbrücken 66123
Germany

Prof. Dr. Eckart Meese
Universität des Saarlandes
Institut für Humangenetik/Geb. 60
Kirrberger Str.
Germany

Cover
Background photo.
Source: fotolia.com
© krishnacreations

Library of Congress Card No.: applied for

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

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.d-nb.de.

Print ISBN: 978-3-527-33556-5
ePDF ISBN: 978-3-527-67223-3
ePub ISBN: 978-3-527-67222-6
Mobi ISBN: 978-3-527-67221-9
oBook ISBN: 978-3-527-67216-5

Cover Design Bluesea Design, McLeese Lake, Canada
Typesetting Thomson Digital, Noida, India
Printing and Binding Markono Print Media Pte Ltd, Singapore

Printed on acid-free paper

Contents

List of Contributors

Saskia Biskup
CeGaT GmbH
Paul-Ehrlich-Strasse 23
72076 Tübingen
Germany

Nikolaus Blin
Wroclaw Medical University
Department of Genetics
Marcinkowskiego 1
50-368 Wroclaw
Poland

Johannes Dietl
University of Würzburg Medical School
Department of Obstetrics and Gynecology
Josef-Schneider-Strasse 4
97080 Würzburg
Germany

Juana Díez
Universitat Pompeu Fabra
Department of Experimental and Health Sciences
Molecular Virology Laboratory
Doctor Aiguader 88
08003 Barcelona
Spain

Claudia Durand
CeGaT GmbH
Paul-Ehrlich-Strasse 23
72076 Tübingen
Germany

Peter J. Goebell
Universitätsklinikum Erlangen
Urologische Klinik
Krankenhausstrasse 12
91054 Erlangen
Germany

Jan Haas
University of Heidelberg
Department of Internal Medicine III
Im Neuenheimer Feld 410
69120 Heidelberg
Germany

and

University of Heidelberg
DZHK (German Centre for Cardiovascular Research)
Im Neuenheimer Feld 410
69120 Heidelberg
Germany

Sebastian F.M. Häusler
University of Würzburg Medical School
Department for Obstetrics and Gynecology
Josef-Schneider-Strasse 4
97080 Würzburg
Germany

Wolfram Henn
Universität des Saarlandes
Institut für Humangenetik
Campus Homburg, Gebäude 68
66421 Homburg
Germany

Pawel Karpinski
Wroclaw Medical University
Department of Genetics
Marcinkowskiego 1
50-368 Wroclaw
Poland

Hugo A. Katus
University of Heidelberg
Department of Internal Medicine III
Im Neuenheimer Feld 410
69120 Heidelberg
Germany

and

University of Heidelberg
DZHK (German Centre for Cardiovascular Research)
Im Neuenheimer Feld 410
69120 Heidelberg
Germany

Stefan Kirsch
Fraunhofer Institut für Toxikologie und Experimentelle Medizin
Personalisierte Tumortherapie
Josef-Engert-Straße 9
93053 Regensburg
Germany

Jan Kirsten
Merck Serono/Merck KGaA
Fertility Technologies
Frankfurter Straße 250
64293 Darmstadt
Germany

and

Merck Serono Diviison of Merck KGaA
Head of Fertility Technologies – Global Business Franchise Fertility
Frankfurter Str. 250
64293 Darmstadt
Germany

Christoph A. Klein
Fraunhofer Institut für Toxikologie und Experimentelle Medizin
Personalisierte Tumortherapie
Josef-Engert-Straße 9
93053 Regensburg
Germany

and

University Regensburg
Chair of Experimental Medicine and Therapy Research
Franz-Josef-Strauß-Allee 11
93053 Regensburg
Germany

Cédric C. Laczny
University of Luxembourg
Luxembourg Centre for Systems Biomedicine
7 Avenue des Hauts-Fourneaux
4362 Esch-sur-Alzette
Luxembourg

Irene Latorre
Saarland University
Human Genetics Department
Kirrberger Strasse, 66424
Homburg
Germany

and

Universitat Pompeu Fabra
Infection Biology Laboratory
Department of Experimental and Health Sciences
Doctor Aiguader 88, 08003
Barcelona
Spain

and

CIBER Enfermedades Respiratorias (CIBERES)
Barcelona
Spain

Hans-Peter Lenhof
Universität des Saarlandes
Zentrum für Bioinformatik
Gebäude E2.1
66041 Saarbrücken
Germany

Benjamin Meder
University of Heidelberg
Department of Internal Medicine III
Im Neuenheimer Feld 410
69120 Heidelberg
Germany

and

University of Heidelberg
DZHK (German Centre for Cardiovascular Research)
Im Neuenheimer Feld 410
69120 Heidelberg
Germany

Andreas Meyerhans
Universitat Pompeu Fabra
Infection Biology Laboratory
Department of Experimental and Health Sciences
Doctor Aiguader 88, 08003
Barcelona
Spain

and

Institució Catalana de Recerca i Estudis Avançats (ICREA)
Barcelona
Spain

Christian P. Pallasch
University of Cologne
Department I of Internal Medicine and Center of Integrated Oncology (CIO)
Kerpener Strasse 62
50937 Cologne
Germany

Bernhard Polzer
Fraunhofer Institut für Toxikologie und Experimentelle Medizin
Personalisierte Tumortherapie
Josef-Engert-Straße 9
93053 Regensburg
Germany

Patrick Roth
University Hospital Zurich
Department of Neurology and Brain Tumor Center
Frauenklinikstrasse 26
8091 Zurich
Switzerland

Verónica Saludes
Universitat Pompeu Fabra
Molecular Virology Laboratory
Department of Experimental and Health Sciences
Doctor Aiguader 88, 08003
Barcelona
Spain

and

CIBER Epidemiología y Salud Pública (CIBERESP)
Barcelona
Spain

Maria M. Sasiadek
Wroclaw Medical University
Department of Genetics
Marcinkowskiego 1
50-368 Wroclaw
Poland

Michael G. Schrauder
Universitätsklinikum Erlangen
Frauenklinik
Universitätsstrasse 21–23
91054 Erlangen
Germany

Janine Schwamb
University of Cologne
Department I of Internal Medicine and Center of Integrated Oncology (CIO)
Kerpener Strasse 62
50937 Cologne
Germany

Daniel Stöckel
Universität des Saarlandes
Zentrum für Bioinformatik
Gebäude E2.1
66041 Saarbrücken
Germany

Reiner Strick
Universitätsklinikum Erlangen
Frauenklinik
Universitätsstrasse 21–23
91054 Erlangen
Germany

Helge Taubert
Universitätsklinikum Erlangen
Urologische Klinik
Krankenhausstrasse 12
91054 Erlangen
Germany

Britta Vogel
University of Heidelberg
Department of Internal Medicine III
Im Neuenheimer Feld 410
69120 Heidelberg
Germany

and

University of Heidelberg
DZHK (German Centre for Cardiovascular Research)
Im Neuenheimer Feld 410
69120 Heidelberg
Germany

Sven Wach
Universitätsklinikum Erlangen
Urologische Klinik
Krankenhausstrasse 12
91054 Erlangen
Germany

Michael Weller
University Hospital Zurich
Department of Neurology and Brain Tumor Center
Frauenklinikstrasse 26
8091 Zurich
Switzerland

Paul Wilmes
University of Luxembourg
Luxembourg Centre for Systems Biomedicine
7 Avenue des Hauts-Fourneaux
4362 Esch-sur-Alzette
Luxembourg

Jörg Wischhusen
University of Würzburg Medical School
Department for Obstetrics and Gynecology
Section for Experimental Tumor Immunology
Josef-Schneider-Strasse 4
97080 Würzburg
German

Bernd Wullich
Universitätsklinikum Erlangen
Urologische Klinik
Krankenhausstrasse 12
91054 Erlangen
Germany

Preface

With the selected contributions presented in this volume we set out to shed light on the role of nucleic acids as molecular diagnostic tools from different perspectives. We invited clinicians, biologists, and bioinformaticians to present their views on this intriguing topic. Their contributions offer a broad coverage of methods, biological targets, and clinical applications.

As for the different biological targets, the diagnostic roles of nucleic acids are addressed on a systemic level (e.g., body fluids), on an organ level (e.g., different cancer tissues), and, most challenging, on the single-cell level. On the molecular level, the different targets include DNA as well as coding and non-coding RNA (ncRNA). From the clinical perspective, the chapters address different human diseases, including the most lethal diseases (i.e., cardiovascular and cancer diseases). The diagnostics of infectious diseases as one of the leading healthcare challenges is addressed with specific emphasis on nucleic acids for the detection of viral and bacterial pathogens. The methods addressed include array-based and next-generation sequencing (NGS)-based techniques. All of the aforementioned topics (i.e., methods, biological targets, and clinical applications) may be used legitimately for an overall book structure; however, we chose the clinic/biology topics for structuring since the clinical application is the crucial endpoint of any nucleic acid-based diagnostic.

The first group of chapters (Chapters 1–8) address cardiovascular and cancer diseases, and the specific challenges for nucleic acid-based diagnoses for these diseases. Chapter 1 by Haas *et al.* describes the application of NGS for the genetic diagnostics of cardiomyopathies. The roles of microRNAs (miRNAs) as biomarkers for cardiomyopathies are described in Chapter 2 by Vogel *et al.*, who specifically address their diagnostic potential in coronary artery disease, cardiac ischemia and necrosis, and heart failure. In Chapter 3, Roth and Weller address the diagnostic potential of miRNAs in various brain tumors, including the generally benign meningiomas. As an example for one of the most common and lethal cancer diseases, in Chapter 4, Karpinski *et al.* focus on sporadic colon cancer and its specific genetic and epigenetic alterations, including chromosomal and microsatellite instability. Wullich *et al.*, in Chapter 5, address biomarkers for the three most prominent urologic malignancies: bladder cancer, prostate cancer, and renal cell carcinoma. In Chapter 6 on molecular markers in breast cancer,

Schrauder and Strick specifically address long intergenic ncRNAs, which are increasingly recognized as important ncRNAs in addition to miRNAs. Other tumors also treated by gynecological oncologists are addressed by Häusler *et al.* in Chapter 7, who summarize the emerging role of DNA-, RNA-, and miRNA-based diagnostics in gynecological oncology. While the aforementioned contributions concern solid tumors, Chapter 8 by Schwamb and Pallasch addresses nucleic acid-based approaches in the diagnosis of hematopoetic malignancies.

The second group of contributions (Chapters 9–11) deals with infectious diseases. Latorre *et al.* give a summary of nucleic acid-based diagnostic methods in the management of bacterial and viral infectious diseases in Chapter 9. Chapter 10 by Saludes *et al.* addresses questions of nucleic acid-based diagnostics in infectious diseases, specifically the diagnostic potential of miRNAs in *Mycobacterium tuberculosis* and chronic hepatitis C virus infections. Laczny and Wilmes take a broader microbiology approach in Chapter 11 in that they address compositional and functional changes in endogenous microbial communities. They use metagenomic data for a microbiome-based diagnostics and modified therapeutic intervention.

Chapters 12–14 focus on technical approaches, including bioinformatics tools. In Chapter 12, Durand and Biskup deal with challenges of sequencing in a clinical setting. Chapter 13 by Kirsch *et al.* is dedicated to one of the ultimate challenges in nucleic acid-based diagnostics – the analysis of single cells. Single-cell analysis allows us to both address challenges associated with a heterogeneous cell population as found in tumor tissues and to utilize circulating tumor cells for diagnostic purposes. Chapter 14 on bioinformatics approaches by Stöckel and Lenhof is dedicated to problems that hamper the routine clinical application of biomarkers. Topics covered in this bioinformatics chapter include dealing with the noise of high-dimensional data produced by the applied biotechnological high-throughput, with batch effects by various experimental environments, and with frequent switchovers of the experimental platforms.

Finally, two chapters (Chapters 15 and 16) are dedicated to general healthcare and ethical questions. Kirsten addresses economical aspects, specifically the key drivers for the companion diagnostics that have become increasingly important beside the primary diagnostic, in Chapter 15. Henn emphasizes the importance of ethical and legal issues for molecular genetic diagnosis in Chapter 16. In his chapter, Henn addresses key aspects such as medical secrecy, data protection, problems associated with informed consent, and predictive/prenatal diagnosis.

Homburg/Saar
July 2014

Andreas Keller
Eckart Meese

1
Next-Generation Sequencing for Clinical Diagnostics of Cardiomyopathies

Jan Haas, Hugo A. Katus, and Benjamin Meder

1.1
Introduction

The vast progress next-generation sequencing (NGS) has undergone during the past few years [1,2] has opened doors for a more advanced genetic diagnostic for many inherited diseases, such as Miller syndrome or Charcot–Marie–Tooth neuropathy [3,4]. Here, we want to describe the paradigm change in genetic diagnostics using the example of cardiomyopathies.

1.2
Cardiomyopathies and Why Genetic Testing is Needed

Cardiomyopathies are a heterogeneous group of cardiac diseases that can either be acquired through, for example, inflammation (myocarditis), be stress-induced (tako-tsubo), or be due to a genetic cause [5,6]. Examples of genetic forms are hypertrophic cardiomyopathy (HCM), dilated cardiomyopathy (DCM), arrhythmogenic right ventricular cardiomyopathy, and left-ventricular non-compaction cardiomyopathy. Together with the channellopathies, such as long-QT syndrome and Brugada syndrome, they account for the most common heart diseases and belong to the most prevalent causes of premature death in western civilizations [7,8]. A point mutation in exon 13 of the β-myosin heavy chain gene was the first detected mutation diagnosed to be relevant for HCM in 1990 [9]. Driven by this finding, genetic research has progressed tremendously over the past two decades. Mutations in genes coding for a diverse set of proteins (e.g., sarcomeric, cytoskeletal, desmosomal, channel and channel-associated, membrane, and nuclear proteins, but also mitochondrial proteins or proteins relevant for mRNA splicing) have now been found to be implicated in disease onset and progression [10]. With currently more than 90 known disease genes with more than 1000 exons and multiple malign mutations per gene, the disease's heterogeneity is high and poses a challenge for classical Sanger-based sequencing. Although Sanger sequencing is able to detect mutations by testing only the

Nucleic Acids as Molecular Diagnostics, First Edition. Edited by Andreas Keller and Eckart Meese.

most heavily affected genes, such as the β-myosin heavy chain gene (*MYH7*), where it is possible to find mutations in up to 30–50% of HCM patients, the mutation frequency in most genes is very low [10–12]. Therefore, new methods were needed to further improve genetic diagnostics in cardiomyopathy patients.

1.3 NGS

In contrast to Sanger sequencing, which is only capable of sequencing a few megabases, NGS is able to sequence hundreds of gigabases per run [13–15]. Currently, the most widely used NGS systems are the "sequencing by synthesis"-based sequencer HiSeq 2000 (Illumina), the "ligation and two-base coding"-based system SOLIDv4 (Life technologies), and the 454 GS FLX (Roche), which relies on "pyrosequencing" technology [7,16]. A detailed comparison of currently used systems including performance benchmarks, such as read lengths and output amounts, has been published recently by Liu *et al.* [1,2,17]. In addition to the mature NGS systems, so-called benchtop sequencers have emerged. Those instruments, such as the MiSeq (Illumina), the Ion Torrent (PGM), or the GS Junior (Roche), benefit from a significantly shorter run time (hours compared to days) and a lower price, taking into account a reduced amount of sequenced bases [17,18].

Originally, NGS was designed to sequence whole genomes. In order to reduce costs, methods for target enrichment were developed to restrict sequencing to the regions of interest only [19–22]. The sequencing of only selected segments of the genome allowed the sequencing of a larger number of individuals per run [23]. For the enrichment of the target regions, array-based, in-solution based, or polymerase chain reaction (PCR)-based approaches exist [7,24]. Depending on the sequence composition of the target region in terms of, for example, GC content or sequence heterogeneity, the efficiency of the different methods might vary [25]. In recent years, the number of NGS applications has grown considerably. In addition to the target-enrichment methods, which can be either used for custom gene panels or whole-exome sequencing (WES), RNA-Seq methods to study messenger RNA as well as microRNA are now also routinely used. Furthermore, Methyl-Seq or Chip-Seq methods to study DNA methylation are frequently used. It can be assumed that the number of applications will grow further together with the need for more advanced bioinformatics analysis tools [26]. Here, a shift in the cost distribution from sequencing to downstream analysis is expected [27].

1.4 NGS for Cardiomyopathies

As mentioned above, whole-genome sequencing (WGS) is an unbiased approach to determine the exact order of every base in a studied genome. In the case of patient studies, the human genome, consisting of more than 3 billion bases,

needs to be analyzed. Although costs have been decreasing dramatically (http://www.genome.gov/sequencingcosts), it is still too expensive to perform WGS with a sufficient coverage for routine diagnostics in cardiomyopathies. Downstream bioinformatics analyses are also much more demanding for WGS compared with WES or partial-exome sequencing (PES), which is preferred to be used instead. Whereas WES is mainly used to discover new disease genes, PES has started to become a standard approach for high-throughput testing of multiple genes and patients. Meder and Haas, for example, applied an array-based enrichment of 47 genes (0.27 Mb) on cardiomyopathy patients, and were able to identify disease-causing mutations in both HCM (80%) and DCM (40%) patients [28]. A similar smaller-scaled array-based approach was used by Mook *et al.* to study 23 genes in cardiomyopathy patients [29]. PCR-based, filter-based, and in-solution-based methods have also successfully been applied in combination with the different NGS systems mentioned above to study cardiomyopathy-relevant genomic loci by NGS [30–34]. Haas *et al.* for example were able to for the first time show the mutational landscape for DCM across a large european cohort of 639 patients using in solution based target enrichment (Haas *et al.* EUR HEAR J. 2014). These studies show the feasibility of using NGS in a clinical environment, but also show the diversity of currently used approaches. Although studies exist which claim NGS to be ready as a stand-alone diagnostic test [35], most centers still rely on Sanger validation of the relevant NGS variants, which are still mainly produced within research projects and are not yet part of the daily routine. Although initial guidelines for clinical NGS testing exist [36,37], it is still difficult to compare results between different centers.

1.5 Sample Preparation

Well-established protocols exist for sample preparation that enable technicians to reproducibly prepare high-quality sequencing libraries, if high-quality DNA is used. Therefore, an initial quality check of the DNA through, for example, Bioanalyzer (Agilent) or Qubit (Invitrogen) measurements is inevitable. Library preparation protocols have been improved tremendously and now only require a few nanograms of input DNA compared with the several micrograms that were needed not so long ago. Also, the time needed for sample preparation has been shortened from a couple of days to a few hours, making it possible to finish preparation in a single working day, achieved, for example, with the Haloplex system (Agilent). Recently, disease-specific enrichment assays were introduced by Haloplex, including an optimized predesigned cardiomyopathy and arrhythmia panel, removing the need for manual target region design. Such panels already exist for cancer, for example, and are expected to be developed for other diseases. Another important aspect in the course of sample preparation is the tracking of the samples to guarantee sample integrity when used in a high-throughput manner. Use of a laboratory information management system

(LIMS) is desired. Here, freely as well as commercially available tools exist that help to reduce manual intervention and lower the overall turnaround time [38,39].

1.6 Bioinformatics Analysis Pipeline

Nowadays, the decreasing costs per base enable researchers to sequence larger target regions, exomes or genomes at a higher depth. However, those high-quality gigabase-scale datasets pose an immense challenge for downstream analyses. One major problem for comparison of results is the variety of mainly "homegrown" analysis strategies that have been developed at individual sites. Briefly, they consist of mapping, variant calling, annotation, filtering, and validation of selected variants. Although most approaches rely on similar strategies for filtering (e.g., filtering variants present in databases like dbSNP or the 1000 Genome Project) caution has to be taken. Andreasen *et al.* showed, for example, that 14% of HCM and 17% of DCM previously disease-causing reported missense and nonsense variants are present within the National Heart, Lung and Blood Institute "Grand Opportunity" Exome Sequencing Project (GO-ESP) cohort, which contains exome data from 6500 individuals [40]. Depending on the chosen filters and also due to differences in pipeline tools, tool combinations, or even versions of the programs, this will lead to a low concordance among the analyses [41]. Despite those drawbacks, it is expected that these hurdles will be overcome by newly developed, improved algorithms. Growing sequencing quality and performance of analysis tools will contribute to provide reliable variant calls, with no need for validation to gain acceptable clinical sensitivity, specificity, and positive (negative) predictive values ready for clinical use in the near future. Such a "routine use" requires an existing infrastructure in both the wet-lab and bioinformatics.

Overall, costs are still high to fully equip a laboratory for NGS-based genetic testing [1]. Benchtop sequencers may become a cheaper solution for some diseases, depending on the required sequencing capacity to adequately cover the target region. Due to the discussed differences in NGS techniques and analysis tools, no ultimate cutoff rule for base coverage exists. However a minimum coverage of 30 times seems to provide genuine results and is applied as a standard for many laboratories [36]. Researchers have to calculate the necessary sequencing capacity for their desired coverage based on their target region and decide which NGS system fits best.

1.7 Interpretation of Results and Translation into Clinical Practice

Apart from any limitation mentioned above, the translation of NGS into daily genetic testing of cardiomyopathies is mainly hindered by the restraints

physicians have in interpretation of the finally reported annotated variants. Disease mutation databases like the Human Genome Mutation Database (http://www.biobase-international.com/product/hgmd) help to identify known mutations and can give a hint to their contribution to disease onset or progression. However, for cardiomyopathies, it is expected that many cases (sporadic and familial) are caused by very rare or private variants. Thus, finding a new disease mutation is like finding a needle in a haystack. Whereas nonsense mutations in a known disease gene are expected to cause a disease, others may be reported as "variants of unknown significance" and need to be investigated further. One way of dissecting the possible influence a variant has on protein function is to apply prediction algorithms. A variety of stand-alone as well as Web-based tools exist [42–50]. Their calculations are based on, for example, conservation of the amino acid, transition frequencies, domain profiles, structural positions, or post-translational modifications. Most of them are trained with certain sets of variants. Depending on the type of tested amino acid (e.g., charging state), the tools all perform differently in terms of sensitivity [51,52]. The degree of uniformity among the tools also differs and some tools tend to predict much more damaging variants then others [53]. Despite of the limitations, these tools are valuable to filter out possibly benign variants and restrict validation to only a smaller candidate set, which should be studied in more depth. Such a further investigation comprises, for example, segregation analysis and functional assays in animal models. For cardiomyopathies, the zebrafish has been proven to be an excellent model organism to test a gene/mutation function by knockdown or overexpression analyses [54]. We have also used the zebrafish in the previously mentioned study to validate novel disease variants. With this approach we were able to prove the malignancy of a variant on zebrafish heart function together with a co-segregation in the index patient's family [28]. Another important method to reveal a variant's effect is genotype–phenotype correlational analyses. Lopes *et al.* used such an approach to compare rare non-synonymous single nucleotide polymorphism (nsSNPs), found in an enrichment-based NGS study on HCM patients, against a whole-exome control and were able to explain 13–53% of HCM cases by studying four sarcomeric genes, which showed a significant excess of rare nsSNPs in the HCM cohort [55]. Van de Meerakker used linkage and haplotype analysis on PES of a DCM family in combination with conservation analyses and functional prediction before they could functionally verify an effect of the studied *TPM1* variant on the binding capacity to its actin partner [56]. These studies exemplify the potential NGS-based studies have to identify (new) disease variants (genes). However, such long-term studies do not have the potential to be of immediate benefit for patients. To ultimately translate all the findings into clinical care, physicians need a detailed report to judge the relevance of the identified known and novel variants in relation to the patient's phenotype. A clearly structured more condensed version of the report should then be given to the patient to explain the diagnosis and subsequent therapy planning. A report should follow general principals of clinical genetic reporting and adhere to widely accepted guidelines for variant descriptions from the

Genome Variation Society (www.hgvs.org) [36]. The generation of such a report still requires a lot of manual work. Automated solutions like the knoSYS100 system from Knome (www.knome.com) have begun to emerge on the market. Similar "homegrown" solutions also already exist at some clinics, but those systems will have to evolve further to be regularly implemented. Currently, only genetic testing of genomic variants is performed. As mentioned above, data from the transcriptome and methylome can also be accessed quickly through NGS technology. In the future it might therefore be reasonable to integrate further molecular data into the course of diagnostics. If this is the case, the complexity will increase and reporting meaningful results will become even more difficult.

In summary, before a clinic decides on how to implement NGS into diagnostics, the following points should be considered. Will NGS be implemented at the local site or will it be used through service providers? If the latter is the case, how can data security of transferred genomic information be guaranteed? If a sequencing facility is installed at the local site, the required bioinformatics hardware has to grow with the increasing amount of data that will be produced [57]. This includes both computational power as well as storage capacity, which should be secure but also quickly accessible within the data analysis pipeline [58]. Enough long-term storage should be taken into account. Currently, the amount of NGS data produced is growing faster than computational power is expected to grow based on Moore's law [59]. To close this gap it is possible to use cloud-based solutions from commercial vendors like the "Amazon Elastic Compute Cloud" (www.aws.amazon.com/ec2) as a computational resource. However, as mentioned above, data security as well as ethical aspects have to be considered before its implementation.

In conclusion, NGS-based diagnostics for cardiomyopathies and other inherited diseases are now technically feasible. A careful weighing of the points discussed will help to successfully implement such diagnostics in routine use in the near future to guide more personalized diagnosis and therapy planning.

References

1 Liu, L. *et al.* (2012) Comparison of next-generation sequencing systems. *Journal of Biomedicine & Biotechnology*, **2012**, 251364.

2 Mestan, K.K., Ilkhanoff, L., Mouli, S., and Lin, S. (2011) Genomic sequencing in clinical trials. *Journal of Translational Medicine*, **9**, 222.

3 Lupski, J.R. *et al.* (2010) Whole-genome sequencing in a patient with Charcot–Marie–Tooth neuropathy. *New England Journal of Medicine*, **362**, 1181–1191.

4 Ng, S.B. *et al.* (2010) Exome sequencing identifies the cause of a Mendelian disorder. *Nature Genetics*, **42**, 30–35.

5 Maron, B.J. *et al.* (2006) Contemporary definitions and classification of the cardiomyopathies: an American Heart Association Scientific Statement from the Council on Clinical Cardiology, Heart Failure and Transplantation Committee; Quality of Care and Outcomes Research and Functional Genomics and Translational Biology Interdisciplinary Working Groups; and Council on Epidemiology and Prevention. *Circulation*, **113**, 1807–1816.

6 Meder, B., Katus, H., and Rottbauer, W. (2009) Genetik der hypertrophischen Kardiomyopathie. *Journal für Kardiologie*, **16**, 274–278.

7 Haas, J., Katus, H.A., and Meder, B. (2011) Next-generation sequencing entering the clinical arena. *Molecular and Cellular Probes*, **25**, 206–211.

8 Watkins, H., Ashrafian, H., and Redwood, C. (2011) Inherited cardiomyopathies. *New England Journal of Medicine*, **364**, 1643–1656.

9 Geisterfer-Lowrance, A.A. *et al.* (1990) A molecular basis for familial hypertrophic cardiomyopathy: a beta cardiac myosin heavy chain gene missense mutation. *Cell*, **62**, 999–1006.

10 Wilde, A.A. and Behr, E.R. (2013) Genetic testing for inherited cardiac disease. *Nature Reviews Cardiology*, **10**, 571–583.

11 Perrot, A. *et al.* (2005) Prevalence of cardiac beta-myosin heavy chain gene mutations in patients with hypertrophic cardiomyopathy. *Journal of Molecular Medicine*, **83**, 468–477.

12 Kaski, J.P. *et al.* (2009) Prevalence of sarcomere protein gene mutations in preadolescent children with hypertrophic cardiomyopathy. *Circulation: Cardiovascular Genetics*, **2**, 436–441.

13 Shendure, J. *et al.* (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*, **309**, 1728–1732.

14 Margulies, M. *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376–380.

15 Harris, T.D. *et al.* (2008) Single-molecule DNA sequencing of a viral genome. *Science*, **320**, 106–109.

16 Metzker, M.L. (2010) Sequencing technologies – the next generation. *Nature Reviews Genetics*, **11**, 31–46.

17 Quail, M.A. *et al.* (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, **13**, 341.

18 Loman, N.J. *et al.* (2012) Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology*, **30**, 434–439.

19 Albert, T.J. *et al.* (2007) Direct selection of human genomic loci by microarray hybridization. *Nature Methods*, **4**, 903–905.

20 Hodges, E. *et al.* (2007) Genome-wide *in situ* exon capture for selective resequencing. *Nature Genetics*, **39**, 1522–1527.

21 Okou, D.T. *et al.* (2007) Microarray-based genomic selection for high-throughput resequencing. *Nature Methods*, **4**, 907–909.

22 Porreca, G.J. *et al.* (2007) Multiplex amplification of large sets of human exons. *Nature Methods*, **4**, 931–936.

23 Stratton, M. (2008) Genome resequencing and genetic variation. *Nature Biotechnology*, **26**, 65–66.

24 Mamanova, L. *et al.* (2010) Target-enrichment strategies for next-generation sequencing. *Nature Methods*, **7**, 111–118.

25 Bodi, K. *et al.* (2013) Comparison of commercially available target enrichment methods for next-generation sequencing. *Journal of Biomolecular Techniques*, **24**, 73–86.

26 Shendure, J. and Lieberman Aiden, E. (2012) The expanding scope of DNA sequencing. *Nature Biotechnology*, **30**, 1084–1094.

27 Sboner, A., Mu, X.J., Greenbaum, D., Auerbach, R.K., and Gerstein, M.B. (2011) The real cost of sequencing: higher than you think! *Genome Biology*, **12**, 125.

28 Meder, B. *et al.* (2011) Targeted next-generation sequencing for the molecular genetic diagnostics of cardiomyopathies. *Circulation: Cardiovascular Genetics*, **4**, 110–122.

29 Mook, O.R. *et al.* (2013) Targeted sequence capture and GS-FLX Titanium sequencing of 23 hypertrophic and dilated cardiomyopathy genes: implementation into diagnostics. *Journal of Medical Genetics*, **50**, 614–626.

30 Dames, S., Durtschi, J., Geiersbach, K., Stephens, J., and Voelkerding, K.V. (2010) Comparison of the Illumina Genome Analyzer and Roche 454 GS FLX for resequencing of hypertrophic cardiomyopathy-associated genes. *Journal of Biomolecular Techniques*, **21**, 73–80.

31 Herman, D.S. *et al.* (2012) Truncations of titin causing dilated cardiomyopathy. *New England Journal of Medicine*, **366**, 619–628.

32 Ware, J.S. *et al.* (2013) Next generation diagnostics in inherited arrhythmia syndromes: a comparison of two approaches. *Journal of Cardiovascular Translational Research*, **6**, 94–103.

33 Haas, J., Barb, I., Katus, H.A., and Meder, B. (2014, in press) Targeted next-generation sequencing - the clinician's stethoscope for genetic disorders. *Biom. Med.*

34 Haas, J., Frese, K.S., Peil, B., Kloos, W., Keller, A., Nietsch, R., Feng, Z., Müller, S., Kayvanpour, E., Vogel, B., Sedaghat-Hamedani, F., Lim, W.K., Zhao, X., Fradkin, D., Köhler, D., Fischer, S., Franke, J., Marquart, S., Barb, I., Amr, A., Ehlermann, P., Mereles, D., Weis, T., Kremer, A., King, V., Wirsz, E., Isnard, R., Komajda, M., Serio, A., Grasso, M., Syrris, P., Wicks, E., Plagnol, V., Lopes, L., Gadgaard, T., Eiskjær, H., Jørgensen, M., Garcia-Giustiniani, D., Ortiz-Genga, M., Crespo-Leiro, M.G., Lekanne Dit Deprez, R.H., Christiaans, I., van Rijsingen, I.A., Wilde, A.A., Waldenstrom, A., Bolognesi, M., Bellazzi, R., Mörner, S., Lorenzo Bermejo, J., Monserrat, L., Villard, E., Mogensen, J., Pinto, Y.M., Charron, P., Elliott, P., Arbustini, E., Katus, H.A., and Meder, B. (2014, in press) Atlas of the clinical genetics of human dilated cardiomyopathy. *Eur Heart J.*

35 Sikkema-Raddatz, B. *et al.* (2013) Targeted next-generation sequencing can replace Sanger sequencing in clinical diagnostics. *Human Mutation*, **34**, 1035–1042.

36 Weiss, M.M. *et al.* (2013) Best practice guidelines for the use of next generation sequencing (NGS) applications in genome diagnostics: a national collaborative study of Dutch genome diagnostic laboratories. *Human Mutation*, **34**, 1313–1321.

37 Rehm, H.L. *et al.* (2013) ACMG clinical laboratory standards for next-generation sequencing. *Genetics in Medicine*, **15**, 733–747.

38 Scholtalbers, J. *et al.* (2013) Galaxy LIMS for next-generation sequencing. *Bioinformatics*, **29**, 1233–1234.

39 GenoLogis (2013) Clarity LIMS. GenoLogics Life Sciences Software, Victoria, BC.

40 Andreasen, C. *et al.* (2013) New population-based exome data are questioning the pathogenicity of previously cardiomyopathy-associated genetic variants. *European Journal of Human Genetics*, **21**, 918–928.

41 O'Rawe, J. *et al.* (2013) Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. *Genome Medicine*, **5**, 28.

42 Adzhubei, I.A. *et al.* (2010) A method and server for predicting damaging missense mutations. *Nature Methods*, **7**, 248–249.

43 Ng, P.C. and Henikoff, S. (2003) SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Research*, **31**, 3812–3814.

44 Kumar, P., Henikoff, S., and Ng, P.C. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, **4**, 1073–1081.

45 Reva, B., Antipin, Y., and Sander, C. (2011) Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Research*, **39**, e118.

46 Gonzalez-Perez, A. and Lopez-Bigas, N. (2011) Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *American Journal of Human Genetics*, **88**, 440–449.

47 Carter, H. *et al.* (2009) Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations. *Cancer Research*, **69**, 6660–6667.

48 Choi, Y., Sims, G.E., Murphy, S., Miller, J.R., and Chan, A.P. (2012) Predicting the functional effect of amino acid substitutions and indels. *PLoS One*, **7**, e46688.

49 McLaren, W. *et al.* (2010) Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics*, **26**, 2069–2070.

50 Dayem Ullah, A.Z., Lemoine, N.R., and Chelala, C. (2012) SNPnexus: a web server for functional annotation of novel and publicly known genetic variants (2012 update). *Nucleic Acids Research*, **40**, W65–W70.

51 Thusberg, J., Olatubosun, A., and Vihinen, M. (2011) Performance of mutation pathogenicity prediction methods on missense variants. *Human Mutation*, **32**, 358–368.

52 Thusberg, J. and Vihinen, M. (2009) Pathogenic or not? And if so, then how? Studying the effects of missense mutations using bioinformatics methods. *Human Mutation*, **30**, 703–714.

53 Castellana, S. and Mazza, T. (2013) Congruency in the prediction of pathogenic missense mutations: state-of-the-art web-based tools. *Briefings in Bioinformatics*, **14**, 448–459.

54 Meder, B. *et al.* (2009) A single serine in the carboxyl terminus of cardiac essential myosin light chain-1 controls cardiomyocyte contractility *in vivo*. *Circulation Research*, **104**, 650–659.

55 Lopes, L.R. *et al.* (2013) Genetic complexity in hypertrophic cardiomyopathy revealed by high-throughput sequencing. *Journal of Medical Genetics*, **50**, 228–239.

56 van deMeerakker, J.B. *et al.* (2013) A novel alpha-tropomyosin mutation associates with dilated and non-compaction cardiomyopathy and diminishes actin binding. *Biochimica et Biophysica Acta*, **1833**, 833–839.

57 Baker, M. (2010) Next-generation sequencing: adjusting to data overload. *Nature Methods*, **7**, 495–499.

58 Richter, B.G. and Sexton, D.P. (2009) Managing and analyzing next-generation sequence data. *PLoS Computational Biology*, **5**, e1000369.

59 Moore, G.E. (1965) Cramming more components onto integrated circuits. *Electronics*, **38**, 114–117.

2
MicroRNAs as Novel Biomarkers in Cardiovascular Medicine

Britta Vogel, Hugo A. Katus, and Benjamin Meder

2.1
Introduction

Cardiovascular diseases (CVDs) are among the predominant causes of morbidity and mortality in western countries. In general, early diagnosis and treatment have been shown to be crucial to improve patient outcome. Hence, current biomarkers are indispensible for the diagnostic and therapeutic regimens and largely improved survival of patients with heart failure and acute coronary syndromes (ACSs). As therapeutic options are often applied in a stepwise manner, correlated with distinct disease stages, biomarkers also help to limit often expensive treatment options to appropriate cases and reduce side-effects in patients where the treatment is not expected to be efficient. To further improve the outcome of patients with CVD, the identification, evaluation, and characterization of new biomarkers are the main topics of current research.

Over recent years it has been shown that there are significant alterations of the human miRNome in response to certain physiological and pathological stimuli, such as cardiac hypertrophy, arrhythmias, or ischemia. It is of note that dysregulation in microRNA (miRNA) [1] expression profiles was found to be closely linked to the underlying cause or pathologic pathway activated, allowing discrimination of different disease entities [2]. Such discriminative miRNAs can be extracted and analyzed from tissues and body fluids (e.g., whole blood, serum and plasma, circulating cells, or saliva). By being bound to specific proteins (e.g., Argonaute 2 and lipoproteins) or packed in cellular vesicles (e.g., microvesicles and exosomes) they even avoid degradation by endogenous RNase activity [3–7]. Further, miRNAs are extraordinary stable as analytes even under long-term storage and multiple freeze–thaw cycles [8,9]. Hence, being easily accessible and stable within the specimen, miRNAs meet important pre-analytic criteria and promise robust association with specific diseases or disease stages.

Nucleic Acids as Molecular Diagnostics, First Edition. Edited by Andreas Keller and Eckart Meese.

2.2 miRNAs are Associated with Cardiovascular Risk Factors

The risk of an individual patient developing atherosclerotic plaques and coronary artery disease can be estimated on the basis of established cardiovascular risk factors and composite scores. Whether miRNAs are also associated with the individual cardiovascular risk and whether they might aid current models is not yet understood.

For arterial hypertension, there are data from a plasma-based analysis comparing 127 high-blood pressure patients with 67 normotensive individuals [10]. They identified miRNAs let7e, miR-296-5p, and UL112 to be differently expressed, with the latter being encoded by human cytomegalovirus (CMV). As they also found elevated CMV titers in the hypertensive patients, the authors speculate that this might indicate a link between arterial hypertension and CMV. In a study by Robertson *et al.* [11], miR-24 was found to be part of the regulation process of aldosterone and cortisol synthesis. Through this pathway, miR-24 might play a role in the development of arterial hypertension. Similarly, diabetes patients suffer from an increased level of circulating glucose leading to progressive damage of the vessel integrity, and hence micro- and macrovascular disease. A subset of studies tried to further elucidate the role of miRNAs in the onset and progression of diabetes. Zampetaki *et al.* identified 12 candidate miRNAs to be associated with diabetes and were able to validate their findings in a cohort of 822 diabetic patients for miR-126 [12]. Using a signature of five candidate miRNAs (miR-126, miR-15a, miR-320, miR-223, and miR-28-3p), they could discriminate between diabetes and non-diabetes with a sensitivity of 70% and specificity of 92%. Relevant for risk prediction, the levels of four of these miRNAs were altered even before disease manifestation. In an independent study, different miRNAs were found to be stage-dependently dysregulated in whole-blood samples from diabetic patients [13]. Interestingly, three of the miRNAs found were also identified as being dysregulated in response to diabetes in plasma samples from diabetic compared with non-diabetic patients (miR-29a, miR-30d, and miR-146a) [14]. Another important risk factor for CVD is aging. Boon *et al.* evaluated the role of miRNAs during aging in a mouse model, comparing aged with young mice [15]. They found that altered miRNA expression contributed to the age-related decline in cardiac function, and identified miR-34a as a potential target to reduce cardiomyocyte death and fibrotic remodeling after acute myocardial infarction (AMI) by diminishing telomere attrition. Hence, miRNAs impact on aging and cardiac function of the aging heart, with miR-34a being one of the most recently discovered candidates. The association of miRNAs with the remaining cardiovascular risk factors is less understood. However, there are reports on cigarette smoking and cholesterol metabolism in smaller observational studies [16]. A recent study by Ramirez *et al.*, for example, showed that miR-144 controls high-density lipoprotein levels in plasma by regulating its biogenesis in the liver [17]. The authors assumed that miR-144 could serve additionally as a potential therapeutic target in the treatment of

Table 2.1 miRNAs associated with cardiovascular risk factors.

miRNAs	Detected in	Associated with	Reference
miR-let7e, miR-296-5p, miR-UL112, hcmv-miR-24	plasma of 127 hypertensive and 67 normotensive patients	arterial hypertension	10
miR-24	human aldosterone-producing adenoma and non-diseased adrenal samples	arterial hypertension	11
miR-126, miR-15a, miR-320, miR-223, miR-28-3p, miR-205, miR-21, miR-24, miR-191, miR-197, miR-486	plasma of 80 diabetic patients and 80 controls; validation of miR-126 in 822 additional diabetic patients	diabetes mellitus	12
miR-144, miR-146a, miR-150, miR-182, miR-192, miR-29a, miR-30d, miR-320a	rat model: blood and tissue samples (pancreas, liver, adipose, skeletal muscle)	diabetes mellitus	13
miR-34a, miR-9, miR-29a, miR-30d, miR-124a, miR-146a, miR-375	serum of 56 diabetic patients	diabetes mellitus	14
miR-34a	mouse model	age	15
miR-144	mouse model	dyslipidemia	17

dyslipidemia, and hence the onset and progression of atherosclerotic vascular disease. Table 2.1 summarizes the miRNAs, the specimens used for their detection, and their association with classical risk factors.

2.3 miRNAs in Coronary Artery Disease

Coronary artery disease (CAD) is one of the most frequent cardiovascular disorders. It is characterized by the formation of atherosclerotic plaques in the coronary arteries, whereas atherosclerosis in most cases not only affects cardiac, but also non-cardiac vessels, leading to a high comorbidity with regard to peripheral artery disease. The formation of atherosclerotic plaques and CAD involves different cell types, starting from dysfunction of the endothelium, inflammatory responses by leucocytes, and proliferation of smooth muscle cells. miRNAs play a crucial role in this complex pathologic process, both actively and passively. Measuring the resulting changes in miRNA abundance could hence serve as a diagnostic tool for CAD.

In the course of the disease, plaque progression might lead to narrowing of the coronary vessel lumen, and consequently impair blood flow and oxygen supply of the corresponding myocardium. Revascularization is then an option to permanently relieve symptoms in patients with typical angina pectoris and improve

outcome. Other atherosclerotic plaques, which are referred to as "unstable," might at some point rupture and lead to acute thrombus formation, subsequent vessel occlusion, and hence AMI. To date, there is no reliable tool to discriminate unstable from stable atherosclerotic plaques, whereas the identification and treatment of such "lesions at risk" could potentially reduce cardiovascular events. A biomarker indicating underlying, maybe asymptomatic CAD, correlating with disease severity and most importantly identifying patients at an increased risk for plaque rupture and cardiac ischemia would tremendously facilitate the therapeutic decisions.

The first studies investigating miRNAs derived from whole-blood samples, plasma, serum, and peripheral blood mononuclear cells (PBMCs) point to a role of miRNAs as novel biomarkers in these patients. In a study analyzing whole-blood samples of 12 CAD and 12 control subjects using microarrays, miR-140-3p and miR-182 were found to be upregulated in response to CAD [18]. The authors extended their study by also investigating miRNA changes after an exercise-driven rehabilitation program following coronary bypass surgery and found miR-92 to be significantly increased. In another whole-blood based miRNA study, Weber *et al.* compared expression levels of candidate miRNAs in 10 CAD patients with 15 healthy controls using real-time polymerase chain reaction (PCR) [19]. They found 11 miRNAs to be reduced in CAD patients (miR-19a, miR-484, miR-155, miR-222, miR-145, miR-29a, miR-378, miR-342, miR-181d, miR-150, and miR-30e-5p), although it is of note that medication with angiotensin-converting enzyme (ACE) inhibitors might have had an influence on miRNA abundance. Further promising data on circulating miRNAs have originated from several studies investigating human serum and plasma. Fichtlscherer *et al.* assessed the expression of eight selected miRNAs in plasma of 36 patients with stable CAD and 17 controls, and found endothelial-derived (miR-126, miR-17, and miR-92a), smooth-muscle-cell-originated (miR-145), as well as inflammatory (miR-155) miRNAs to be downregulated in response to CAD [20]. Interestingly, known cardiac miRNAs (miR-133 and miR-208a) were shown to be increased in CAD patients, potentially reflecting clinically unapparent, recurrent cardiac ischemia. In yet another study, by Taurino *et al.*, miR-92 was shown to be upregulated in patients after bypass surgery and exercise-driven rehabilitation, although the reason for the somewhat conflicting results is not clear [18]. There is also evidence that miRNAs derived from platelets can indicate CAD. In a recent study, a miRNA signature comprising four miRNAs (miR-340*, miR-624*, miR-451, and miR-454) could discriminate premature CAD from controls with high sensitivity and specificity, resulting in an area under the curve (AUC) of 0.71 in receiver operating characteristic (ROC) analysis [21].

As described, the diagnostic potential of miRNAs in identifying CAD patients at an increased risk for cardiovascular events and disease progression is of major interest. The first promising data on this come from a study by Hoekstra *et al.* evaluating miRNAs derived from PBMCs by real-time PCR-based microarrays [22]. They compared 25 patients with stable angina pectoris to 25 patients with unstable angina, as well as 20 controls. They found miR-135 and miR-147 to be

Table 2.2 miRNAs associated with CAD.

miRNAs	Detected in	Reference
miR-140-3p, miR-182, miR-92a, miR-92b	whole-blood samples of 12 CAD patients and 12 controls, as well as 10 additional CAD patients after surgical coronary revascularization	18
miR-19a, miR-484, miR-155, miR-222, miR-145, miR-29a, miR-378, miR-342, miR-181d, miR-150, miR-30e-5p	whole-blood samples of 10 CAD patients and 15 controls	19
miR-126, miR-17, miR-92a, miR-145, miR-155, miR-133, miR-208a	plasma of 36 CAD patients and 17 controls	20
miR-340*, miR-624*, miR-451, miR-454, miR545, miR615-5p, miR-1280	platelets of 12 patients with premature CAD and 12 controls; validation of miR-340* and miR-624* in 67 CAD patients and 80 controls	21
miR-134, miR-370, miR-198, miR-135, miR-147	PBMCs of 25 patients with unstable and 25 patients with stable angina, as well as 20 controls	22
miR-146a, miR-146b	PBMCs of 66 patients with stable CAD and 33 controls	23

downregulated in both patient groups when compared with healthy individuals. However, most importantly, they also identified miR-134, miR-370, and miR-198 to be increased significantly in PBMCs of unstable angina patients in comparison with those with stable angina pectoris. Supporting the hypothesis that miRNAs can indicate patients with unstable CAD, Takahashi *et al.* showed that miR-146a derived from PBMCs can serve as an independent risk factor for cardiac events [23]. They assessed expression of the inflammatory-related miR-146a and miR-146b in 66 patients with stable CAD, as well as 33 controls, and found both to be upregulated in affected patients. During the 1-year follow-up period, 13 patients experienced a cardiac event, with miR-146a indicating patients at risk.

Taken together, there are now several studies, although mostly on small patient cohorts, that indicate a potential role of miRNAs as biomarkers for CAD. These miRNAs are derived from different blood components with only little overlap, indicating a certain specificity of miRNAs to the investigated sample types. Table 2.2 summarizes CAD-associated miRNAs.

2.4 miRNAs in Cardiac Ischemia and Necrosis

The pathogenesis of AMI is complex and involves a series of events. Contingent rupture of an unstable atherosclerotic plaque is followed by thrombus formation and occlusion of a coronary vessel, leading to ischemia and hence necrosis of the

downstream myocardium. It is conceivable that miRNAs play a role in the different phases of AMI and that they might be released from or changed in any tissues and cells involved during this process.

Today's gold standard biomarkers for cardiac necrosis are cardiac troponins. Troponins, as constituents of cardiac sarcomeres, regulate the excitation–contraction coupling process and are necessary for appropriate muscle function. Troponin I and T are expressed as specific cardiac isoforms and can be assessed after their release from cardiomyocytes in human blood samples with antibody-dependent diagnostic tools [24]. The intrinsic cardiac specificity is of great importance, as former non-cardiac-specific biomarkers often failed to discriminate cardiac from non-cardiac injury, reducing their diagnostic accuracy. As such, creatine kinase or creatine kinase myocardial bound (CK-MB), which served as the gold standard AMI biomarkers until the late 1970s, are today dispensable due to their inferior diagnostic performance compared with troponins. Over the past decades, such assays have been reworked, modified, enhanced, and consequently improved with regard to their analytical performance [25]. Assays with a total imprecision at the 99th percentile of 10% or less and measurable normal values below the 99th percentile in at least 90% of healthy individuals are classified as "high-sensitivity" assays [26]. These high-sensitivity troponin T assays have been shown to be superior to the former-generation assays, especially with regard to diagnostic sensitivity and prediction of risk and outcome [27,28]. Based on strong clinical and experimental evidence, troponins became part of the universal definition of AMI as stated in current guidelines [26]. The release of these proteins follows a time-dependent manner after the initial event, with current high-sensitivity assays allowing early diagnosis and treatment by measuring even very low abundances. This high sensitivity in detecting myocardial damage results in a comparable low specificity for AMI if taken as single values, since troponin levels can also be altered in response to other CVDs, such as pulmonary embolism, myocarditis, acute tachycardia, or heart failure. New biomarkers based on miRNAs could potentially circumvent such limitations and add precision to the diagnosis of ACS.

miRNAs that are known to be expressed in cardiomyocytes served as candidate biomarkers for AMI in a series of studies. Among them is miR-208, which was shown to be already present in plasma of AMI rats 3 h after AMI [29], with confirming results also coming from other animal models [30,31]. Additional muscle-specific/enriched miRNAs, such as miR-1, miR-133a, and miR-499, were also shown to be dysregulated in response to cardiac necrosis in a time-dependent manner in, for example, rats and pigs [32–34]. There are also several studies underlining the diagnostic potential of cardiac- or muscle-specific circulating miRNAs in patients with AMI. D'Allesandra *et al.*, for example, were able to show that miR-1, miR-133a and b, and miR-499-5p are increased during AMI [31]. They compared miRNA expression of 25 AMI patients and 17 controls. However, they also found significant downregulation of miR-122 and miR-375. It is of note that within this study, samples were collected at 9 h after the onset of symptoms. This rather late sample collection might explain why miR-208 was

not (or maybe no longer) detectable in most patients. Similar problems might have occurred in a study by Kuwabara *et al.*, where miR-208 was not detectable in AMI patients [35]. These results again underline the necessity of consistent and early sample collection in the course of AMI. That miR-208 is indeed dysregulated in response to AMI was shown in a study by Wang *et al.*, where it remained undetectable in healthy controls, as well as in patients with stable CAD [30]. Strikingly, the authors observed increased amounts of miR-208 even in those (few) AMI patients where cardiac troponin I was still negative. Its family member miR-208b, as well as miR-499, were also shown to correlate with troponin T [33,36]. Similar results were obtained in a larger cohort of 510 AMI patients and 87 controls. Further insight into the release kinetics of miRNAs in human patients after myocardial injury comes from a recently published study evaluating the kinetic changes of muscle-enriched miRNAs after transcoronary ablation of septal hypertrophy (TASH) – a procedure to treat patients with hypertrophic obstructive cardiomyopathy by iatrogenic induction of a regional myocardial necrosis [37]. The authors found miR-1 and miR-133a to be significantly increased already after 15 min and miR-208a to be elevated 105 min after the procedure. Hence, miRNA expression seems to be altered at a very early stage during AMI.

Viewed retrospectively, the huge progress in diagnosis of myocardial infarction was driven by the development of strictly cardiac-specific markers and their investigation in serum/plasma. For a single marker, this will likely produce the highest sensitivity; however, as discussed above, this results in clinically relevant problems regarding the positive predictive power and specificity. An alternative idea is to use biomarker signatures, which may include hundreds of features such as miRNAs and proteins. For such a composite marker, it will be important to incorporate molecularly independent features to gain biological information. A whole-blood sample may provide additional details about the pathophysiological cascade of AMI. In an early study, our group compared genome-wide miRNA levels from AMI patients with non-infarction controls [38]. We found 121 miRNAs to be significantly dysregulated during AMI, with miR-1291 and miR-663b showing the highest discriminative power. This diagnostic potential of single markers could be enhanced by combining the informational content of different miRNAs, with a signature of 20 miRNAs leading to an AUC of 0.99. Additionally, it could be shown that miR-145 and miR-30c highly correlate with troponin T levels, and hence infarct size, disease severity, and presumably prognosis. Both miRNAs were found repeatedly in successful studies on CAD and ACS/AMI [32, 33, 39]. To investigate the time course of these miRNAs in AMI, Vogel *et al.* conducted a subsequent study on the release kinetics of whole-blood miRNAs [40]. They collected blood samples of AMI patients at the initial presentation to hospital (high-sensitivity troponin T below 50 pg/nl) as well as during the first 24 h after admission and compared them with controls. They found a signature of seven miRNAs (miR-636, miR-7-1*, miR-380*, miR-1254, miR-455-3p, miR-566, and miR-1291) able to discriminate diseased from non-affected individuals with high statistical power, especially during the early stages of AMI.

The correlation of miRNAs with clinical outcome and prognosis is starting to be the focus of newer studies. A study from Widera *et al.* compared candidate miRNA levels from 327 AMI patients with 117 patients with unstable angina pectoris [41]. Individuals were followed for 6 months. miR-133a and miR-208b were shown to significantly correlate with all-cause mortality, although they could not increase the discriminatory power of high-sensitivity troponin T. Very recent data from Matsumoto *et al.* show that miRNAs are associated with the development of ischemic heart failure after myocardial infarction and, presumably, outcome [42]. The authors assessed expression levels of 377 miRNAs in 21 patients who developed heart failure within 1 year after AMI and 65 controls without subsequent cardiovascular events after hospital discharge. Blood samples were collected 18 days after the ischemic event. They found the p53-responsive miR-192, miR-194, and miR-34a to be increased in patients developing ischemic heart failure, indicating that these miRNAs possibly take part in the regulation of heart failure development via the p53 pathway.

In summary, it is now evident that miRNAs are novel biomarkers for AMI (Table 2.3). However, their superiority to gold standard markers needs to be proven, and technical challenges such as appropriate and timely assays need to be addressed.

Table 2.3 miRNAs associated with AMI.

miRNAs	Detected in	Reference
miR-208, miR-1, miR-133a, miR-499	animal models	29–35
miR-1, miR-133a, miR-133b, miR-499-5p, miR-122, miR-375	plasma of 25 AMI patients and 17 controls	31
miR-1, miR-133a	serum of 29 patients with ACS and 42 patients with non-ACS	35
miR-208a, miR-1, miR-133a, miR-499	plasma of 33 AMI patients, 16 patients with non-AMI coronary heart disease, 17 patients with other cardiovascular diseases and 30 controls	30
miR-208b, miR-1, miR-133a, miR-499	plasma of 25 AMI patients	33
miR-208b, miR-499	plasma of 32 AMI patients and 36 controls	36
miR-1, miR-133a, miR-208a	serum of 21 patients after TASH	37
miR-1291, miR-663b, miR-145, miR-30c	whole-blood samples of 20 AMI patients and 20 controls	38
miR-636, miR-7-1*, miR-380*, miR-1254, miR-455-3p, miR-566, miR-1291	whole-blood samples of 18 AMI patients and 21 controls	40
miR-133a, miR-208b	plasma of 327 AMI patients and 117 patients with unstable angina	41
miR-192, miR-194, miR-34a	serum of 86 AMI patients	42

2.5
miRNAs as Biomarkers of Heart Failure

Heart failure is a disease entity that is characterized by the inability of the heart to maintain sufficient oxygen supply to the organs. Different CVDs can result in heart failure, such as inherited cardiomyopathies, severe CAD, exposure to cardiotoxic agents such as chemotherapeutics, valvular heart disease, or myocarditis. In the current guidelines on the management of patients with acute and chronic heart failure [43], biomarkers have gained importance in the diagnosis and estimation of disease severity, with natriuretic peptides as well as troponins being the preferred biomarkers to assess. The increasing need for easily accessible biomarkers is based on the fact that prognosis-improving therapeutic strategies are applied according to disease severity and the estimated cardiovascular risk.

The potential of circulating miRNAs as biomarkers for heart failure was investigated in a recent study by Tijsen *et al.* [44]. The authors compared the expressions of 16 selected miRNAs from the plasma of 39 healthy controls with 50 patients with dyspnea (30 of them due to heart failure and 20 of them due to other causes). They found several miRNAs to be upregulated in response to heart failure, with miR-423-5p as their top candidate. Assuming this miRNA as a single diagnostic marker, they were able to discriminate heart failure from control subjects with an AUC of 0.91 and from other causes of dyspnea with an AUC of 0.83. Expression levels of miR-423-5p also correlated with left ventricular ejection fraction (LVEF) and hence disease severity, as well as symptoms of patients as indicated by the New York Heart Association functional class. In addition, the authors showed correlation of miR-423-5p with NTproBNP (N-terminal of the prohormone brain natriuretic peptide) levels. These promising results on miR-423-5p are supported by Goren *et al.* [45]. However, the authors could not replicate the correlation of miR-423-5p with LVEF.

Different miRNAs were identified as promising new biomarkers for heart failure in whole-blood samples. In a study published recently by our group, a subset of miRNAs was shown to be dysregulated in response to non-ischemic systolic heart failure [46]. Some of these miRNAs have a high diagnostic potential already as single markers (miR-558, miR-122*, and miR-520d-5p), but the statistical power could be improved when combining different miRNAs as a signature (miR-200b*, miR-622, miR-519e*, miR-1231, and miR-1228*), reaching an AUC value of 0.81. Interestingly, we could find a potential prognostic value of miR-519e* that indicates occurrence of a combined endpoint comprising cardiovascular death, rehospitalization due to heart failure, and heart transplantation.

A remaining question is where the miRNAs are actually derived from. First, there is certainly an overlap between miRNAs from whole-blood and serum or plasma samples. miR-423-5p and miR-622, for instance, which are promising serum/plasma heart failure markers, can also be detected in whole-blood samples [44]. Often, this is referred to as convergence of serum and cellular miRNAs, as miRNAs released from tissues can be taken up by different cell types and vice versa. Secondly, the miRNAs may be released from different tissues into the blood

Table 2.4 miRNAs associated with heart failure.

miRNAs	Detected in	Reference
miR-122, miR-208b	ischemic porcine cardiogenic shock model	47
miR-423-5p	plasma of 30 patients with heart failure, 39 controls 20 patients with dyspnea due to other underlying disease	44
miR-423-5p, miR-320a, miR-22, miR-92b	serum of 30 heart failure patients and 30 controls	45
miR-519e*, miR-558, miR-122*, miR-520d-5p, miR-200b*, miR-622, miR-1231, miR-1228*	whole-blood samples of 53 heart failure patients and 39 controls	46
miR-208b, miR-499, miR-1, miR-133a, miR-122	plasma of 33 heart failure patients and 20 controls	36

stream. miR-122, for example, was shown to be enriched in plasma of acute heart failure [36] and its minor form (miR-122*) in whole-blood samples [46]. In a porcine cardiogenic shock model, miR-122 was also found to be upregulated in blood [47] and the authors speculated that this might be due to release from the liver due to insufficient oxygen supply caused by reduced cardiac output.

All of the studies published so far are, from a clinical perspective, based on relatively small patient numbers and have to be considered with caution. Future studies will have to confirm the capability of the described miRNAs (Table 2.4) as heart failure biomarkers in adequately sized cohorts.

2.6 Future Challenges

The field of miRNA biomarker research is rapidly evolving, from first observational studies to well-powered clinical trials. This development points the spotlight on novel questions, such as pre-analytical considerations, technologies to measure miRNAs in a clinical setting, and strategies to easily measure as well as interpret multivariate signatures. In addition to the diagnostic use of miRNAs, future studies are needed to characterize the association of miRNAs with disease progression, individual risk, and prognosis. Finally, miRNAs have to be compared with current diagnostic gold standards to underline their added value.

Acknowledgments

This work was supported by grants from the German Ministry of Education and Research (BMBF): German Centre for Cardiovascular Research (DZHK), the

University of Heidelberg (Innovationsfond FRONTIER) and the European Union (FP7 BestAgeing and INHERITANCE).

References

1 Shi, K.H., Tao, H., Yang, J.J., Wu, J.X., Xu, S.S., and Zhan, HY. (2013) Role of microRNAs in atrial fibrillation: new insights and perspectives. *Cellular Signalling*, **25** (11), 2079–2084. doi: 10.1016/j.cellsig.2013.06.009

2 Keller, A., Leidinger, P., Bauer, A., Elsharawy, A., Haas, J., Backes, C., Wendschlag, A., Giese, N., Tjaden, C., Ott, K., Werner, J., Hackert, T., Ruprecht, K., Huwer, H., Huebers, J., Jacobs, G., Rosenstiel, P., Dommisch, H., Schaefer, A., Muller-Quernheim, J., Wullich, B., Keck, B., Graf, N., Reichrath, J., Vogel, B., Nebel, A., Jager, S.U., Staehler, P., Amarantos, I., Boisguerin, V., Staehler, C., Beier, M., Scheffler, M., Buchler, M.W., Wischhusen, J., Haeusler, S.F., Dietl, J., Hofmann, S., Lenhof, H.P., Schreiber, S., Katus, H.A., Rottbauer, W., Meder, B., Hoheisel, J.D., Franke, A., and Meese, E. (2011) Toward the blood-borne miRNome of human diseases. *Nature Methods*, **8** (10), 841–843. doi: 10.1038/nmeth.1682.

3 Fichtlscherer, S., Zeiher, A.M., and Dimmeler, S. (2011) Circulating microRNAs: biomarkers or mediators of cardiovascular diseases? *Arteriosclerosis, Thrombosis, and Vascular Biology*, **31** (11), 2383–2390. doi: 10.1161/ATVBAHA.111.226696

4 Mitchell, P.S., Parkin, R.K., Kroh, E.M., Fritz, B.R., Wyman, S.K., Pogosova-Agadjanyan, E.L., Peterson, A., Noteboom, J., O'Briant, K.C., Allen, A., Lin, D.W., Urban, N., Drescher, C.W., Knudsen, B.S., Stirewalt, D.L., Gentleman, R., Vessella, R. L., Nelson, P.S., Martin, D.B., and Tewari, M. (2008) Circulating microRNAs as stable blood-based markers for cancer detection. *Proceedings of the National Academy of Sciences of the United States of America*, **105** (30), 10513–10518. doi: 10.1073/pnas.0804549105

5 Valadi, H., Ekstrom, K., Bossios, A., Sjostrand, M., Lee, J.J., and Lotvall, J.O. (2007) Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nature Cell Biology*, **9** (6), 654–659. doi: 10.1038/ncb1596

6 Arroyo, J.D., Chevillet, J.R., Kroh, E.M., Ruf, I.K., Pritchard, C.C., Gibson, D.F., Mitchell, P.S., Bennett, C.F., Pogosova-Agadjanyan, E.L., Stirewalt, D.L., Tait, J.F., and Tewari, M. (2011) Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proceedings of the National Academy of Sciences of the United States of America*, **108** (12), 5003–5008. doi: 10.1073/pnas.1019055108

7 Wang, K., Zhang, S., Weber, J., Baxter, D., and Galas, D.J. (2010) Export of microRNAs and microRNA-protective protein by mammalian cells. *Nucleic Acids Research*, **38** (20), 7248–7259. doi: 10.1093/nar/gkq601

8 Tijsen, A.J., Pinto, Y.M., and Creemers, E. E. (2012) Circulating microRNAs as diagnostic biomarkers for cardiovascular diseases. *American Journal of Physiology Heart and Circulatory Physiology*, **303** (9), H1085–H1095. doi: 10.1152/ajpheart.00191.2012

9 Chen, X., Ba, Y., Ma, L., Cai, X., Yin, Y., Wang, K., Guo, J., Zhang, Y., Chen, J., Guo, X., Li, Q., Li, X., Wang, W., Zhang, Y., Wang, J., Jiang, X., Xiang, Y., Xu, C., Zheng, P., Zhang, J., Li, R., Zhang, H., Shang, X., Gong, T., Ning, G., Wang, J., Zen, K., Zhang, J., and Zhang, C.Y. (2008) Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Research*, **18** (10), 997–1006. doi: 10.1038/cr.2008.282

10 Li, S., Zhu, J., Zhang, W., Chen, Y., Zhang, K., Popescu, L.M., Ma, X., Lau, W.B., Rong, R., Yu, X., Wang, B., Li, Y., Xiao, C., Zhang, M., Wang, S., Yu, L., Chen, A.F., Yang, X., and Cai, J. (2011) Signature

microRNA expression profile of essential hypertension and its novel link to human cytomegalovirus infection. *Circulation*, **124** (2), 175–184. doi: 10.1161/CIRCULATIONAHA.110.012237

11 Robertson, S., Mackenzie, S.M., Alvarez-Madrazo, S., Diver, L.A., Lin, J., Stewart, P. M., Fraser, R., Connell, J.M., and Davies, E. (2013) MicroRNA-24 is a novel regulator of aldosterone and cortisol production in the human adrenal cortex. *Hypertension*, **62** (3), 572–578. doi: 10.1161/HYPERTENSIONAHA.113.01102

12 Zampetaki, A., Kiechl, S., Drozdov, I., Willeit, P., Mayr, U., Prokopi, M., Mayr, A., Weger, S., Oberhollenzer, F., Bonora, E., Shah, A., Willeit, J., and Mayr, M. (2010) Plasma microRNA profiling reveals loss of endothelial miR-126 and other microRNAs in type 2 diabetes. *Circulation Research*, **107** (6), 810–817. doi: 10.1161/CIRCRESAHA.110.226357

13 Karolina, D.S., Armugam, A., Tavintharan, S., Wong, M.T., Lim, S.C., Sum, C.F., and Jeyaseelan, K. (2011) MicroRNA 144 impairs insulin signaling by inhibiting the expression of insulin receptor substrate 1 in type 2 diabetes mellitus. *PloS One*, **6** (8), e22839. doi: 10.1371/journal.pone.0022839

14 Kong, L., Zhu, J., Han, W., Jiang, X., Xu, M., Zhao, Y., Dong, Q., Pang, Z., Guan, Q., Gao, L., Zhao, J., and Zhao, L. (2011) Significance of serum microRNAs in pre-diabetes and newly diagnosed type 2 diabetes: a clinical study. *Acta Diabetologica*, **48** (1), 61–69. doi: 10.1007/s00592-010-0226-0

15 Boon, R.A., Iekushi, K., Lechner, S., Seeger, T., Fischer, A., Heydt, S., Kaluza, D., Treguer, K., Carmona, G., Bonauer, A., Horrevoets, A.J., Didier, N., Girmatsion, Z., Biliczki, P., Ehrlich, J.R., Katus, H.A., Muller, O.J., Potente, M., Zeiher, A.M., Hermeking, H., and Dimmeler, S. (2013) MicroRNA-34a regulates cardiac ageing and function. *Nature*, **495** (7439), 107–110. doi: 10.1038/nature11919

16 Takahashi, K., Yokota, S.I., Tatsumi, N., Fukami, T., Yokoi, T., and Nakajima, M. (2013) Cigarette smoking substantially alters plasma microRNA profiles in healthy subjects. *Toxicology and Applied Pharmacology*, **272** (1), 154–160. doi: 10.1016/j.taap.2013.05.018

17 Ramirez, C.M., Rotllan, N., Vlassov, A.V., Davalos, A., Li, M., Goedeke, L., Aranda, J. F., Cirera-Salinas, D., Araldi, E., Salerno, A., Wanschel, A., Zavadil, J., Castrillo, A., Kim, J., Suarez, Y., and Fernandez-Hernando, C. (2013) Control of cholesterol metabolism and plasma high-density lipoprotein levels by microRNA-144. *Circulation Research*, **112** (12), 1592–1601. doi: 10.1161/CIRCRESAHA.112.300626

18 Taurino, C., Miller, W.H., McBride, M.W., McClure, J.D., Khanin, R., Moreno, M.U., Dymott, J.A., Delles, C., and Dominiczak, A.F. (2010) Gene expression profiling in whole blood of patients with coronary artery disease. *Clinical Science*, **119** (8), 335–343. doi: 10.1042/CS20100043

19 Weber, J.A., Baxter, D.H., Zhang, S., Huang, D.Y., Huang, K.H., Lee, M.J., Galas, D.J., and Wang, K. (2010) The microRNA spectrum in 12 body fluids. *Clinical Chemistry*, **56** (11), 1733–1741. doi: 10.1373/clinchem.2010.147405

20 Fichtlscherer, S., De Rosa, S., Fox, H., Schwietz, T., Fischer, A., Liebetrau, C., Weber, M., Hamm, C.W., Roxe, T., Muller-Ardogan, M., Bonauer, A., Zeiher, A.M., and Dimmeler, S. (2010) Circulating microRNAs in patients with coronary artery disease. *Circulation Research*, **107** (5), 677–684. doi: 10.1161/CIRCRESAHA.109.215566

21 Sondermeijer, B.M., Bakker, A., Halliani, A., de Ronde, M.W., Marquart, A.A., Tijsen, A.J., Mulders, T.A., Kok, M.G., Battjes, S., Maiwald, S., Sivapalaratnam, S., Trip, M.D., Moerland, P.D., Meijers, J.C., Creemers, E.E., and Pinto-Sietsma, S.J. (2011) Platelets in patients with premature coronary artery disease exhibit upregulation of miRNA340* and miRNA624*. *PloS One*, **6** (10), e25946. doi: 10.1371/journal.pone.0025946

22 Hoekstra, M., van der Lans, C.A., Halvorsen, B., Gullestad, L., Kuiper, J., Aukrust, P., van Berkel, T.J., and Biessen, E.A. (2010) The peripheral blood mononuclear cell microRNA signature of coronary artery disease. *Biochemical and Biophysical Research Communications*,

394 (3), 792–797. doi: 10.1016/j.bbrc.2010.03.075

23 Takahashi, Y., Satoh, M., Minami, Y., Tabuchi, T., Itoh, T., and Nakamura, M. (2010) Expression of miR-146a/b is associated with the Toll-like receptor 4 signal in coronary artery disease: effect of renin-angiotensin system blockade and statins on miRNA-146a/b and Toll-like receptor 4 levels. *Clinical Science*, **119** (9), 395–405. doi: 10.1042/CS20100003

24 Katus, H.A., Looser, S., Hallermayer, K., Remppis, A., Scheffold, T., Borgya, A., Essig, U., and Geuss, U. (1992) Development and *in vitro* characterization of a new immunoassay of cardiac troponin T. *Clinical Chemistry*, **38** (3), 386–393.

25 Mueller, M., Vafaie, M., Biener, M., Giannitsis, E., and Katus, H.A. (2013) Cardiac troponin T: from diagnosis of myocardial infarction to cardiovascular risk prediction. *Circulation Journal*, **77** (7), 1653–1661.

26 Thygesen, K., Alpert, J.S., Jaffe, A.S., Simoons, M.L., Chaitman, B.R., White, H. D., Joint ESC/ACCF/AHA/WHF Task Force for the Universal Definition of Myocardial and Clemmensen, P. (2012) Third universal definition of myocardial infarction. *European Heart Journal*, **33** (20), 2551–2567. doi: 10.1093/eurheartj/ehs184

27 Giannitsis, E., Becker, M., Kurz, K., Hess, G., Zdunek, D., and Katus, H.A. (2010) High-sensitivity cardiac troponin T for early prediction of evolving non-ST-segment elevation myocardial infarction in patients with suspected acute coronary syndrome and negative troponin results on admission. *Clinical Chemistry*, **56** (4), 642–650. doi: 10.1373/clinchem.2009.134460

28 Giannitsis, E., Kurz, K., Hallermayer, K., Jarausch, J., Jaffe, A.S., and Katus, H.A. (2010) Analytical validation of a high-sensitivity cardiac troponin T assay. *Clinical Chemistry*, **56** (2), 254–261. doi: 10.1373/clinchem.2009.132654

29 Ji, X., Takahashi, R., Hiura, Y., Hirokawa, G., Fukushima, Y., and Iwai, N. (2009) Plasma miR-208 as a biomarker of myocardial injury. *Clinical Chemistry*, **55** (11), 1944–1949. doi: 10.1373/clinchem.2009.125310

30 Wang, G.K., Zhu, J.Q., Zhang, J.T., Li, Q., Li, Y., He, J., Qin, Y.W., and Jing, Q. (2010) Circulating microRNA: a novel potential biomarker for early diagnosis of acute myocardial infarction in humans. *European Heart Journal*, **31** (6), 659–666. doi: 10.1093/eurheartj/ehq013

31 D'Alessandra, Y., Devanna, P., Limana, F., Straino, S., Di Carlo, A., Brambilla, P.G., Rubino, M., Carena, M.C., Spazzafumo, L., De Simone, M., Micheli, B., Biglioli, P., Achilli, F., Martelli, F., Maggiolini, S., Marenzi, G., Pompilio, G., and Capogrossi, M.C. (2010) Circulating microRNAs are new and sensitive biomarkers of myocardial infarction. *European Heart Journal*, **31** (22), 2765–2773. doi: 10.1093/eurheartj/ehq167

32 Cheng, Y., Tan, N., Yang, J., Liu, X., Cao, X., He, P., Dong, X., Qin, S., and Zhang, C. (2010) A translational study of circulating cell-free microRNA-1 in acute myocardial infarction. *Clinical Science*, **119** (2), 87–95. doi: 10.1042/CS20090645

33 Gidlof, O., Andersson, P., van der Pals, J., Gotberg, M., and Erlinge, D. (2011) Cardiospecific microRNA plasma levels correlate with troponin and cardiac function in patients with ST elevation myocardial infarction, are selectively dependent on renal elimination, and can be detected in urine samples. *Cardiology*, **118** (4), 217–226. doi: 10.1159/000328869

34 Adachi, T., Nakanishi, M., Otsuka, Y., Nishimura, K., Hirokawa, G., Goto, Y., Nonogi, H., and Iwai, N. (2010) Plasma microRNA 499 as a biomarker of acute myocardial infarction. *Clinical Chemistry*, **56** (7), 1183–1185. doi: 10.1373/clinchem.2010.144121

35 Kuwabara, Y., Ono, K., Horie, T., Nishi, H., Nagao, K., Kinoshita, M., Watanabe, S., Baba, O., Kojima, Y., Shizuta, S., Imai, M., Tamura, T., Kita, T., and Kimura, T. (2011) Increased microRNA-1 and microRNA-133a levels in serum of patients with cardiovascular disease indicate myocardial damage. *Circulation Cardiovascular Genetics*, **4** (4), 446–454. doi: 10.1161/CIRCGENETICS.110.958975

36 Corsten, M.F., Dennert, R., Jochems, S., Kuznetsova, T., Devaux, Y., Hofstra, L., Wagner, D.R., Staessen, J.A., Heymans, S., and Schroen, B. (2010) Circulating MicroRNA-208b and MicroRNA-499

reflect myocardial damage in cardiovascular disease. *Circulation Cardiovascular Genetics*, **3** (6), 499–506. doi: 10.1161/CIRCGENETICS.110.957415

37 Liebetrau, C., Mollmann, H., Dorr, O., Szardien, S., Troidl, C., Willmer, M., Voss, S., Gaede, L., Rixe, J., Rolf, A., Hamm, C., and Nef, H. (2013) Release kinetics of circulating muscle-enriched microRNAs in patients undergoing transcoronary ablation of septal hypertrophy. *Journal of the American College of Cardiology*, **62** (11), 992–998. doi: 10.1016/j.jacc.2013.05.025

38 Meder, B., Keller, A., Vogel, B., Haas, J., Sedaghat-Hamedani, F., Kayvanpour, E., Just, S., Borries, A., Rudloff, J., Leidinger, P., Meese, E., Katus, H.A., and Rottbauer, W. (2011) MicroRNA signatures in total peripheral blood as novel biomarkers for acute myocardial infarction. *Basic Research in Cardiology*, **106** (1), 13–23. doi: 10.1007/s00395-010-0123-2

39 Mayorga, M.E. and Penn, M.S. (2012) miR-145 is differentially regulated by TGF-β1 and ischaemia and targets Disabled-2 expression and wnt/β-catenin activity. *Journal of Cellular and Molecular Medicine*, **16** (5), 1106–1113. doi: 10.1111/j.1582-4934.2011.01385

40 Vogel, B., Keller, A., Frese, K.S., Kloos, W., Kayvanpour, E., Sedaghat-Hamedani, F., Hassel, S., Marquart, S., Beier, M., Giannitis, E., Hardt, S., Katus, H.A., and Meder, B. (2013) Refining diagnostic microRNA signatures by whole-miRNome kinetic analysis in acute myocardial infarction. *Clinical Chemistry*, **59** (2), 410–418. doi: 10.1373/clinchem.2011.181370

41 Widera, C., Gupta, S.K., Lorenzen, J.M., Bang, C., Bauersachs, J., Bethmann, K., Kempf, T., Wollert, K.C., and Thum, T. (2011) Diagnostic and prognostic impact of six circulating microRNAs in acute coronary syndrome. *Journal of Molecular and Cellular Cardiology*, **51** (5), 872–875. doi: 10.1016/j.yjmcc.2011.07.011

42 Matsumoto, S., Sakata, Y., Suna, S., Nakatani, D., Usami, M., Hara, M., Kitamura, T., Hamasaki, T., Nanto, S., Kawahara, Y., and Komuro, I. (2013) Circulating p53-responsive microRNAs are predictive indicators of heart failure after acute myocardial infarction. *Circulation Research*, **113** (3), 322–326. doi: 10.1161/CIRCRESAHA.113.301209

43 McMurray, J.J., Adamopoulos, S., Anker, S.D., Auricchio, A., Bohm, M., Dickstein, K., Falk, V., Filippatos, G., Fonseca, C., Gomez-Sanchez, M.A., Jaarsma, T., Kober, L., Lip, G.Y., Maggioni, A.P., Parkhomenko, A., Pieske, B.M., Popescu, B.A., Ronnevik, P.K., Rutten, F.H., Schwitter, J., Seferovic, P., Stepinska, J., Trindade, P.T., Voors, A.A., Zannad, F., and Zeiher, A. (2012) ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2012. The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. *European Journal of Heart Failure*, **14** (8), 803–869. doi: 10.1093/eurjhf/hfs105

44 Tijsen, A.J., Creemers, E.E., Moerland, P.D., de Windt, L.J., van der Wal, A.C., Kok, W.E., and Pinto, Y.M. (2010) MiR423-5p as a circulating biomarker for heart failure. *Circulation Research*, **106** (6), 1035–1039. doi: 10.1161/CIRCRESAHA.110.218297

45 Goren, Y., Kushnir, M., Zafrir, B., Tabak, S., Lewis, B.S., and Amir, O. (2012) Serum levels of microRNAs in patients with heart failure. *European Journal of Heart Failure*, **14** (2), 147–154. doi: 10.1093/eurjhf/hfr155.

46 Vogel, B., Keller, A., Frese, K.S., Leidinger, P., Sedaghat-Hamedani, F., Kayvanpour, E., Kloos, W., Backe, C., Thanaraj, A., Brefort, T., Beier, M., Hardt, S., Meese, E., Katus, H.A., and Meder, B. (2013) Multivariate miRNA signatures as biomarkers for non-ischaemic systolic heart failure. *European Heart Journal*. doi: 10.1093/eurheartj/eht256

47 Andersson, P., Gidlof, O., Braun, O.O., Gotberg, M., van der Pals, J., Olde, B., and Erlinge, D. (2012) Plasma levels of liver-specific miR-122 is massively increased in a porcine cardiogenic shock model and attenuated by hypothermia. *Shock*, **37** (2), 234–238. doi: 10.1097/SHK.0b013e31823f1811

3
MicroRNAs in Primary Brain Tumors: Functional Impact and Potential Use for Diagnostic Purposes

Patrick Roth and Michael Weller

3.1
Background

The incidence rate of all primary central nervous system (CNS) tumors is 20.6 cases per 100 000 in the United States with a slight preponderance of females [1]. It is 7.3 per 100 000 for malignant tumors and 13.3 per 100 000 for non-malignant neoplasms. Benign tumors may still cause significant morbidity because of their compressing properties on the surrounding healthy tissue. Both benign and malignant tumors can cause various neurological symptoms and signs, such as headaches, seizures, personality changes, and focal neurological deficits. The diagnostic steps are mainly based on imaging techniques, including computed tomography (CT) and magnetic resonance imaging (MRI). Positron emission tomography (PET) and cerebrospinal fluid (CSF) analysis may sometimes help to narrow the differential diagnosis. For the vast majority of CNS tumors, however, tumor tissue obtained by a resection or biopsy is required to establish the definitive diagnosis. The histopathological classification of the tumor is a prerequisite for any decision on the ultimate therapeutic management. The distribution of primary brain and CNS tumors by histology is shown in Figure 3.1. Benign tumors can often be cured by surgical resection, whereas malignant tumors frequently require additional therapeutic measures, such as radiation therapy or chemotherapy. Owing to the dismal prognosis of many malignant CNS tumors and the insufficient activity of the currently available treatment options, a better understanding of the biology of these tumors is required to allow for the development of novel treatment approaches. The involvement of microRNAs (miRNAs) in various physiological and pathological conditions has become clear only within the last 10 years. miRNAs are small, non-coding RNA molecules that regulate the stability or translational efficiency of their target messenger RNAs (mRNAs). The characterization of miRNA expression and, much more importantly, their functional contribution to the biological phenotype of tumor cells has therefore gained rapidly increasing

Nucleic Acids as Molecular Diagnostics, First Edition. Edited by Andreas Keller and Eckart Meese.

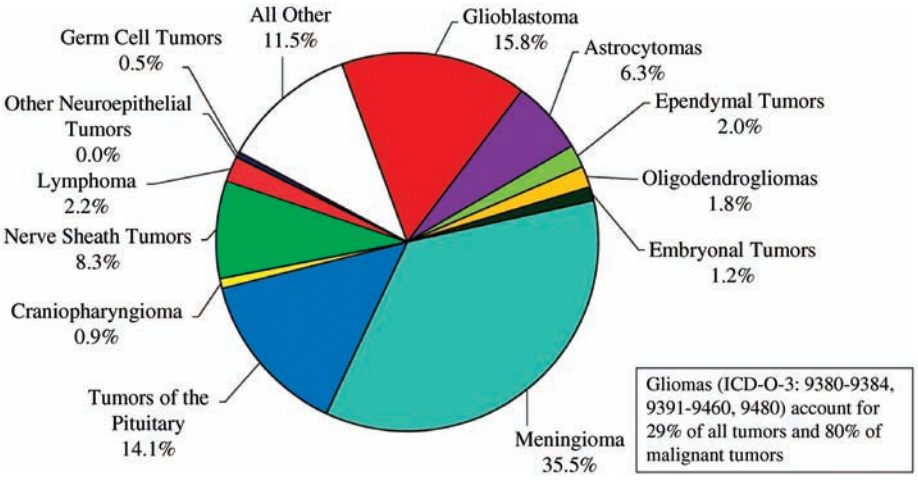


Figure 3.1 Distribution of primary brain and CNS tumors by histology ($N = 311\ 202$). ICD, International Classification of Diseases. (Reproduced with kind permission of Oxford Journals [1]. © 2012 Oxford University Press.)

interest [2]. Glioblastoma, the most common subtype of malignant brain tumors, has been in the focus of most miRNA studies within recent years. The current knowledge on miRNA function in glioblastoma and other primary CNS tumors will be summarized within this chapter. Furthermore, the ongoing attempts to use miRNA as novel diagnostic tools and biomarkers will be highlighted.

3.2 Gliomas

Among the most common primary (intrinsic) brain tumors are gliomas. It was traditionally thought that they were derived from the subpopulation of glial cells – the housekeeping cells maintaining function and integrity in the normal brain. Among the gliomas, glioblastoma is the most common and most malignant subtype. More than half of patients diagnosed with glioblastoma die from their cancer within the first 12 months of diagnosis. This tumor is more frequent in the elderly, but may develop across all age groups, including children [3]. Malignant glioma cells display several characteristics, including infiltrating growth pattern and resistance to different apoptotic stimuli [4]. The first reports on miRNA expression and function in glioblastoma cells were published in 2005. An initial microarray analysis of glioblastoma tissue and cell lines revealed a deregulation of various miRNAs [5]. In parallel, the first reports were published indicating a functional role for miRNAs in glioblastoma cells. Chan *et al.* described that miR-21 is overexpressed in glioblastoma cells.

miR-21 was identified subsequently as an antiapoptotic factor in these cells. Silencing of miR-21 in glioma cells triggered activation of caspases and resulted in the induction of apoptotic cell death [6]. miR-21 was the focus of numerous reports that emerged in the following years highlighting its contribution to the malignant phenotype of glioma cells and therefore it was also termed an "oncomiR." Similarly, the functional impact of other miRNAs on cell death pathways in glioma cells became apparent. Within recent years, numerous miRNAs have been described that affect various other biological properties of glioma cells, including (i) proliferation, migration, and invasiveness, (ii) resistance to immune cell attack as well as apoptotic stimuli, and (iii) metabolic activity and features of "stemness." The following provides examples for the diverse miRNA actions in glioma cells.

Among the best-studied miRNAs in glioblastoma is miR-10b, which is upregulated in glioblastoma according to several studies [7]. miR-10b may facilitate the invasiveness and growth of glioma cells. It may also sustain the stem cell phenotype of glioblastoma stem cell lines [8]. Inhibition of miR-10b resulted in reduced growth of experimental gliomas *in vivo* [9]. miRNAs acting as "tumor suppressors" are typically downregulated in glioma cells, including miR-326 that regulates glioma cell survival through its target pyruvate kinase type M2 (PKM2) [10,11]. The putative tumor suppressor calmodulin-binding transcription activator 1 (CAMTA1) is a target of miR-9/9* and miR-17, which are both abundantly expressed in glioma stem cells. miR-9/9* and miR-17 may therefore prevent the growth-suppressing function of CAMTA1 [12]. Other miRNAs have been linked to metabolic changes in glioma cells. As an example, miR-451 expression decreases in response to low glucose levels, which results in reduced proliferation, but increased migration and survival [13]. miR-124 targets the transcription factor STAT3 (signal transducer and activator of transcription 3) and is typically absent in gliomas. The introduction of miR-124 into glioma cells resulted in an inhibition of the STAT3 pathway with subsequent reversion of glioma-mediated suppression of T cell proliferation, indicating that miRNAs have the potential to modulate the complex interaction between the tumor and its microenvironment [14]. Other miRNA-mediated mechanisms that preclude immune cell responses against gliomas include the expression of miR-222 and miR-339 by glioma cells. Both miRNAs suppress the expression of ICAM (intercellular adhesion molecule)-1 on tumor cells, which confers resistance to T cell attacks against the tumor [15]. Diffuse astrocytomas (World Health Organization (WHO) grade II) may spontaneously progress to anaplastic astrocytomas (WHO grade III) and secondary glioblastomas (WHO grade IV). Such a progression is associated with an increased expression of a set of 12 miRNAs (miR-9, miR-15a, miR-16, miR-17, miR-19a, miR-20a, miR-21, miR-25, miR-28, miR-130b, miR-140, and miR-210) and two miRNAs (miR-184 and miR-328) that display reduced expression levels upon progression [16]. In summary, these data demonstrate that miRNAs are involved in various biological pathways in glioma cells and may also be implicated in the malignant progression of these tumors.

3.2.1 miRNA as Biomarkers in Glioma Tissue

The current WHO classification for gliomas is based on histopathological features of the tumor tissue. However, several molecular markers have turned out to be of diagnostic and prognostic significance. Oligodendroglial tumors are characterized by the frequent loss of genetic material from the short arm of chromosome 1 (1p) and the long arm of chromosome 19 (19q) [17]. Mutations in the isocitrate dehydrogenase *IDH* genes are found frequently in WHO grade II and III gliomas. It can be expected that these molecular markers will be included in the next WHO classification of CNS tumors [18]. It must be assumed that these are only the first steps towards a more detailed molecular profiling of gliomas also for standard diagnostics. As outlined above, many miRNAs are deregulated in glioblastoma. Specific miRNA patterns allow for a differentiation between glioblastoma and healthy brain tissue. Furthermore, single miRNAs or rather complex miRNA profiles may be used as prognostic and predictive biomarkers. Basically, prognostic biomarkers are indicators for the course of the disease independent of the applied treatment. In contrast, predictive biomarkers should predict the response to a particular treatment type. miRNAs have not yet entered the clinical routine as biomarkers. However, compared with mRNA, miRNAs are characterized by their stability, which is a prerequisite to use them for diagnostic purposes. Whether miRNA expression patterns in the tumor tissue may be helpful for a more detailed classification of gliomas needs to be determined.

Several reports indicate that high miR-21 levels in glioblastoma tissue are associated with a poor prognosis [19,20]. Similarly, high expression of miR-182 was described as a marker of poor outcome [21]. Other studies emphasized the combined upregulation of miR-195 and miR-196b as indicators for long-term survival [22]. Limitations of these studies include the assessment of heterogeneous patient populations as well as univariate analyses which exclude many other important parameters. A large-scale miRNA ($n = 756$) expression profiling of glioblastoma, anaplastic astrocytoma, and normal brain samples identified several differentially regulated miRNAs between these groups. The authors of the publication determined a discriminatory 23-miRNA expression signature that allowed for a differentiation between glioblastoma from anaplastic astrocytoma with an accuracy of 95% [23]. Using an *in silico* approach, data from The Cancer Genome Atlas (TCGA; cancergenome.nih.gov) dataset were analyzed to identify a miRNA expression signature that could predict survival of glioblastoma patients. The authors defined seven miRNAs that were associated with poor survival, whereas three miRNAs were correlated with long-term survival. This 10-miRNA signature was subsequently validated in an additional set, and permitted a classification of patients into groups with "high-risk" and overall poor survival compared to patients with "low-risk" scores. These results indicate that combined miRNA profiles may be suitable for the prognostic classification of glioblastoma patients [24]. In a similar approach using the TCGA database,

another study described miRNA expression profiles that permitted a characterization of several glioblastoma subclasses. The authors suggested that developmental miRNA expression patterns characterize and contribute to the phenotypic variety of glioblastoma subgroups [25]. Further analyses with larger datasets which include important molecular markers such as MGMT (O^6-methylguanine-DNA methyltransferase) promoter methylation status and IDH mutation status in a multivariate model are required to confirm these findings.

Whether miRNAs in glioma tissue can be exploited as predictive biomarkers is another field of research. One study reported that the expression of miR-181d was a predictive marker for response to temozolomide treatment in glioblastoma patients [26]. The underlying mechanism may be an inhibiting effect of miR-181d on the expression of the DNA repair protein MGMT, which confers resistance to glioblastoma cells to the deleterious effects of temozolomide [27]. However, these results also need confirmation in a prospective study and larger patient cohorts. Only such comprehensive studies will determine whether tissue-derived miRNA expression patterns may be valuable tools as biomarkers that eventually enter clinical practice.

3.2.2 Circulating miRNA as Biomarkers

miRNAs are not only present in tumor tissue, but also in body fluids such as blood, urine or cerebrospinal fluid (CSF). Here, they are easily accessible and their expression can be monitored continuously over time. Therefore, several efforts have aimed at exploiting the presence of miRNAs in blood or CSF for diagnostic purposes in cancer patients [28–30]. In the blood, miRNAs can be extracted from the serum or from the cellular components. Glioblastoma-specific expression patterns have been characterized in both ways. Some miRNAs that are upregulated in glioblastoma tissue are found at high levels in the blood, such as miR-21 [31]. However, single miRNAs typically do not discriminate well between different samples. Therefore, more comprehensive analyses are required that take advantage of the numerous miRNAs that can be detected by high-throughput techniques. A first proof-of-principle study that tested 1158 miRNAs in blood samples of glioblastoma patients and healthy controls demonstrated that 52 miRNAs were significantly deregulated. The most deregulated candidates were miR-128 (upregulated) and miR-342-3p (downregulated). Furthermore, a comprehensive miRNA signature was obtained that allowed for the discrimination between blood samples of glioblastoma patients and healthy controls with an accuracy of 81%, specificity of 79%, and sensitivity of 83% [32]. Serum-derived miRNA expression patterns were analyzed in another study from patients with WHO grade III and IV tumors as well as healthy controls. The analysis of miRNAs in the serum by Solexa sequencing followed by validation by real-time polymerase chain reaction (PCR) revealed a marked difference in serum miRNA profiles between patients with high-grade gliomas and healthy controls. miRNAs that were decreased in glioma patients included miR-15b*,

miR-23a, miR-133a, miR-150*, miR-197, miR-497, and miR-548b-5p. This miRNA panel had a high sensitivity (88%) and specificity (97.9%) for the prediction of a malignant glioma [33]. Again, these studies must be broadened using homogenous patient cohorts examined in a prospective manner.

A small study assessed the expression levels of predefined miRNAs in the CSF of glioblastoma patients. An upregulation of miR-15b and miR-21 compared with control CSF was reported [34]. In a similar approach, Teplyuk *et al.* examined miRNA profiles in the CSF in patients diagnosed with different types of brain tumors and non-neoplastic CNS diseases by real-time PCR. miR-10b and miR-21 levels were increased in patients with glioblastoma and brain metastasis of breast and lung cancer. High expression levels of members of the miR-200 family were only observed in patients with brain metastases and allowed for a differentiation from glioblastoma samples. Furthermore, preliminary findings suggest that longitudinal miRNA profiles in the CSF may allow for a monitoring of the response to treatment in these patients [35].

3.3 Meningiomas

Meningiomas account for approximately 36% of all primary brain tumors (www.cbtrus.org). They originate from the meninges and may cause neurological deficits by their mass effect on the adjacent brain parenchyma. The vast majority of meningiomas are classified as WHO grade I tumors, indicating the overall benign behavior of these neoplasms. A small percentage of meningiomas are referred to as atypical (WHO grade II) and anaplastic (WHO grade III) tumors. Atypical and anaplastic meningiomas are characterized by a more infiltrative growth pattern, and are associated with a higher risk of recurrence after initial resection. In sharp contrast to their high incidence, little is known about the role of miRNAs in the biology of meningioma pathogenesis. Indeed, only a very limited number of publications have addressed the functional meaning of miRNAs in meningioma cells. Still, it must be assumed that miRNAs are involved in various intracellular processes of these tumor cells, and that miRNAs are deregulated in atypical and anaplastic meningiomas. Expression levels of miR-145 are significantly reduced in atypical and anaplastic tumors as compared with benign meningiomas [36]. Overexpression of this miRNA in meningioma cells was associated with impaired proliferation, enhanced susceptibility to apoptosis, and decreased tumor size upon orthotopic inoculation in nude mice as compared with control cells. miR-335 is frequently overexpressed in meningiomas and may enhance meningioma cell growth [37]. miR-200a is downregulated, which results in increased expression of β-catenin and cyclin D_1. When miR-200a was overexpressed in meningioma cells, it inhibited β-catenin expression and subsequent Wnt/β-catenin signaling, which suggests a role for miR-200a as a tumor suppressor miRNA in meningioma [38]. Out of 200 miRNAs assessed in meningioma tissue samples, high expression of miR-190a and low expression of

miR-29c-3p and miR-219-5p were associated with an increased risk for recurrence [39]. In summary, miRNAs are involved in various biological mechanisms of meningioma cells; however, their exact role needs further investigation.

3.4 Pituitary Adenomas

Approximately 14% of all intracranial tumors are neoplasms of the pituitary gland (www.cbtrus.org). A first report on deregulated miRNAs in pituitary adenomas was published in 2005, describing that miR-15 and miR-16 are downregulated in these tumors, which is associated with greater tumor diameters [40]. A more comprehensive analysis revealed different miRNA expression profiles in normal pituitary as compared with pituitary adenomas [41]. In a similar study that focused exclusively on adrenocorticotropic hormone-secreting pituitary tumor samples, an altered miRNA expression profile was detected compared with normal pituitary tissue [42]. A bioinformatics analysis of the miRNA expression pattern in a set of pituitary adenomas determined five overexpressed miRNAs that target Smad3: miR-135a, miR-140-5p, miR-582-3p, miR-582-5p, and miR-93. The authors speculated that a suppression of Smad3-dependent transforming growth factor-β signaling by miRNAs may be involved in the tumorigenesis of these neoplasms [43]. *HMGA* (high mobility group A) genes are overexpressed in pituitary adenomas. Palmieri *et al.* described a downregulation of several miRNAs in pituitary adenomas of different histological subtypes that target *HMGA* genes [44]. A study focusing on growth hormone-secreting pituitary adenomas reported similar findings [45]. These results suggest that a downregulation of HMGA-targeting miRNAs could contribute to HMGA-dependent pituitary tumorigenesis. miR-26a,which is frequently overexpressed in human pituitary adenoma, targets protein kinase Cδ (PRKCD). miR-26a inhibition resulted in a cell cycle delay in the G_1 phase. miR-26a may therefore be implicated in the cell cycle control of pituitary cells [46]. Furthermore, Palumbo *et al.* demonstrated that miR-26b and miR-128 regulate the activity of the PTEN/AKT pathway in growth hormone-producing pituitary tumors [47]. Overall, similar to other tumor entities, miRNAs are involved in various intracellular pathways in pituitary adenomas. Their potential use for diagnostic and therapeutic purposes, however, has not yet been addressed.

3.5 Medulloblastomas

Medulloblastoma occurs typically in children and young adults. On a molecular level, four distinct subgroups can be distinguished: tumors with activated Wingless pathway signaling (WNT tumors), tumors with Sonic Hedgehog pathway activation (SHH tumors), and group 3 and 4 tumors which are molecularly less

well characterized. The first reports on miRNA function in medulloblastoma were only published in 2008. A set of three miRNAs is downregulated in medulloblastoma, which results in high expression of their target (i.e., the SHH pathway activator *smoothened*) [48]. Further studies revealed a functional collaboration between the miRNA cluster miR-17–92 and the SHH pathway in medulloblastoma pathogenesis, resulting in increased tumor growth [49,50]. In non-SHH medulloblastoma, miR-182 may contribute to leptomeningeal metastatic dissemination [51]. Several miRNAs that may act as tumor suppressors that could inhibit the growth of medulloblastoma are downregulated in these tumors, including miR-124 and miR-218 [52,53].

High-throughput miRNA expression profile analyses were used to compare human primary medulloblastoma specimens. Several specific miRNA expression patterns were obtained which allowed for a differentiation between different histological subtypes of medulloblastoma, such as anaplastic, classic, and desmoplastic tumors. Furthermore, the miRNA expression profiles obtained allowed for a differentiation between medulloblastoma and normal cerebellar tissue [54]. A combined analysis of medulloblastoma tissue by high-density single nucleotide polymorphism arrays and miRNA analysis identified six molecular subgroups associated with different clinical outcomes [55]. It must be awaited whether approaches that incorporate miRNA analyses reach the clinic and can be exploited for routine diagnostics as well as prognostic or predictive purposes within the next few years.

3.6 Other Brain Tumors

Apart from gliomas, meningiomas, pituitary adenomas, and medulloblastomas, the list of primary brain tumors comprises schwannomas, primary CNS lymphomas (PCNSL), and other rare entities. Overall, only limited data on miRNA expression and function in these tumors are available.

3.6.1 Schwannomas

The first miRNA described in the context of schwannomas was miR-21. It is overexpressed when compared with normal vestibular nerve tissue. Inhibition of miR-21 decreased the growth of vestibular schwannoma cells and increased apoptotic cell death [56]. miR-21 was therefore suggested as a putative target for therapies limiting the growth of these tumors. A high-throughput miRNA expression profiling of human vestibular schwannomas described miR-7 as one of the most downregulated miRNAs. Overexpression of miR-7 reduced schwannoma cell growth *in vitro* and *in vivo*, indicating that it might act as a tumor suppressor and could be used as a therapeutic tool [57].

3.6.2 PCNSLs

For PCNSLs, a divergent miRNA expression pattern compared with nodal lymphomas was determined using low-density PCR arrays [58]. Furthermore, the expression of miR-19, miR-21, and miR-90a was described to be upregulated significantly in the CSF of PCNSL patients with a specificity of 96.7% and sensitivity of 95.7% when used for a combined analysis [59]. miRNA expression in the CSF may also allow for monitoring of disease progression [60]. Confirmation of these findings in a prospective study is, however, required.

3.7 Summary and Outlook

Within less than a decade, miRNAs have emerged as one of the most important regulators of post-transcriptional gene expression. A multitude of studies, particularly in glioblastoma cells, has highlighted the contribution of deregulated miRNA expression to the biological phenotype of primary brain tumors. Owing to their differential regulation in tumor cells, their long-lasting stability as well as rapidly improving methods for their measurement, miRNAs from tumor tissue and circulating miRNAs in the blood or CSF may represent excellent biomarker candidates. Complex miRNA expression profiles usually contain more diagnostic information than single biomarkers. Microarray and sequencing technologies have become more and more accessible on a routine base, and allow for a rapid profiling of complex miRNA expression patterns. Open questions that need to be addressed include the assessment of miRNAs from fresh-frozen compared with formalin-fixed paraffin-embedded tissue. Furthermore, the ongoing characterization of novel miRNAs allows for even more detailed analyses, but represents an increasing technical challenge that requires more comprehensive analyses. Whether miRNAs from blood or CSF can be used to monitor disease progression needs to be assessed in large prospective patient cohorts, which will be time- and cost-consuming. Obviously, glioblastoma as the most common malignant brain tumor will be the most appropriate tumor entity that is suitable for such studies. In other tumor types, it may be difficult to gather sufficient tissue specimens from homogeneous patient populations within a reasonable time period. Finally, miRNAs may also be exploited for therapeutic approaches in the future, either by antagonizing their function with antisense oligonucleotides or by a potentiation of their function aiming at inhibiting the expression of target genes. Although promising, such strategies are still at a very early stage of clinical development and will require intense research efforts over the coming years.

References

1 Dolecek, T.A., Propp, J.M., Stroup, N.E., and Kruchko, C. (2012) CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2005–2009. *Journal of Neuro-Oncology*, **14** (Suppl. 5), v1–v49.

2 Farazi, T.A., Hoell, J.I., Morozov, P., and Tuschl, T. (2013) MicroRNAs in human cancer. *Advances in Experimental Medicine and Biology*, **774**, 1–20.

3 Ohgaki, H. and Kleihues, P. (2005) Population-based studies on incidence, survival rates, and genetic alterations in astrocytic and oligodendroglial gliomas. *Journal of Neuropathology and Experimental Neurology*, **64** (6), 479–489.

4 Ziegler, D.S., Kung, A.L., and Kieran, M.W. (2008) Anti-apoptosis mechanisms in malignant gliomas. *Journal of Clinical Oncology*, **26** (3), 493–500.

5 Ciafre, S.A., Galardi, S., Mangiola, A. *et al.* (2005) Extensive modulation of a set of microRNAs in primary glioblastoma. *Biochemical and Biophysical Research Communications*, **334** (4), 1351–1358.

6 Chan, J.A., Krichevsky, A.M., and Kosik, K.S. (2005) MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. *Cancer Research*, **65** (14), 6029–6033.

7 Sasayama, T., Nishihara, M., Kondoh, T., Hosoda, K., and Kohmura, E. (2009) MicroRNA-10b is overexpressed in malignant glioma and associated with tumor invasive factors, uPAR and RhoC. *International Journal of Cancer*, **125** (6), 1407–1413.

8 Guessous, F., Alvarado-Velez, M., Marcinkiewicz, L. *et al.* (2013) Oncogenic effects of miR-10b in glioblastoma stem cells. *Journal of Neuro-Oncology*, **112** (2), 153–163.

9 Gabriely, G., Yi, M., Narayan, R.S. *et al.* (2011) Human glioma growth is controlled by microRNA-10b. *Cancer Research*, **71** (10), 3563–3572.

10 Kefas, B., Comeau, L., Floyd, D.H. *et al.* (2009) The neuronal microRNA miR-326 acts in a feedback loop with notch and has therapeutic potential against brain tumors. *Journal of Neuroscience*, **29** (48), 15161–15168.

11 Kefas, B., Comeau, L., Erdle, N., Montgomery, E., Amos, S., and Purow, B. (2010) Pyruvate kinase M2 is a target of the tumor-suppressive microRNA-326 and regulates the survival of glioma cells. *Journal of Neuro-Oncology*, **12** (11), 1102–1112.

12 Schraivogel, D., Weinmann, L., Beier, D. *et al.* (2011) CAMTA1 is a novel tumour suppressor regulated by miR-9/9* in glioblastoma stem cells. *EMBO Journal*, **30** (20), 4309–4322.

13 Godlewski, J., Nowicki, M.O., Bronisz, A. *et al.* (2010) MicroRNA-451 regulates LKB1/AMPK signaling and allows adaptation to metabolic stress in glioma cells. *Molecular Cell*, **37** (5), 620–632.

14 Wei, J., Wang, F., Kong, L.Y. *et al.* (2013) miR-124 inhibits STAT3 signaling to enhance T cell-mediated immune clearance of glioma. *Cancer Research*, **73** (13), 3913–3926.

15 Ueda, R., Kohanbash, G., Sasaki, K. *et al.* (2009) Dicer-regulated microRNAs 222 and 339 promote resistance of cancer cells to cytotoxic T-lymphocytes by down-regulation of ICAM-1. *Proceedings of the National Academy of Sciences of the United States of America*, **106** (26), 10746–10751.

16 Malzkorn, B., Wolter, M., Liesenberg, F. *et al.* (2010) Identification and functional characterization of microRNAs involved in the malignant progression of gliomas. *Brain Pathology (Zurich, Switzerland)*, **20** (3), 539–550.

17 Roth, P., Wick, W., and Weller, M. (2013) Anaplastic oligodendroglioma: a new treatment paradigm and current controversies. *Current Treatment Options in Oncology*, **14** (4), 505–513.

18 Weller, M., Pfister, S.M., Wick, W., Hegi, M.E., Reifenberger, G., and Stupp, R. (2013) Molecular neuro-oncology in clinical practice: a new horizon. *Lancet Oncology*, **14** (9), e370–379.

19 Zhi, F., Chen, X., Wang, S. *et al.* (2010) The use of hsa-miR-21, hsa-miR-181b and hsa-miR-106a as prognostic indicators of

astrocytoma. *European Journal of Cancer*, **46** (9), 1640–1649.
20 Hermansen, S.K., Dahlrot, R.H., Nielsen, B.S., Hansen, S., and Kristensen, B.W. (2013) MiR-21 expression in the tumor cell compartment holds unfavorable prognostic value in gliomas. *Journal of Neuro-Oncology*, **111** (1), 71–81.
21 Jiang, L., Mao, P., Song, L. *et al.* (2010) miR-182 as a prognostic marker for glioma progression and patient survival. *American Journal of Pathology*, **177** (1), 29–38.
22 Lakomy, R., Sana, J., Hankeova, S. *et al.* (2011) MiR-195, miR-196b, miR-181c, miR-21 expression levels and *O*-6-methylguanine-DNA methyltransferase methylation status are associated with clinical outcome in glioblastoma patients. *Cancer Science*, **102** (12), 2186–2190.
23 Rao, S.A., Santosh, V., and Somasundaram, K. (2010) Genome-wide expression profiling identifies deregulated miRNAs in malignant astrocytoma. *Modern Pathology*, **23** (10), 1404–1417.
24 Srinivasan, S., Patric, I.R., and Somasundaram, K. (2011) A ten-microRNA expression signature predicts survival in glioblastoma. *PLoS One*, **6** (3), e17438.
25 Kim, T.M., Huang, W., Park, R., Park, P.J., and Johnson, M.D. (2011) A developmental taxonomy of glioblastoma defined and maintained by microRNAs. *Cancer Research*, **71** (9), 3387–3399.
26 Zhang, W., Zhang, J., Hoadley, K. *et al.* (2012) miR-181d: a predictive glioblastoma biomarker that downregulates MGMT expression. *Journal of Neuro-Oncology*, **14** (6), 712–719.
27 Weller, M., Stupp, R., Reifenberger, G. *et al.* (2010) MGMT promoter methylation in malignant gliomas: ready for personalized medicine? *Nature Reviews Neurology*, **6** (1), 39–51.
28 Chen, X., Ba, Y., Ma, L. *et al.* (2008) Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Research*, **18** (10), 997–1006.
29 Gilad, S., Meiri, E., Yogev, Y. *et al.* (2008) Serum microRNAs are promising novel biomarkers. *PLoS One*, **3** (9), e3148.
30 Mitchell, P.S., Parkin, R.K., Kroh, E.M. *et al.* (2008) Circulating microRNAs as stable blood-based markers for cancer detection. *Proceedings of the National Academy of Sciences of the United States of America*, **105** (30), 10513–10518.
31 Ilhan-Mutlu, A., Wagner, L., Wohrer, A. *et al.* (2012) Plasma MicroRNA-21 concentration may be a useful biomarker in glioblastoma patients. *Cancer Investigation*, **30** (8), 615–621.
32 Roth, P., Wischhusen, J., Happold, C. *et al.* (2011) A specific miRNA signature in the peripheral blood of glioblastoma patients. *Journal of Neurochemistry*, **118** (3), 449–457.
33 Yang, C., Wang, C., Chen, X. *et al.* (2013) Identification of seven serum microRNAs from a genome-wide serum microRNA expression profile as potential noninvasive biomarkers for malignant astrocytomas. *International Journal of Cancer*, **132** (1), 116–127.
34 Baraniskin, A., Kuhnhenn, J., Schlegel, U. *et al.* (2012) Identification of microRNAs in the cerebrospinal fluid as biomarker for the diagnosis of glioma. *Journal of Neuro-Oncology*, **14** (1), 29–33.
35 Teplyuk, N.M., Mollenhauer, B., Gabriely, G. *et al.* (2012) MicroRNAs in cerebrospinal fluid identify glioblastoma and metastatic brain cancers and reflect disease activity. *Journal of Neuro-Oncology*, **14** (6), 689–700.
36 Kliese, N., Gobrecht, P., Pachow, D. *et al.* (2012) miRNA-145 is downregulated in atypical and anaplastic meningiomas and negatively regulates motility and proliferation of meningioma cells. *Oncogene*, **32** (39), 4712–4720.
37 Shi, L., Jiang, D., Sun, G. *et al.* (2012) miR-335 promotes cell proliferation by directly targeting Rb1 in meningiomas. *Journal of Neuro-Oncology*, **110** (2), 155–162.
38 Saydam, O., Shen, Y., Wurdinger, T. *et al.* (2009) Downregulated microRNA-200a in meningiomas promotes tumor growth by reducing E-cadherin and activating the Wnt/beta-catenin signaling pathway. *Molecular and Cellular Biology*, **29** (21), 5923–5940.
39 Zhi, F., Zhou, G., Wang, S. *et al.* (2013) A microRNA expression signature predicts

meningioma recurrence. *International Journal of Cancer*, **132** (1), 128–136.

40 Bottoni, A., Piccin, D., Tagliati, F., Luchin, A., Zatelli, M.C., and degli Uberti, E.C. (2005) miR-15a and miR-16-1 down-regulation in pituitary adenomas. *Journal of Cellular Physiology*, **204** (1), 280–285.

41 Bottoni, A., Zatelli, M.C., Ferracin, M. *et al.* (2007) Identification of differentially expressed microRNAs by microarray: a possible role for microRNA genes in pituitary adenomas. *Journal of Cellular Physiology*, **210** (2), 370–377.

42 Amaral, F.C., Torres, N., Saggioro, F. *et al.* (2009) MicroRNAs differentially expressed in ACTH-secreting pituitary tumors. *Journal of Clinical Endocrinology and Metabolism*, **94** (1), 320–323.

43 Butz, H., Liko, I., Czirjak, S. *et al.* (2011) MicroRNA profile indicates downregulation of the TGFbeta pathway in sporadic non-functioning pituitary adenomas. *Pituitary*, **14** (2), 112–124.

44 Palmieri, D., D'Angelo, D., Valentino, T. *et al.* (2012) Downregulation of HMGA-targeting microRNAs has a critical role in human pituitary tumorigenesis. *Oncogene*, **31** (34), 3857–3865.

45 D'Angelo, D., Palmieri, D., Mussnich, P. *et al.* (2012) Altered microRNA expression profile in human pituitary GH adenomas: down-regulation of miRNA targeting HMGA1, HMGA2, and E2F1. *Journal of Clinical Endocrinology and Metabolism*, **97** (7), E1128–1138.

46 Gentilin, E., Tagliati, F., Filieri, C. *et al.* (2013) miR-26a plays an important role in cell cycle regulation in ACTH-secreting pituitary adenomas by modulating protein kinase Cdelta. *Endocrinology*, **154** (5), 1690–1700.

47 Palumbo, T., Faucz, F.R., Azevedo, M., Xekouki, P., Iliopoulos, D., and Stratakis, C.A. (2013) Functional screen analysis reveals miR-26b and miR-128 as central regulators of pituitary somatomammotrophic tumor growth through activation of the PTEN–AKT pathway. *Oncogene*, **32** (13), 1651–1659.

48 Ferretti, E., De Smaele, E., Miele, E. *et al.* (2008) Concerted microRNA control of Hedgehog signalling in cerebellar neuronal progenitor and tumour cells. *EMBO Journal*, **27** (19), 2616–2627.

49 Uziel, T., Karginov, F.V., Xie, S. *et al.* (2009) The miR-17∼92 cluster collaborates with the Sonic Hedgehog pathway in medulloblastoma. *Proceedings of the National Academy of Sciences of the United States of America*, **106** (8), 2812–2817.

50 Northcott, P.A., Fernandez, L.A., Hagan, J.P. *et al.* (2009) The miR-17/92 polycistron is up-regulated in sonic hedgehog-driven medulloblastomas and induced by N-*myc* in sonic hedgehog-treated cerebellar neural precursors. *Cancer Research*, **69** (8), 3249–3255.

51 Bai, A.H., Milde, T., Remke, M. *et al.* (2012) MicroRNA-182 promotes leptomeningeal spread of non-sonic hedgehog-medulloblastoma. *Acta Neuropathologica*, **123** (4), 529–538.

52 Silber, J., Hashizume, R., Felix, T. *et al.* (2013) Expression of miR-124 inhibits growth of medulloblastoma cells. *Journal of Neuro-Oncology*, **15** (1), 83–90.

53 Venkataraman, S., Birks, D.K., Balakrishnan, I. *et al.* (2013) MicroRNA 218 acts as a tumor suppressor by targeting multiple cancer phenotype-associated genes in medulloblastoma. *Journal of Biological Chemistry*, **288** (3), 1918–1928.

54 Ferretti, E., De Smaele, E., Po, A. *et al.* (2009) MicroRNA profiling in human medulloblastoma. *International Journal of Cancer*, **124** (3), 568–577.

55 Cho, Y.J., Tsherniak, A., Tamayo, P. *et al.* (2011) Integrative genomic analysis of medulloblastoma identifies a molecular subgroup that drives poor clinical outcome. *Journal of Clinical Oncology*, **29** (11), 1424–1430.

56 Cioffi, J.A., Yue, W.Y., Mendolia-Loffredo, S., Hansen, K.R., Wackym, P.A., and Hansen, M.R. (2010) MicroRNA-21 overexpression contributes to vestibular schwannoma cell proliferation and survival. *Otology & Neurotology*, **31** (9), 1455–1462.

57 Saydam, O., Senol, O., Wurdinger, T. *et al.* (2011) miRNA-7 attenuation in Schwannoma tumors stimulates growth by

upregulating three oncogenic signaling pathways. *Cancer Research*, **71** (3), 852–861.

58 Fischer, L., Hummel, M., Korfel, A., Lenze, D., Joehrens, K., and Thiel, E. (2011) Differential micro-RNA expression in primary CNS and nodal diffuse large B-cell lymphomas. *Journal of Neuro-Oncology*, **13** (10), 1090–1098.

59 Baraniskin, A., Kuhnhenn, J., Schlegel, U. *et al.* (2011) Identification of microRNAs in the cerebrospinal fluid as marker for primary diffuse large B-cell lymphoma of the central nervous system. *Blood*, **117** (11), 3140–3146.

60 Baraniskin, A., Kuhnhenn, J., Schlegel, U., Schmiegel, W., Hahn, S., and Schroers, R. (2012) MicroRNAs in cerebrospinal fluid as biomarker for disease course monitoring in primary central nervous system lymphoma. *Journal of Neuro-Oncology*, **109** (2), 239–244.

4
Genetic and Epigenetic Alterations in Sporadic Colorectal Cancer: Clinical Implications

Pawel Karpinski, Nikolaus Blin, and Maria M. Sasiadek

4.1
Introduction

Current data from the EUROCARE program indicate that as a result of advances in the early detection and treatment of colorectal cancer (CRC), the 5-year relative survival rate has improved substantially during the last 30 years [1]. However, CRC is the most frequently diagnosed cancer for both genders combined in the European population and its incidence has increased significantly in recent years, with a growth rate 0.5% per year likely due to the increase in life expectancy [2]. An increasing frequency of CRC and relatively high mortality underscores the urgency to develop reliable early biomarkers as well as novel targeted therapies and to integrate them into routine clinical practice. Over the last 15 years, great progress has been made in our understanding of the molecular changes involved in the etiology of CRC. A constantly expanding catalog of molecular markers identified in colorectal adenocarcinoma has led to progressive changes in the therapeutic approach paradigm from selecting therapy on the basis of the anatomical site and histopathological data to incorporating molecular classification [3,4]. Although interest in the molecular classification of tumors has been increasing, the heterogeneity of CRCs together with the high number of complex and interconnected changes affecting the regulation of gene expression have made it very difficult to identify molecular abnormalities that directly influence pathophysiology and outcome of diseases [5]. Sequential distortion of genetic and epigenetic landscapes over time results in alterations of gene expression patterns leading to a malignant and invasive cancer stage. During this transformation, tumor cells acquire mutational signatures that reflect their evolution by the presence of global genomic or epigenomic alternations as well as single molecular defects [3,6]. "Driver" mutational pathways and randomly accumulated genetic events both contribute to tumor development; however, the relative impact they have on the CRC phenotype and clinical behavior still remains uncertain. Moreover, tumor heterogeneity leads to biological plasticity of the neoplasm. This chapter covers our current knowledge on genetic

Nucleic Acids as Molecular Diagnostics, First Edition. Edited by Andreas Keller and Eckart Meese.

and epigenetic alterations in sporadic CRC with a particular emphasis on their potential impact on phenotype and prognosis.

4.2 Chromosomal Instability

Over the past 30 years, chromosomal instability (CIN) has been recognized as one of the most common genetic alterations in human neoplasias and has been attributed to up to 80% in CRCs [7]. Despite broad CIN prevalence in cancer, the mechanisms of both its occurrence and its contribution to cancer development and progression still remain elusive [8]. Moreover, a precise definition of CIN is still not available. Therefore, many studies on CRCs use CIN to describe cancers with either structural chromosome rearrangements (translocations, deletions, amplifications, inversions) or frequent loss of heterozygosity (LOH), while others define CIN tumors that display aneuploid/polyploid karyotypes. Geigl *et al.* and other authors recently tentatively defined CIN as cell-to-cell variability in gain or loss of either part or whole chromosomes; however, due to technical obstacles, few if any published data on CRC provide reliable information about cell-to-cell variability in chromosome number and structural rearrangements [7,9]. For the purpose of this chapter, we adhere to the less rigorous definition of CIN that includes stable (over cell generations) chromosomal and segmental aneuploidies (including copy number alterations).

Whole-chromosome aneuploidy is a consequence of chromosome missegregation and can be defined as alteration of chromosome number that is not a multiple of the haploid complement (loss and gain of whole chromosomes) [10]. Although there is strong evidence for a high frequency of aneuploidy in cancer, its contribution to tumorigenesis is largely unknown [11,12]. In fact, aneuploidy is also found in normal human cells, including hepatocytes and neurons [13,14]. Furthermore, using a mouse model, Weaver *et al.* demonstrated that aneuploidy under a specific genetic context may inhibit tumorigenesis [15]. Thus, it is hypothesized that aneuploidy may create a favorable context (i.e., by triggering cell stress response) for the acquisition of mutations and copy number changes of a few particularly dosage-sensitive genes that may have direct consequences on tumorigenesis [11,16,17]. Moreover, it has been demonstrated that while CIN is associated with poor cancer outcome, an extremely high level of CIN is associated with better prognosis, probably resulting from increased tumor cell lethality [18].

The mechanisms responsible for reduced fidelity of chromosome segregation include defects in sister chromatid cohesion, faulty spindle assembly checkpoints, and centrosome copy number errors. Although recent studies employing yeast, cell lines and mouse models have greatly extended the list of genes involved in aneuploidy, such as mitotic spindle genes (*Bub1*, *Mad2*), sister chromatid cohesion-related genes (*SMC1L1*, *NIPBL*, *STAG2*) and cycle checkpoint genes (*Cdc4*), only a small fraction of CRCs (around 20%) display any of the above-mentioned mutations [8,19,20].

In most sporadic CRC cases (up to 80%), CIN is a predominant genetic feature. It is speculated that mutations in such genes as *APC*, *KRAS*, and *TP53*, which are frequent in sporadic CRCs (greater than 40% each), may also contribute to aneuploidy; however, not all tumors with mutations in the listed genes manifest aneuploidy [21–24]. Moreover, *in vivo* experiments have shown that only *APC* mutants become aneuploid [25]. This may suggest that mutations in *APC*, *KRAS*, and *TP53* may indirectly promote chromosome missegregation in ways that had not been fully recognized [26]. Two recent studies by Thompson *et al.* and Li *et al.* using human cell line and mouse models provide elegant evidence for how inactivation of *TP53* may contribute to aneuploidy [27,28]. Both groups have shown that non-diploid genomes are not tolerated in normal cells; however, loss of *TP53* permits survival and proliferation of aneuploid cells. This likely explains the correlation of *TP53* and CIN in CRC [29].

According to the Mitelman Database of Chromosome Aberrations in Cancer (http://cgap.nci.nih.gov/Chromosomes; June 2013), the most frequent trisomies detected in adenocarcinomas of the large intestine are +7, +13, and +20; among the most frequent monosomies are −18, −14, −22, −17, −4, and −15. The most recurrent aberration found in almost all CRC karyotyping studies is deletion of chromosome 18 [30]. Aneuploidy may also reflect tumor progression. As shown in a meta-analysis by Diep *et al.*, who studied a large set of CRC studies, loss of chromosome 18 and gain of chromosome 20 belong to early events in colorectal carcinogenesis, whereas losses of chromosomes 11 and 19 constitute late events [31].

Previous studies including meta-analyses have indicated that diploid tumor content measured by flow cytometry favors better prognosis of CRC patients when compared with aneuploid tumors [32,33]. However, ploidy assessed by cytometric methods does not reflect more complex structural changes in tumor karyotype (segmental gains and/or deletions), which, as shown many times, are stronger, independent predictors of CRC patient prognosis than ploidy itself [32].

Unlike whole-chromosome aneuploidy, somatic copy number alterations (SCNAs) have a well-documented impact on tumor development and progression through the activation of oncogenes and the inactivation of tumor suppressors (by LOH). A recent high-resolution study by Beroukhim *et al.* on 3131 cancer specimens demonstrated that chromosomal arm-level SCNAs are more frequent than focal SCNAs (median length 1.8 Mb) [12]. At least two categories of DNA repair mechanisms are known to give rise to SCNAs: non-allelic homologous recombination (NAHR) and microhomology-mediated break-induced replication (MMBIR) [34]. Both NAHR and MMBIR are error-prone and both belong to mechanisms engaged in cells during stress [35,36]. Numerous reports present evidence that stress-inducing downregulation of recombinational repair genes may cause a switch from high-fidelity homologous repair to error-prone MMBIR. Indeed, Bindra *et al.* found that *RAD51* and *BRCA1* (both components of homologous repair) are repressed by hypoxia in mammalian cells [37,38]. The biological consequence of hypoxia-induced DNA repair switch is a likely induction of chromosomal instability [39]. Supporting this idea are observations that

RAD51D-deficient mouse embryos display extensive chromosomal instability [40]. Surprisingly, genes that are involved directly in induction of SCNAs in CRC remain unknown [41]. Genes such as *APC*, *KRAS*, *PIK3CA*, *TP53*, *SMAD2*, and *SMAD4* are frequently mutated in CRCs that show a high level of SCNAs. It has previously been suggested that such genetic changes may indirectly promote the generation of SCNAs by still unknown mechanisms [41]. As mentioned before, *TP53* inactivation permits aneuploid cells to continue cell cycling [27,28]. If it is assumed that chromosome segregation errors may cause structural chromosomal aberrations, it can be hypothesized that frequent *TP53* loss of function may be at least partially responsible for acquisition of a CIN phenotype in CRC [42,43]. Recently Baba *et al.* reported that Aurora-A kinase (*AURKA*) overexpression is directly associated with CIN [44]. Given the role of *AURKA* in the regulation of the function of centrosomes, spindles, and kinetochores, this could also explain CIN in a fraction (20%) of CRCs [44].

Over the past decade, enormous progress has been made in defining SCNAs in CRC due to development of high-resolution array techniques. The Cancer Genome Atlas (TCGA; cancergenome.nih.gov/) program combined detailed, high-resolution analysis of SCNAs in 257 CRCs, confirming data obtained in a large set of previous studies [45]. Most frequent arm-level changes include gains of 1q, 7p and q, 8p and q, 12q, 13q, 19q, and 20p and q, whereas losses mainly cover 18p and q, 17p and q, 1p, 4q, 5q, 8p, 14q, 15q, 20p, and 22q. The TCGA study has identified 28 recurrent focal deletions and 17 focal amplifications that contain 874 and 212 genes, respectively [45]. It is likely that among these genes only a few may have a direct impact on CRCs tumorigenesis (see [38] for more details).

It is also well proven that arm-level and focal SCNAs associate with adenoma-to-carcinoma progression. By studying 112 colorectal adenomas and 82 colorectal carcinomas, Hermsen *et al.* found that 17p and 18q loss and 8q, 13q, and 20q gains are early events in CRC carcinogenesis [46]. Their data were confirmed in 859 CRCs of various Dukes' stages [31]. This study also linked loss of 4p with Dukes transition from A to B–D stages, and revealed 14q loss and 1q and 12p gains as late events [31]. Additional investigations disclosed an association of 8p loss and 7p and 17q gains with liver metastasis [31,47].

As mentioned before, arm-level and focal SCNAs are strong predictors of patient survival and prognosis. A study by Xie *et al.* on stage II and III CRCs has associated 20q gain with improved survival [48]. Several studies have related deletions at 18q, 8p, 4p, and 15q to poor prognosis [32,49]. A recent report by Watanabe *et al.* connected the CIN level (measured by LOH in four chromosomes) to survival in 750 patients classified with stage II and III CRC [50]. Patients displaying the so-called CIN-high phenotype (at least three chromosomes out of four affected by LOH) showed significantly poorer survival than patients with less pronounced LOH [50]. Growing evidence suggests that poor outcome associated with CIN may be related to increased tumor adaptation to environmental stresses. Lee *et al.* have recently demonstrated that chromosomally unstable CRC cell lines exhibit significant multidrug resistance when

compared with chromosomally stable lines [51]. The authors supplemented their findings by a meta-analysis of CRC outcome following adjuvant therapy with 5-fluorouracil (5-FU), showing that patients with CIN (measured by flow cytometry) display worse progression-free or disease-free survival when compared with diploid CRC patients [51]. The high prevalence of CIN in CRC has also explained the discouraging results associated with the introduction of taxane-based therapies for CRC [52].

Recently, a novel phenomenon (termed chromothripsis) was described in CRC and other cancers as a possible consequence of CIN [53,54]. Chromothripsis likely results from massive chromosome fragmentation and subsequent inaccurate DNA repair, which leads to an extensive concentration of rearrangements in a single chromosome, whole chromosome arm, or segment [53]. Chromothripsis may contribute to carcinogenesis by the disruption of tumor suppressors and generation of fusion genes [54]. Interestingly, a recent study pointed to a strong association between somatic *TP53* mutations and chromothripsis in acute myeloid leukemia [55]. Details of the extent, importance, and impact on prognosis of chromothripsis in CRC remain to be elucidated.

Due to multiple forms of CIN, none of the methods used to study CIN provide satisfactory information [56]. Labor-intensive measurements of cell-to-cell variability in gain or loss of either part or whole chromosomes became an obstacle for implementation of such methods into routine practice [7]. Among strategies to target CIN in CRC, the most popular methods enable the assessment of stable (over cell generations) chromosomal and segmental aneuploidies. These strategies include identification of ploidy level by flow cytometry, determination of LOH status of selected loci, and, finally, comparative genomic hybridization array-based profiling [45,50,57]

4.3 Microsatellite Instability

Unlike CIN, microsatellite instability (MSI) in CRC has been molecularly and clinically well-characterized during the past two decades [58]. Instability of microsatellite sequences refers to tumor-specific length alterations within short, repetitive sequences that occur abundantly and randomly in the human genome, including introns and exons of genes [59]. During DNA replication, microsatellite sequences are often affected by one nucleotide expansion or contraction due to polymerase slippage that results in formation of a deletion/insertion loop. Subsequently, after replication these loops are processed by a specific DNA repair system (mismatch repair (MMR)), conserved during evolution from bacteria to humans [60]. When the MMR system is inactivate, alterations in microsatellite length remain unrepaired and result in microsatellite instability (expansion or contraction). Consequently, MSI leads to gene inactivation by accumulation of frameshift mutations within exonic microsatellites [61]. In about 10–15% of sporadic CRCs, the MMR system is inactivated exclusively by

epigenetic silencing of the *hMLH1* gene while somatic mutations MMR genes are rarely detected [62,63]. MSI has been assessed by polymerase chain reaction (PCR) using the so-called Bethesda marker panel that includes five microsatellite loci: two mononucleotides and three dinucleotides. However, recent data supports the use of a pentaplex panel of mononucleotide repeats as it is more sensitive and performs better on non-white populations [64–66]. Until recently, three distinct CRC subsets were recognized according to MSI: MSI-high, MSI-low and microsatellite stable (MSS); however, molecular and clinical differences between MSI-low and MSS seemed to highly overlap [60]. Thus, in most recent publications, CRC tumors are divided simply into MSI and MSS. Recently, a distinct type of MSI has been recognized in CRC which is characterized by elevated microsatellite alterations at selected tetranucleotide repeats (EMASTs) [67]. EMASTs have been attributed to loss of *MSH3* expression and have been detected in 60% of sporadic CRCs (including MSI-high, MSI-low, and MSS tumors) [68]. A recent study by Campregher *et al.* revealed that although *MSH3* deficiency induces EMASTs, it does not lead to oncogenic transformation [69]. Thus, the role of EMASTs in CRC remains controversial.

Tumors displaying MSI exhibit a specific molecular and clinical phenotype that is characterized by a higher prevalence in women, proximal location (Figure 4.1), low pathological stage, mucinous features, and tumor-infiltrating lymphocytes [60,70]. MSI CRCs possibly arise in sessile serrated adenomas/polyps [71]. At the molecular level, MSI CRCs strongly correlate with $BRAF^{V600E}$ mutation, CIMP-high (see Section 4.7), and a low frequency of CIN and LOH events [72,73]. Several studies have addressed the genome-wide differences in gene expression and SCNAs between MSI and MSS tumors, indicating significant differences with respect to expression patterns and type of chromosomal aberrations [74,75]. The inverse association of MSI with CIN suggests that these are two distinct pathways to colon cancer; however, both forms of genetic instability are not mutually exclusive as the MSI tumors display some degree of CIN [75–77]. In addition, recent advances in high-throughput sequencing have led researchers to gain insight into the spectrum of somatic mutations in MSI and MSS CRCs [78,79]. Timmermann *et al.* concluded that MSI CRCs, due to their nature, harbor up to eightfold more coding somatic mutations than MSS cancers [78]. A subsequent study by Alhopuro *et al.* aimed at distinguishing driver frameshift mutations from passengers in a large group of MSI tumors. This study identified driver frameshift mutations in 10 genes, including *TGFBR2*, *ACVR2*, *MSH3*, *GLYR1*, and *ABCC5* [79].

Multiple studies have been published investigating the impact of MSI on prognosis in CRC patients. A meta-analysis including 7642 CRC cases published in 2005 confirmed the more favorable prognosis for MSI CRC patients when compared with MSS patients [81]. This finding has subsequently been confirmed by other research groups; however, the association between MSI and positive prognosis has been limited to stage II CRCs [60,82]. The reason for a positive prognostic effect of MSI has as yet not been fully understood. Recent research by Williams *et al.* has shed some light on this issue [83]. MSI gives rise to a variety

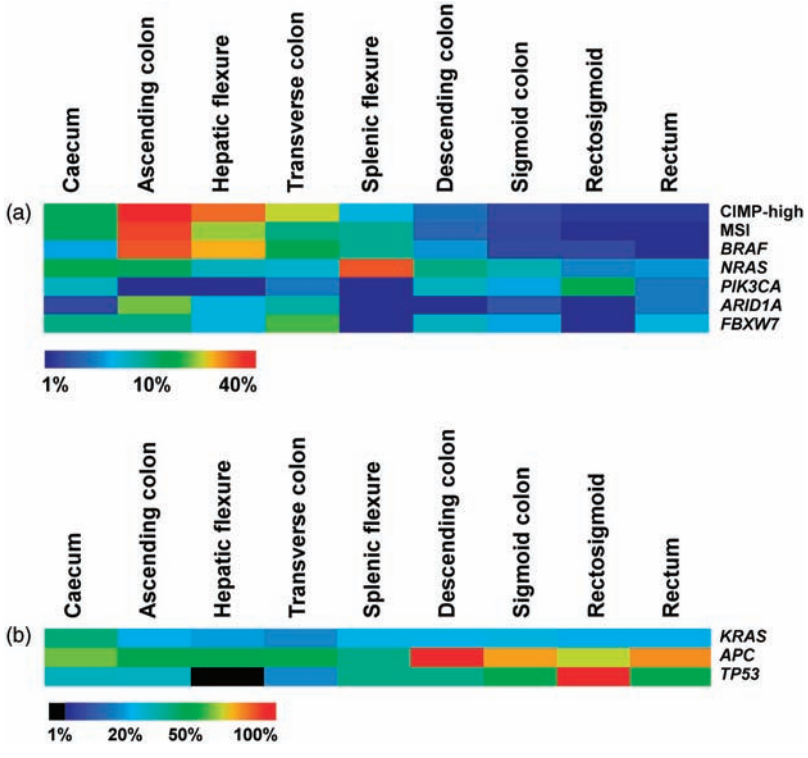


Figure 4.1 Distribution of important molecular changes along large intestine subsites Diagram illustrates frequencies (%) of (a) CIMP-high, MSI, mutations in *BRAF*, *PIK3CA*, *NRAS*, *ARID1A*, and *FBXW7* genes; and (b) mutations in *KRAS*, *APC*, and *TP53* genes. Note the gradual change of frequencies of CIMP-high, MSI, $BRAF^{V600E}$, and *TP53* mutations along the length of the colon. (Data from Yamauchi *et al.* [80] and The Cancer Genome Atlas Network [45].)

of mutant proteins that appear less degradable and are a source of tumor-specific antigens. These in turn stimulate tumor-infiltrating T cells, which likely confer an improved clinical outcome in MSI patients, as was also shown by Nosho *et al.* in a large group of CRCs [83,84].

The influence of MSI on outcome of adjuvant chemotherapy (including 5-FU and irinotecan) remains disputable. Despite conflicting results, current evidence proves that MSI patients do not benefit from 5-FU-based therapy; however, studies have been limited to stage II CRCs and such a conclusion for other stages has not yet been postulated [71,85]. Lack of response for 5-FU-based therapy has recently been attributed to recquirement of a fully functional MMR system to trigger 5-FU cytotoxicity [86,87]. In contrast, two independent studies have shown that MSI might be a positive prognostic marker of irinotecan regimen in advanced CRC; however, such evidence is still considered preliminary [88,89].

In summary, MSI has been recognized as a well-characterized CRC phenotype as it is relatively easy to detect and is clearly defining in Lynch syndrome cases; however, MSI testing still awaits implementation into clinical practice [60,70,90].

Recent progress in the detection of a heritable predisposition to CRC (Lynch syndrome) has triggered universal testing for MSI in all patients newly diagnosed with CRC [91].

Several assays have been used for the detection of MSI in CRC, including mutation screening of MMR genes, MMR protein detection by immunohistochemistry, and assessment of length changes in a set of microsatellites by PCR and subsequent capillary electrophoresis. Several panels of microsatellites have been used to detect MSI. An older panel approved by an international consensus meeting in 1997 consists of three dinucleotide repeats and two mononucleotide repeats [66]. However, the use of this panel requires normal tissue to be compared with tumor tissue. In recent years, a pentaplex panel of quasimonomorphic mononucleotide repeat markers has gained popularity as it offers better sensitivity, specificity, and robustness, and, importantly, obviates the need for normal DNA testing [64,92]. It can be speculated that in the near future this particular method will supersede other methodologies used to determine MSI [92].

4.4 Driver Somatic Mutations in CRC

In the era of deep, high-throughput sequencing, each tumor can be characterized by hundreds of somatic mutations, of which only a small fraction contribute to the tumor's progression by conferring a growth advantage (referred to as driver mutations) [93]. Here, we will focus on a short characterization of genes in which somatic mutations exert a well-documented impact on sporadic CRC tumorigenesis and/or prognosis. The frequencies of major driver somatic mutations along various colon subsites are shown in Figure 4.1.

4.4.1 APC

The *APC* (adenomatous polyposis coli) protein interacts with a number of other cancer-related proteins, and thus it functions as a signaling platform. Numerous studies have demonstrated that APC is capable of interacting within several fundamental cellular processes, including Wnt signaling, cell adhesion and migration, organization of microtubules, spindle formation, and chromosome segregation [94]. *APC* mutations or loss of the *APC* locus are frequently detected in adenomas and CRCs (around 60%), thus APC inactivation occurs early in colorectal carcinogenesis [95,96]. Many investigators have implicated *APC* mutations in the occurrence of CIN in CRC tumors; however, careful analysis of current data suggests a modulatory role for *APC* in this process [22]. *APC*

mutations alone do not affect the patient's survival, as often reported [97–99]. The relation of *APC* mutations to 5-FU-based therapy is not well known [98].

4.4.2 TP53

Similarly to APC, the *TP53* protein is involved in various cellular processes, including DNA damage repair, cell cycle regulation, apoptosis, and centrosome clustering. *TP53*-inactivating mutations are relatively frequent among CRC cases (around 40%) [100]. A large study of 3583 CRC cases reported *TP53* mutations associated with worse prognosis in advanced CRCs [100]. Other published data have shown insufficient evidence regarding a predictive value of *TP53* mutations on prognosis and response to chemotherapy [101,102]. It is currently speculated that the specific localizations of mutations across *TP53* protein domains may directly apply to the prognosis of cancer patients [103,104].

4.4.3 KRAS

The *KRAS* protooncogene participates in a signal transduction cascade dependent on binding of peptide growth factors by epidermal growth factor receptor (EGFR); this subsequently regulates multiple effector molecules, which leads to cell proliferation [105]. The majority of *KRAS* mutations are located in codons 12, 13, and 61, resulting in constitutive *KRAS* activation. *KRAS* mutations, arising early during colon tumorigenesis, are present in about 40% of CRCs with the most frequent location at codon 12 [106]. In recent large trial studies, *KRAS* mutations were found to have no prognostic value for patients with II and III stage CRC treated with 5-FU-based therapies [107]. However, a study encompassing 1075 CRC cases has shown that mutations in codon 12 confer poor survival when compared with *KRAS* wild-type patients [106]. However, due to the functional effect of *KRAS* mutations (permanent activation of EGFR-dependent pathways) they are very useful predictors of patient lack of responsiveness to anti-EGFR therapies in metastatic CRC [108].

4.4.4 BRAF

The *BRAF* protooncogene is one of the most important downstream effectors of *KRAS* [109]. *BRAF* is frequently activated by a hotspot mutation V600E that accounts for up to 90% of all *BRAF* mutations detected in around 12% CRCs [110]. Interestingly, *BRAF* mutations are mutually exclusive with *KRAS* mutations in CRCs [109]. $BRAF^{V600E}$ is thought to arise in a subtype of serrated polyps called sessile serrated adenomas [111]. This mutation displays significant correlation with proximal localization, MSI, and methylator phenotype in CRCs [112]. Interestingly, in microsatellite stable CRCs, $BRAF^{V600E}$ significantly

correlates with CIN and *TP53* mutations [113,114]. Studies on a large number of CRC cases have shown that $BRAF^{V600E}$ is associated with a poor prognosis in microsatellite stable CRCs [115,116]. Similar to somatic mutations in *KRAS*, $BRAF^{V600E}$ is associated with a lack of response to anti-EGFR therapies in metastatic CRC [117].

4.4.5
PIK3CA

The *PIK3CA* protooncogene belongs to a family of signaling lipid kinases that transduces signals from activated tyrosine kinase receptors as well as activated RAS proteins (including KRAS). Downstream signaling of activated *PIK3CA* protein triggers various cellular processes, including cell growth and survival pathways [118]. Mutations in *PIK3CA* are present around 15% of CRCs and exhibit significant correlation with proximal location, *KRAS* mutation, and methylator phenotype [119]. A recent study by Liao *et al.* on 1170 CRC cases demonstrated that patients with *PIK3CA* mutations experience significantly worse survival than *PIK3CA* wild-type cases [120]. Current data suggest that *PIK3CA* mutations confer resistance to anti-EGFR therapies [121].

4.4.6
Other Mutations

In addition to known, albeit infrequently mutated genes (e.g., *NRAS*), recent high-throughput sequencing studies have identified numerous recurrently mutated genes that may have significance for colorectal tumorigenesis [45,122,123]. The list of novel genes includes *POLE* (polymerase ε), *ARID1A*, *SOX9*, *FAM123B*, and many more (see [45] for the details). However, their functional relevance and pathogenic role remain to be elucidated.

A variety of classical strategies have been utilized to assess the mutation status of important genes in CRC, including immunohistochemistry (specifically for *TP53*) and Sanger sequencing (*TP53*, *APC*, and other genes); however, both techniques require a substantial amount of hands-on tech time [124,125]. The clinical relevance of *KRAS* and *BRAF* mutations for anti-EGFR therapies in metastatic CRC has recently stimulated the development of sensitive, cost-effective, and high-throughput assays that include pyrosequencing or the multiplex SNaPshot assay [126,127]. The latter method has been frequently utilized to assess selected mutations in *PIK3CA* and *NRAS* genes [126,128].

4.5
Epigenetic Instability in CRC

In recent years, an increasing amount of data concerning epigenetic changes as one of major contributing factors in CRC have been published. During colorectal

carcinogenesis, tumor genomes undergo global loss of methylation (hypomethylation) along with focal hypomethylation of tumor suppressor gene promoters. Perturbations of methylation are accompanied by dynamic changes in histone modification, which subsequently lead to changes in gene expression patterns [129]. Here, we focus on global epigenetic alternations in CRC, such as hypomethylation and methylator phenotype, and their impact on the pathogenesis of CRC.

4.6 Hypomethylation

DNA methylation means a covalent modification of DNA by addition of methyl residues to cytosine bases in DNA (me^5C) catalyzed by DNA methyltransferases (DNMTs). In humans, me^5C is located exclusively within CpG dinucleotides. In eukaryotic cells, me^5C is required for control of tissue-specific transcriptional profiles, allele-specific expression of imprinted genes, and the transcriptional repression of retrotransposons [130]. Although the DNA methylation pattern constitutes a stable modification of the genomic DNA that can be inherited, it can also undergo dynamic changes. Normally, methylated sequences are located in imprinted and tissue-specific genes, the inactive X chromosome, pericentromeric sequences, intragenic regions, and repetitive elements. In fact, a large fraction of CpGs (45%) is located in repetitive elements that are heavily methylated [131].

The significant decrease of me^5C content in cancer genomes including CRC was recognized about 30 years ago [132,133]. Subsequently, hypomethylation has been associated with increased chromosomal instability, global loss of imprinting, and activation of protooncogenes [134]. Early methods of assessing the global me^5C content were laborious and required large amounts of DNA [135]. As the retrotransposable LINE-1 elements are highly repetitive and represent a substantial portion of human genome (18%), they have been in use as a surrogate marker of global methylation in CRC since 2004 [135]. Thus, the relationship between LINE-1 methylation and various genomic events in colorectal carcinogenesis has been widely investigated. LINE-1 demethylation occurs very early and progresses continuously thorough progression of CRC [136]. LINE-1 demethylation displays an inverse association with methylator phenotype and MSI [137]. Significantly lower mean methylation of LINE-1 has been reported for LOH at 18q, which supports a possible link between global hypomethylation and chromosomal instability [137]. Hur *et al.* have recently demonstrated LINE-1 methylation-dependent activation of the *MET* protooncogene in CRC metastases [134]. It has been shown conclusively that a higher LINE-1 methylation level is a marker for better prognosis in patients treated by resection alone. Kawakami *et al.* recently reported that LINE-1 methylation is not a predictive value of survival benefit of 5-FU-based chemotherapy; however, due to the small sample size, caution must be applied to this finding [138].

4.7
CpG Island Methylator Phenotype

In normal cells, me^5C is generally absent apart from short stretches of CpG-rich sequences known as CpG islands (CGIs). However, in cancers, focal hypermethylation of CGIs leading to silencing of tumor suppressors is a common key to the tumourigenic process. In 1999, Toyota *et al.* demonstrated that a subset of CRCs appears to possess an excessive number of methylated genes and consequently termed this phenomenon the CpG island methylator phenotype (CIMP) [139]. The concept of CIMP was criticized initially, with the assertion that division of CIMP-positive and -negative CRCs is arbitrary and the selection of CIMP-sensitive markers is likely biased [76]. While this criticism was upheld for several years, an increasing number of studies confirming the notion that CIMP is characteristic for a unique subgroup of around 15% CRCs presenting with a distinct molecular etiology has been published. The unique molecular signature and clinical characteristics of CIMP CRCs include: extensive methylation of CIMP-related CpG islands (CIMP-high), extensive methylation of the 3p22 and 2q14 regions, high prevalence of MSI and $BRAF^{V600E}$ mutation, serrated polyps, proximal localization (Figure 4.1), and older age. CIMP-high tumors display an inverse correlation with *TP53*, *KRAS*, and *APC* mutations, chromosomal instability, and LINE-1 demethylation [72,124,137,140,141]. The underlying molecular mechanisms for CIMP have not been identified [73]. Recent studies reported reduced survival for CIMP-high patients with microsatellite stable tumors and CIMP-high, microsatellite stable rectal cancer [142,143]. According to a preliminary study by Jover *et al.*, patients with CIMP-high colorectal tumors do not benefit from 5-FU-based adjuvant chemotherapy; however, there is not yet sufficient data to recommend its clinical use [71,144].

Currently, numerous unsupervised hierarchical clustering analyses have proposed at least three (instead of two) classes of CIMP (epigenotypes): CIMP-high, CIMP-low, and CIMP-0 depending on the extent of methylation of specific marker loci [73,145,146]. CIMP-low (also designated as CIMP-2 or IME) is noted for enriched *KRAS* mutations, while tumors with infrequent CIMP-related marker methylation (designated as CIMP-negative, CIMP-0, or LME) have been characterized by a frequent LOH at 18q and *TP53* mutations [72,146]. However, methodological inconsistencies exist in these studies in attribution of tumors to specific CIMP-low and CIMP-0 epigenotypes, and thus further research is necessary in the area of epigenotyping CRCs [147].

Although CIMP was reported more than decade ago, no universal method and marker panel to measure CIMP have been developed. Various marker panels and analytical approaches used to detect CIMP are summarized in reviews by Hughes *et al.* [148,149]

To date, at least two marker panels have been repeatedly used to assess CIMP status in CRC. The so-called "classic panel," which includes *MINT1*, *MINT2*, *MINT31*, *CDKN2A* (p16), and *MLH1*, has been utilized mainly by the means of methylation-specific PCR [139]. A more recent marker set, the so-called

"Weisenberger panel," includes *CACA1G*, *IGF2*, *NEUROG1*, *RUNX3*, and *SOCS*, which are assessed by quantitative technology (MethyLight) [112].

4.8 Concluding Remarks

Although significant advances have been made in the molecular characterization of CRC, as well as its treatment, the mortality in patients with this particular cancer is still rising. Along with the increasing number of chemotherapeutic drugs, precise molecular genotyping and subgroup identification is essential for the prognosis and prediction of therapeutic responses, as has been demonstrated for *KRAS*, *BRAF*, and *PIK3CA* mutations in metastatic CRC [150]. The dynamically expanding knowledge of molecular heterogeneity of CRC may be illustrated by a recent study by Yamauchi *et al.* who combined known molecular factors with a multisegmental approach to bowel subsites, demonstrating that the frequencies of key tumor molecular features change along the length of the colon (Figure 4.1) [80]. However, despite the promising advances in molecular pathology of CRC, its biological complexity still remains a challenge for an efficient approach to personalized therapy.

References

1 Brenner, H., Bouvier, A.M., Foschi, R., Hackl, M., Larsen, I.K., Lemmens, V., Mangone, L., Francisci, S., and Group, E. W. (2012) Progress in colorectal cancer survival in Europe from the late 1980s to the early 21st century: the EUROCARE study. *International Journal of Cancer*, **131** (7), 1649–1658.

2 Zavoral, M., Suchanek, S., Zavada, F., Dusek, L., Muzik, J., Seifert, B., and Fric, P. (2009) Colorectal cancer screening in Europe. *World Journal of Gastroenterology*, **15** (47), 5907–5915.

3 Ogino, S. and Goel, A. (2008) Molecular classification and correlates in colorectal cancer. *Journal of Molecular Diagnostics*, **10** (1), 13–27.

4 Greystoke, A. and Mullamitha, S.A. (2012) How many diseases are colorectal cancer? *Gastroenterology Research and Practice*, **2012**, 564741.

5 Pancione, M., Remo, A., and Colantuoni, V. (2012) Genetic and epigenetic events generate multiple pathways in colorectal cancer progression. *Pathology Research International*, **2012**, 509348.

6 Jass, J.R. (2007) Molecular heterogeneity of colorectal cancer: Implications for cancer control. *Journal of Surgical Oncology*, **16** (Suppl. 1), S7–S9.

7 Geigl, J.B., Obenauf, A.C., Schwarzbraun, T., and Speicher, M.R. (2008) Defining 'chromosomal instability'. *Trends in Genetics*, **24** (2), 64–69.

8 Thompson, S.L., Bakhoum, S.F., and Compton, D.A. (2010) Mechanisms of chromosomal instability. *Current Biology*, **20** (6), R285–R295.

9 Bakhoum, S.F. and Compton, D.A. (2012) Chromosomal instability and cancer: a complex relationship with therapeutic potential. *Journal of Clinical Investigation*, **122** (4), 1138–1143.

10 Williams, B.R. and Amon, A. (2009) Aneuploidy: cancer's fatal flaw? *Cancer Research*, **69** (13), 5289–5291.

11 Gordon, D.J., Resio, B., and Pellman, D. (2012) Causes and consequences of

aneuploidy in cancer. *Nature Reviews Genetics*, **13** (3), 189–203.

12 Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., Mc Henry, K.T., Pinchback, R.M., Ligon, A.H., Cho, Y.J., Haery, L., Greulich, H., Reich, M., Winckler, W., Lawrence, M.S., Weir, B. A., Tanaka, K.E., Chiang, D.Y., Bass, A.J., Loo, A., Hoffman, C., Prensner, J., Liefeld, T., Gao, Q., Yecies, D., Signoretti, S., Maher, E., Kaye, F.J., Sasaki, H., Tepper, J. E., Fletcher, J.A., Tabernero, J., Baselga, J., Tsao, M.S., Demichelis, F., Rubin, M.A., Janne, P.A., Daly, M.J., Nucera, C., Levine, R.L., Ebert, B.L., Gabriel, S., Rustgi, A.K., Antonescu, C.R., Ladanyi, M., Letai, A., Garraway, L.A., Loda, M., Beer, D.G., True, L.D., Okamoto, A., Pomeroy, S.L., Singer, S., Golub, T.R., Lander, E.S., Getz, G., Sellers, W.R., and Meyerson, M. (2010) The landscape of somatic copy-number alteration across human cancers. *Nature*, **463** (7283), 899–905.

13 Yurov, Y.B., Iourov, I.Y., Vorsanova, S.G., Liehr, T., Kolotii, A.D., Kutsev, S.I., Pellestor, F., Beresheva, A.K., Demidova, I.A., Kravets, V.S., Monakhov, V.V., and Soloviev, I.V. (2007) Aneuploidy and confined chromosomal mosaicism in the developing human brain. *PLoS One*, **2** (6), e558.

14 Duncan, A.W., Hanlon Newell, A.E., Smith, L., Wilson, E.M., Olson, S.B., Thayer, M.J., Strom, S.C., and Grompe, M. (2012) Frequent aneuploidy among normal human hepatocytes. *Gastroenterology*, **142** (1), 25–28.

15 Weaver, B.A., Silk, A.D., Montagna, C., Verdier-Pinard, P., and Cleveland, D.W. (2007) Aneuploidy acts both oncogenically and as a tumor suppressor. *Cancer Cell*, **11** (1), 25–36.

16 Torres, E.M., Williams, B.R., and Amon, A. (2008) Aneuploidy: cells losing their balance. *Genetics*, **179** (2), 737–746.

17 Sheltzer, J.M., Torres, E.M., Dunham, M. J., and Amon, A. (2012) Transcriptional consequences of aneuploidy. *Proceedings of the National Academy of Sciences of the United States of America*, **109** (31), 12644–12649.

18 Birkbak, N.J., Eklund, A.C., Li, Q., McClelland, S.E., Endesfelder, D., Tan, P., Tan, I.B., Richardson, A.L., Szallasi, Z., and Swanton, C. (2011) Paradoxical relationship between chromosomal instability and survival outcome in cancer. *Cancer Research*, **71** (10), 3447–3452.

19 Barber, T.D., McManus, K., Yuen, K.W., Reis, M., Parmigiani, G., Shen, D., Barrett, I., Nouhi, Y., Spencer, F., Markowitz, S., Velculescu, V.E., Kinzler, K.W., Vogelstein, B., Lengauer, C., and Hieter, P. (2008) Chromatid cohesion defects may underlie chromosome instability in human colorectal cancers. *Proceedings of the National Academy of Sciences of the United States of America*, **105** (9), 3443–3448.

20 Wang, Z., Cummins, J.M., Shen, D., Cahill, D.P., Jallepalli, P.V., Wang, T.L., Parsons, D.W., Traverso, G., Awad, M., Silliman, N., Ptak, J., Szabo, S., Willson, J. K., Markowitz, S.D., Goldberg, M.L., Karess, R., Kinzler, K.W., Vogelstein, B., Velculescu, V.E., and Lengauer, C. (2004) Three classes of genes mutated in colorectal cancers with chromosomal instability. *Cancer Research*, **64** (9), 2998–3001.

21 Rao, C.V. and Yamada, H.Y. (2013) Genomic instability and colon carcinogenesis: from the perspective of genes. *Frontiers in Oncology*, **3**, 130.

22 Rusan, N.M. and Peifer, M. (2008) Original CIN: reviewing roles for APC in chromosome instability. *Journal of Cell Biology*, **181** (5), 719–726.

23 Calistri, D., Rengucci, C., Seymour, I., Leonardi, E., Truini, M., Malacarne, D., Castagnola, P., and Giaretti, W. (2006) KRAS, p53 and BRAF gene mutations and aneuploidy in sporadic colorectal cancer progression. *Cellular Oncology*, **28** (4), 161–166.

24 Kamata, T. and Pritchard, C. (2011) Mechanisms of aneuploidy induction by RAS and RAF oncogenes. *American Journal of Cancer Research*, **1** (7), 955–971.

25 Fodde, R., Kuipers, J., Rosenberg, C., Smits, R., Kielman, M., Gaspar, C., van Es, J.H., Breukel, C., Wiegant, J., Giles, R.H., and Clevers, H. (2001) Mutations in the APC tumour suppressor gene cause chromosomal instability. *Nature Cell Biology*, **3** (4), 433–438.

26 Boland, C.R., Komarova, N.L., and Goel, A. (2009) Chromosomal instability and cancer: not just one CINgle mechanism. *Gut*, **58** (2), 163–164.

27 Thompson, S.L. and Compton, D.A. (2010) Proliferation of aneuploid human cells is limited by a p53-dependent mechanism. *Journal of Cell Biology*, **188** (3), 369–381.

28 Li, M., Fang, X., Baker, D.J., Guo, L., Gao, X., Wei, Z., Han, S., van Deursen, J.M., and Zhang, P. (2010) The ATM–p53 pathway suppresses aneuploidy-induced tumorigenesis. *Proceedings of the National Academy of Sciences of the United States of America*, **107** (32), 14188–14193.

29 Leslie, A., Pratt, N.R., Gillespie, K., Sales, M., Kernohan, N.M., Smith, G., Wolf, C.R., Carey, F.A., and Steele, R.J. (2003) Mutations of APC, K-ras, and p53 are associated with specific chromosomal aberrations in colorectal adenocarcinomas. *Cancer Research*, **63** (15), 4656–4661.

30 Migliore, L., Migheli, F., Spisni, R., and Coppedè, F. (2011) Genetics, cytogenetics, and epigenetics of colorectal cancer. *Journal of Biomedicine & Biotechnology*, **2011**, 792362.

31 Diep, C.B., Kleivi, K., Ribeiro, F.R., Teixeira, M.R., Lindgjaerde, O.C., and Lothe, R.A. (2006) The order of genetic events associated with colorectal cancer progression inferred from meta-analysis of copy number changes. *Genes, Chromosomes & Cancer*, **45** (1), 31–41.

32 Bardi, G., Fenger, C., Johansson, B., Mitelman, F., and Heim, S. (2004) Tumor karyotype predicts clinical outcome in colorectal cancer patients. *Journal of Clinical Oncology*, **22** (13), 2623–2634.

33 Araujo, S.E., Bernardo, W.M., Habr-Gama, A., Kiss, D.R., and Cecconello, I. (2007) DNA ploidy status and prognosis in colorectal cancer: a meta-analysis of published data. *Diseases of the Colon and Rectum*, **50** (11), 1800–1810.

34 Hastings, P.J., Lupski, J.R., Rosenberg, S. M., and Ira, G. (2009) Mechanisms of change in gene copy number. *Nature Reviews Genetics*, **10** (8), 551–564.

35 Arlt, M.F., Mulle, J.G., Schaibley, V.M., Ragland, R.L., Durkin, S.G., Warren, S.T., and Glover, T.W. (2009) Replication stress induces genome-wide copy number changes in human cells that resemble polymorphic and pathogenic variants. *American Journal of Human Genetics*, **84** (3), 339–350.

36 Hastings, P.J., Ira, G., and Lupski, J.R. (2009) A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genetics*, **5** (1), e1000327.

37 Bindra, R.S. and Glazer, P.M. (2007) Repression of RAD51 gene expression by E2F4/p130 complexes in hypoxia. *Oncogene*, **26** (14), 2048–2057.

38 Bindra, R.S., Gibson, S.L., Meng, A., Westermark, U., Jasin, M., Pierce, A.J., Bristow, R.G., Classon, M.K., and Glazer, P.M. (2005) Hypoxia-induced down-regulation of BRCA1 expression by E2Fs. *Cancer Research*, **65** (24), 11597–11604.

39 Huang, L.E., Bindra, R.S., Glazer, P.M., and Harris, A.L. (2007) Hypoxia-induced genetic instability – a calculated mechanism underlying tumor progression. *Journal of Molecular Medicine*, **85** (2), 139–148.

40 Smiraldo, P.G., Gruver, A.M., Osborn, J.C., and Pittman, D.L. (2005) Extensive chromosomal instability in Rad51d-deficient mouse cells. *Cancer Research*, **65** (6), 2089–2096.

41 Pino, M.S. and Chung, D.C. (2010) The chromosomal instability pathway in colon cancer. *Gastroenterology*, **138** (6), 2059–2072.

42 Janssen, A., van der Burg, M., Szuhai, K., Kops, G.J., and Medema, R.H. (2011) Chromosome segregation errors as a cause of DNA damage and structural chromosome aberrations. *Science*, **333** (6051), 1895–1898.

43 Nosho, K., Shima, K., Kure, S., Irahara, N., Baba, Y., Chen, L., Kirkner, G.J., Fuchs, C.S., and Ogino, S. (2009) JC virus

T-antigen in colorectal cancer is associated with p53 expression and chromosomal instability, independent of CpG island methylator phenotype. *Neoplasia*, **11** (1), 87–95.

44 Baba, Y., Nosho, K., Shima, K., Irahara, N., Kure, S., Toyoda, S., Kirkner, G.J., Goel, A., Fuchs, C.S., and Ogino, S. (2009) Aurora-A expression is independently associated with chromosomal instability in colorectal cancer. *Neoplasia*, **11** (5), 418–425.

45 The Cancer Genome Atlas Network (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, **487** (7407), 330–337.

46 Hermsen, M., Postma, C., Baak, J., Weiss, M., Rapallo, A., Sciutto, A., Roemen, G., Arends, J.W., Williams, R., Giaretti, W., De Goeij, A., and Meijer, G. (2002) Colorectal adenoma to carcinoma progression follows multiple pathways of chromosomal instability. *Gastroenterology*, **123** (4), 1109–1119.

47 Muñoz-Bellvis, L., Fontanillo, C., González-González, M., Garcia, E., Iglesias, M., Esteban, C., Gutierrez, M.L., Abad, M.M., Bengoechea, O., De Las Rivas, J., Orfao, A., and Sayagués, J.M. (2012) Unique genetic profile of sporadic colorectal cancer liver metastasis versus primary tumors as defined by high-density single-nucleotide polymorphism arrays. *Modern Pathology*, **25** (4), 590–601.

48 Xie, T., D'Ario, G., Lamb, J.R., Martin, E., Wang, K., Tejpar, S., Delorenzi, M., Bosman, F.T., Roth, A.D., Yan, P., Bougel, S., Di Narzo, A.F., Popovici, V., Budinská, E., Mao, M., Weinrich, S.L., Rejto, P.A., and Hodgson, J.G. (2012) A comprehensive characterization of genome-wide copy number aberrations in colorectal cancer reveals novel oncogenes and patterns of alterations. *PLoS One*, **7** (7), e42001.

49 Sheffer, M., Bacolod, M.D., Zuk, O., Giardina, S.F., Pincas, H., Barany, F., Paty, P.B., Gerald, W.L., Notterman, D.A., and Domany, E. (2009) Association of survival and disease progression with chromosomal instability: a genomic exploration of colorectal cancer. *Proceedings of the National Academy of Sciences of the United States of America*, **106** (17), 7131–7136.

50 Watanabe, T., Kobunai, T., Yamamoto, Y., Matsuda, K., Ishihara, S., Nozawa, K., Yamada, H., Hayama, T., Inoue, E., Tamura, J., Iinuma, H., Akiyoshi, T., and Muto, T. (2012) Chromosomal instability (CIN) phenotype, CIN high or CIN low, predicts survival for colorectal cancer. *Journal of Clinical Oncology*, **30** (18), 2256–2264.

51 Lee, A.J., Endesfelder, D., Rowan, A.J., Walther, A., Birkbak, N.J., Futreal, P.A., Downward, J., Szallasi, Z., Tomlinson, I. P., Howell, M., Kschischo, M., and Swanton, C. (2011) Chromosomal instability confers intrinsic multidrug resistance. *Cancer Research*, **71** (5), 1858–1870.

52 Swanton, C., Nicke, B., Schuett, M., Eklund, A.C., Ng, C., Li, Q., Hardcastle, T., Lee, A., Roy, R., East, P., Kschischo, M., Endesfelder, D., Wylie, P., Kim, S.N., Chen, J.G., Howell, M., Ried, T., Habermann, J.K., Auer, G., Brenton, J.D., Szallasi, Z., and Downward, J. (2009) Chromosomal instability determines taxane response. *Proceedings of the National Academy of Sciences of the United States of America*, **106** (21), 8671–8676.

53 Forment, J.V., Kaidi, A., and Jackson, S.P. (2012) Chromothripsis and cancer: causes and consequences of chromosome shattering. *Nature Reviews Cancer*, **12** (10), 663–670.

54 Kloosterman, W.P., Hoogstraat, M., Paling, O., Tavakoli-Yaraki, M., Renkens, I., Vermaat, J.S., van Roosmalen, M.J., van Lieshout, S., Nijman, I.J., Roessingh, W., van't Slot, R., van de Belt, J., Guryev, V., Koudijs, M., Voest, E., and Cuppen, E. (2011) Chromothripsis is a common mechanism driving genomic rearrangements in primary and metastatic colorectal cancer. *Genome Biology*, **12** (10), R103.

55 Rausch, T., Jones, D.T., Zapatka, M., Stütz, A.M., Zichner, T., Weischenfeldt, J., Jäger, N., Remke, M., Shih, D., Northcott, P.A., Pfaff, E., Tica, J., Wang,

Q., Massimi, L., Witt, H., Bender, S., Pleier, S., Cin, H., Hawkins, C., Beck, C., von Deimling, A., Hans, V., Brors, B., Eils, R., Scheurlen, W., Blake, J., Benes, V., Kulozik, A.E., Witt, O., Martin, D., Zhang, C., Porat, R., Merino, D.M., Wasserman, J., Jabado, N., Fontebasso, A., Bullinger, L., Rücker, F.G., Döhner, K., Döhner, H., Koster, J., Molenaar, J.J., Versteeg, R., Kool, M., Tabori, U., Malkin, D., Korshunov, A., Taylor, M.D., Lichter, P., Pfister, S.M., and Korbel, J.O. (2012) Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. *Cell*, **148** (1–2), 59–71.

56 Gaasenbeek, M., Howarth, K., Rowan, A. J., Gorman, P.A., Jones, A., Chaplin, T., Liu, Y., Bicknell, D., Davison, E.J., Fiegler, H., Carter, N.P., Roylance, R.R., and Tomlinson, I.P. (2006) Combined array-comparative genomic hybridization and single-nucleotide polymorphism-loss of heterozygosity analysis reveals complex changes and multiple forms of chromosomal instability in colorectal cancers. *Cancer Research*, **66** (7), 3471–3479.

57 Walther, A., Houlston, R., and Tomlinson, I. (2008) Association between chromosomal instability and prognosis in colorectal cancer: a meta-analysis. *Gut*, **57** (7), 941–950.

58 Vilar, E. and Tabernero, J. (2013) Molecular dissection of microsatellite instable colorectal cancer. *Cancer Discovery*, **3** (5), 502–511.

59 Ellegren, H. (2004) Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, **5** (6), 435–445.

60 Vilar, E. and Gruber, S.B. (2010) Microsatellite instability in colorectal cancer – the stable evidence. *Nature Reviews Clinical Oncology*, **7** (3), 153–162.

61 Jiricny, J. (2006) The multifaceted mismatch-repair system. *Nature Reviews Molecular Cell Biology*, **7** (5), 335–346.

62 Jeong, S.Y., Shin, K.H., Shin, J.H., Ku, J.L., Shin, Y.K., Park, S.Y., Kim, W.H., and Park, J.G. (2003) Microsatellite instability and mutations in DNA mismatch repair genes in sporadic colorectal cancers. *Diseases of the Colon and Rectum*, **46** (8), 1069–1077.

63 Børresen, A.L., Lothe, R.A., Meling, G.I., Lystad, S., Morrison, P., Lipford, J., Kane, M.F., Rognum, T.O., and Kolodner, R.D. (1995) Somatic mutations in the hMSH2 gene in microsatellite unstable colorectal carcinomas. *Human Molecular Genetics*, **4** (11), 2065–2072.

64 Buhard, O., Cattaneo, F., Wong, Y.F., Yim, S.F., Friedman, E., Flejou, J.F., Duval, A., and Hamelin, R. (2006) Multipopulation analysis of polymorphisms in five mononucleotide repeats used to determine the microsatellite instability status of human tumors. *Journal of Clinical Oncology*, **24** (2), 241–251.

65 Xicola, R.M., Llor, X., Pons, E., Castells, A., Alenda, C., Piñol, V., Andreu, M., Castellví-Bel, S., Payá, A., Jover, R., Bessa, X., Girós, A., Duque, J.M., Nicolás-Pérez, D., Garcia, A.M., Rigau, J., Gassull, M.A., and Gastrointestinal Oncology Group of the Spanish Gastroenterological Association (2007) Performance of different microsatellite marker panels for detection of mismatch repair-deficient colorectal tumors. *Journal of the National Cancer Institute*, **99** (3), 244–252.

66 Boland, C.R., Thibodeau, S.N., Hamilton, S.R., Sidransky, D., Eshleman, J.R., Burt, R.W., Meltzer, S.J., Rodriguez-Bigas, M. A., Fodde, R., Ranzani, G.N., and Srivastava, S. (1998) A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Research*, **58** (22), 5248–5257.

67 Haugen, A.C., Goel, A., Yamada, K., Marra, G., Nguyen, T.P., Nagasaka, T., Kanazawa, S., Koike, J., Kikuchi, Y., Zhong, X., Arita, M., Shibuya, K., Oshimura, M., Hemmi, H., Boland, C.R., and Koi, M. (2008) Genetic instability caused by loss of MutS homologue 3 in human colorectal cancer. *Cancer Research*, **68** (20), 8465–8472.

68 Boland, C.R. and Goel, A. (2010) Microsatellite instability in colorectal

cancer. *Gastroenterology*, **138** (6), 2073–2087.

69 Campregher, C., Schmid, G., Ferk, F., Knasmüller, S., Khare, V., Kortüm, B., Dammann, K., Lang, M., Scharl, T., Spittler, A., Roig, A.I., Shay, J.W., Gerner, C., and Gasche, C. (2012) MSH3-deficiency initiates EMAST without oncogenic transformation of human colon epithelial cells. *PLoS One*, **7** (11), e50541.

70 Funkhouser, W.K., Lubin, I.M., Monzon, F.A., Zehnbauer, B.A., Evans, J.P., Ogino, S., and Nowak, J.A. (2012) Relevance, pathogenesis, and testing algorithm for mismatch repair-defective colorectal carcinomas: a report of the association for molecular pathology. *Journal of Molecular Diagnostics*, **14** (2), 91–103.

71 Pritchard, C.C. and Grady, W.M. (2011) Colorectal cancer molecular biology moves into clinical practice. *Gut*, **60** (1), 116–129.

72 Ogino, S., Kawasaki, T., Kirkner, G.J., Ohnishi, M., and Fuchs, C.S. (2007) 18q loss of heterozygosity in microsatellite stable colorectal cancer is correlated with CpG island methylator phenotype-negative (CIMP-0) and inversely with CIMP-low and CIMP-high. *BMC Cancer*, **7**, 72.

73 Hinoue, T., Weisenberger, D.J., Lange, C.P., Shen, H., Byun, H.M., Van Den Berg, D., Malik, S., Pan, F., Noushmehr, H., van Dijk, C.M., Tollenaar, R.A., and Laird, P.W. (2012) Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Research*, **22** (2), 271–282.

74 Watanabe, T., Kobunai, T., Toda, E., Yamamoto, Y., Kanazawa, T., Kazama, Y., Tanaka, J., Tanaka, T., Konishi, T., Okayama, Y., Sugimoto, Y., Oka, T., Sasaki, S., Muto, T., and Nagawa, H. (2006) Distal colorectal cancers with microsatellite instability (MSI) display distinct gene expression profiles that are different from proximal MSI cancers. *Cancer Research*, **66** (20), 9804–9808.

75 Jorissen, R.N., Lipton, L., Gibbs, P., Chapman, M., Desai, J., Jones, I.T., Yeatman, T.J., East, P., Tomlinson, I.P., Verspaget, H.W., Aaltonen, L.A., Kruhøffer, M., Orntoft, T.F., Andersen, C. L., and Sieber, O.M. (2008) DNA copy-number alterations underlie gene expression differences between microsatellite stable and unstable colorectal cancers. *Clinical Cancer Research*, **14** (24), 8061–8069.

76 Grady, W.M. (2007) CIMP and colon cancer gets more complicated. *Gut*, **56** (11), 1498–1500.

77 Trautmann, K., Terdiman, J.P., French, A.J., Roydasgupta, R., Sein, N., Kakar, S., Fridlyand, J., Snijders, A.M., Albertson, D.G., Thibodeau, S.N., and Waldman, F.M. (2006) Chromosomal instability in microsatellite-unstable and stable colon cancer. *Clinical Cancer Research*, **12** (21), 6379–6385.

78 Timmermann, B., Kerick, M., Roehr, C., Fischer, A., Isau, M., Boerno, S.T., Wunderlich, A., Barmeyer, C., Seemann, P., Koenig, J., Lappe, M., Kuss, A.W., Garshasbi, M., Bertram, L., Trappe, K., Werber, M., Herrmann, B.G., Zatloukal, K., Lehrach, H., and Schweiger, M.R. (2010) Somatic mutation profiles of MSI and MSS colorectal cancer identified by whole exome next generation sequencing and bioinformatics analysis. *PLoS One*, **5** (12), e15661.

79 Alhopuro, P., Sammalkorpi, H., Niittymäki, I., Biström, M., Raitila, A., Saharinen, J., Nousiainen, K., Lehtonen, H.J., Heliövaara, E., Puhakka, J., Tuupanen, S., Sousa, S., Seruca, R., Ferreira, A.M., Hofstra, R.M., Mecklin, J. P., Järvinen, H., Ristimäki, A., Orntoft, T. F., Hautaniemi, S., Arango, D., Karhu, A., and Aaltonen, L.A. (2012) Candidate driver genes in microsatellite-unstable colorectal cancer. *International Journal of Cancer*, **130** (7), 1558–1566.

80 Yamauchi, M., Morikawa, T., Kuchiba, A., Imamura, Y., Qian, Z.R., Nishihara, R., Liao, X., Waldron, L., Hoshida, Y., Huttenhower, C., Chan, A.T., Giovannucci, E., Fuchs, C., and Ogino, S. (2012) Assessment of colorectal cancer molecular features along bowel subsites challenges the conception of distinct dichotomy of proximal versus distal colorectum. *Gut*, **61** (6), 847–854.

81 Popat, S., Hubner, R., and Houlston, R.S. (2005) Systematic review of microsatellite instability and colorectal cancer prognosis. *Journal of Clinical Oncology*, **23** (3), 609–618.

82 Merok, M.A., Ahlquist, T., Røyrvik, E.C., Tufteland, K.F., Hektoen, M., Sjo, O.H., Mala, T., Svindland, A., Lothe, R.A., and Nesbakken, A. (2013) Microsatellite instability has a positive prognostic impact on stage II colorectal cancer after complete resection: results from a large, consecutive Norwegian series. *Annals of Oncology*, **24** (5), 1274–1282.

83 Williams, D.S., Bird, M.J., Jorissen, R.N., Yu, Y.L., Walker, F., Zhang, H.H., Nice, E.C., and Burgess, A.W. (2010) Nonsense mediated decay resistant mutations are a source of expressed mutant proteins in colon cancer cell lines with microsatellite instability. *PLoS One*, **5** (12), e16012.

84 Nosho, K., Baba, Y., Tanaka, N., Shima, K., Hayashi, M., Meyerhardt, J.A., Giovannucci, E., Dranoff, G., Fuchs, C.S., and Ogino, S. (2010) Tumour-infiltrating T-cell subsets, molecular changes in colorectal cancer, and prognosis: cohort study and literature review. *Journal of Pathology*, **222** (4), 350–366.

85 Des Guetz, G., Schischmanoff, O., Nicolas, P., Perret, G.Y., Morere, J.F., and Uzzan, B. (2009) Does microsatellite instability predict the efficacy of adjuvant chemotherapy in colorectal cancer? A systematic review with meta-analysis. *European Journal of Cancer*, **45** (10), 1890–1896.

86 Tajima, A., Iwaizumi, M., Tseng-Rogenski, S., Cabrera, B.L., and Carethers, J.M. (2011) Both hMutSα and hMutSβ DNA mismatch repair complexes participate in 5-fluorouracil cytotoxicity. *PLoS One*, **6** (12), e28117.

87 Park, J.M., Huang, S., Tougeron, D., and Sinicrope, F.A. (2013) MSH3 mismatch repair protein regulates sensitivity to cytotoxic drugs and a histone deacetylase inhibitor in human colon carcinoma cells. *PLoS One*, **8** (5), e65369.

88 Bertagnolli, M.M., Niedzwiecki, D., Compton, C.C., Hahn, H.P., Hall, M., Damas, B., Jewell, S.D., Mayer, R.J., Goldberg, R.M., Saltz, L.B., Warren, R.S., and Redston, M. (2009) Microsatellite instability predicts improved response to adjuvant therapy with irinotecan, fluorouracil, and leucovorin in stage III colon cancer: Cancer and Leukemia Group B Protocol 89803. *Journal of Clinical Oncology*, **27** (11), 1814–1821.

89 Fallik, D., Borrini, F., Boige, V., Viguier, J., Jacob, S., Miquel, C., Sabourin, J.C., Ducreux, M., and Praz, F. (2003) Microsatellite instability is a predictive factor of the tumor response to irinotecan in patients with advanced colorectal cancer. *Cancer Research*, **63** (18), 5738–5744.

90 de la Chapelle, A. and Hampel, H. (2010) Clinical relevance of microsatellite instability in colorectal cancer. *Journal of Clinical Oncology*, **28** (20), 3380–3387.

91 Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group (2009) Recommendations from the EGAPP Working Group: genetic testing strategies in newly diagnosed individuals with colorectal cancer aimed at reducing morbidity and mortality from Lynch syndrome in relatives. *Genetics in Medicine*, **11** (1), 35–41.

92 Goel, A., Nagasaka, T., Hamelin, R., and Boland, C.R. (2010) An optimized pentaplex PCR for detecting DNA mismatch repair-deficient colorectal cancers. *PLoS One*, **5** (2), e9393.

93 Ryslik, G.A., Cheng, Y., Cheung, K.H., Modis, Y., and Zhao, H. (2013) Utilizing protein structure to identify non-random somatic mutations. *BMC Bioinformatics*, **14**, 190.

94 Aoki, K. and Taketo, M.M. (2007) Adenomatous polyposis coli (APC): a multi-functional tumor suppressor gene. *Journal of Cell Science*, **120** (19), 3327–3335.

95 Samowitz, W.S., Slattery, M.L., Sweeney, C., Herrick, J., Wolff, R.K., and Albertsen, H. (2007) APC mutations and other genetic and epigenetic changes in colon cancer. *Molecular Cancer Research*, **5** (2), 165–170.

96 Voorham, Q.J., Carvalho, B., Spiertz, A.J., van Grieken, N.C., Mongera, S., Rondagh, E.J., van de Wiel, M.A., Jordanova, E.S.,

Ylstra, B., Kliment, M., Grabsch, H., Rembacken, B.J., Arai, T., deBruïne, A.P., Sanduleanu, S., Quirke, P., Mulder, C.J., van Engeland, M., and Meijer, G.A. (2012) Chromosome 5q loss in colorectal flat adenomas. *Clinical Cancer Research*, **18** (17), 4560–4569.

97 Conlin, A., Smith, G., Carey, F.A., Wolf, C.R., and Steele, R.J. (2005) The prognostic significance of K-ras, p53, and APC mutations in colorectal carcinoma. *Gut*, **54** (9), 1283–1286.

98 Chen, S.P., Wu, C.C., Lin, S.Z., Kang, J.C., Su, C.C., Chen, Y.L., Lin, P.C., Chiu, S.C., Pang, C.Y., and Harn, H.J. (2009) Prognostic significance of interaction between somatic APC mutations and 5-fluorouracil adjuvant chemotherapy in Taiwanese colorectal cancer subjects. *American Journal of Clinical Oncology*, **32** (2), 122–126.

99 Chen, T.H., Chang, S.W., Huang, C.C., Wang, K.L., Yeh, K.T., Liu, C.N., Lee, H., Lin, C.C., and Cheng, Y.W. (2013) The prognostic significance of APC gene mutation and miR-21 expression in advanced stage colorectal cancer. *Colorectal Disease*, **15** (11), 1367–1374.

100 Iacopetta, B., Russo, A., Bazan, V., Dardanoni, G., Gebbia, N., Soussi, T., Kerr, D., Elsaleh, H., Soong, R., Kandioler, D., Janschek, E., Kappel, S., Lung, M., Leung, C.S., Ko, J.M., Yuen, S., Ho, J., Leung, S.Y., Crapez, E., Duffour, J., Ychou, M., Leahy, D.T., O'Donoghue, D.P., Agnese, V., Cascio, S., Di Fede, G., Chieco-Bianchi, L., Bertorelle, R., Belluco, C., Giaretti, W., Castagnola, P., Ricevuto, E., Ficorella, C., Bosari, S., Arizzi, C.D., Miyaki, M., Onda, M., Kampman, E., Diergaarde, B., Royds, J., Lothe, R.A., Diep, C.B., Meling, G.I., Ostrowski, J., Trzeciak, L., Guzinska-Ustymowicz, K., Zalewski, B., Capellá, G.M., Moreno, V., Peinado, M.A., Lönnroth, C., Lundholm, K., Sun, X.F., Jansson, A., Bouzourene, H., Hsieh, L.L., Tang, R., Smith, D.R., Allen-Mersh, T.G., Khan, Z.A., Shorthouse, A.J., Silverman, M.L., Kato, S., Ishioka, C., and Group, T.-C.C. (2006) Functional categories of TP53 mutation in colorectal cancer: results of an International Collaborative Study. *Annals of Oncology*, **17** (5), 842–847.

101 Russo, A., Bazan, V., Iacopetta, B., Kerr, D., Soussi, T., Gebbia, N., and Group, T.-C.C.S. (2005) The TP53 colorectal cancer international collaborative study on the prognostic and predictive significance of p53 mutation: influence of tumor site, type of mutation, and adjuvant treatment. *Journal of Clinical Oncology*, **23** (30), 7518–7528.

102 Duffy, M.J., van Dalen, A., Haglund, C., Hansson, L., Holinski-Feder, E., Klapdor, R., Lamerz, R., Peltomaki, P., Sturgeon, C., and Topolcan, O. (2007) Tumour markers in colorectal cancer: European Group on Tumour Markers (EGTM) guidelines for clinical use. *European Journal of Cancer*, **43** (9), 1348–1360.

103 Petitjean, A., Mathe, E., Kato, S., Ishioka, C., Tavtigian, S.V., Hainaut, P., and Olivier, M. (2007) Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Human Mutation*, **28** (6), 622–629.

104 Lee, M.K., Teoh, W.W., Phang, B.H., Tong, W.M., Wang, Z.Q., and Sabapathy, K. (2012) Cell-type, dose, and mutation-type specificity dictate mutant p53 functions *in vivo*. *Cancer Cell*, **22** (6), 751–764.

105 Liu, X., Jakubowski, M., and Hunt, J.L. (2011) KRAS gene mutation in colorectal cancer is correlated with increased proliferation and spontaneous apoptosis. *American Journal of Clinical Pathology*, **135** (2), 245–252.

106 Imamura, Y., Morikawa, T., Liao, X., Lochhead, P., Kuchiba, A., Yamauchi, M., Qian, Z.R., Nishihara, R., Meyerhardt, J.A., Haigis, K.M., Fuchs, C.S., and Ogino, S. (2012) Specific mutations in KRAS codons 12 and 13, and patient prognosis in 1075 BRAF wild-type colorectal cancers. *Clinical Cancer Research*, **18** (17), 4753–4763.

107 Tejpar, S., Bertagnolli, M., Bosman, F., Lenz, H.J., Garraway, L., Waldman, F., Warren, R., Bild, A., Collins-Brennan, D., Hahn, H., Harkin, D.P., Kennedy, R., Ilyas, M., Morreau, H., Proutski, V.,

Swanton, C., Tomlinson, I., Delorenzi, M., Fiocca, R., Van Cutsem, E., and Roth, A. (2010) Prognostic and predictive biomarkers in resected colon cancer: current status and future perspectives for integrating genomics into biomarker discovery. *Oncologist*, **15** (4), 390–404.

108 Peeters, M., Douillard, J.Y., Van Cutsem, E., Siena, S., Zhang, K., Williams, R., and Wiezorek, J. (2013) Mutant KRAS codon 12 and 13 alleles in patients with metastatic colorectal cancer: assessment as prognostic and predictive biomarkers of response to panitumumab. *Journal of Clinical Oncology*, **31** (6), 759–765.

109 Michaloglou, C., Vredeveld, L.C., Mooi, W.J., and Peeper, D.S. (2008) BRAF (E600) in benign and malignant human tumours. *Oncogene*, **27** (7), 877–895.

110 Wang, L., Cunningham, J.M., Winters, J.L., Guenther, J.C., French, A.J., Boardman, L.A., Burgart, L.J., McDonnell, S.K., Schaid, D.J., and Thibodeau, S.N. (2003) BRAF mutations in colon cancer are not likely attributable to defective DNA mismatch repair. *Cancer Research*, **63** (17), 5209–5212.

111 Young, J., Jenkins, M., Parry, S., Young, B., Nancarrow, D., English, D., Giles, G., and Jass, J. (2007) Serrated pathway colorectal cancer in the population: genetic consideration. *Gut*, **56** (10), 1453–1459.

112 Weisenberger, D.J., Siegmund, K.D., Campan, M., Young, J., Long, T.I., Faasse, M.A., Kang, G.H., Widschwendter, M., Weener, D., Buchanan, D., Koh, H., Simms, L., Barker, M., Leggett, B., Levine, J., Kim, M., French, A.J., Thibodeau, S.N., Jass, J., Haile, R., and Laird, P.W. (2006) CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nature Genetics*, **38** (7), 787–793.

113 Bond, C.E., Umapathy, A., Ramsnes, I., Greco, S.A., Zhen Zhao, Z., Mallitt, K.A., Buttenshaw, R.L., Montgomery, G.W., Leggett, B.A., and Whitehall, V.L. (2012) p53 mutation is common in microsatellite stable, BRAF mutant colorectal cancers. *International Journal of Cancer*, **130** (7), 1567–1576.

114 Bond, C.E., Umapathy, A., Buttenshaw, R. L., Wockner, L., Leggett, B.A., and Whitehall, V.L. (2012) Chromosomal instability in BRAF mutant, microsatellite stable colorectal cancers. *PLoS One*, **7** (10), e47483.

115 Samowitz, W.S., Sweeney, C., Herrick, J., Albertsen, H., Levin, T.R., Murtaugh, M. A., Wolff, R.K., and Slattery, M.L. (2005) Poor survival associated with the BRAF V600E mutation in microsatellite-stable colon cancers. *Cancer Research*, **65** (14), 6063–6069.

116 Phipps, A.I., Buchanan, D.D., Makar, K. W., Burnett-Hartman, A.N., Coghill, A.E., Passarelli, M.N., Baron, J.A., Ahnen, D.J., Win, A.K., Potter, J.D., and Newcomb, P.A. (2012) BRAF mutation status and survival after colorectal cancer diagnosis according to patient and tumor characteristics. *Cancer Epidemiology, Biomarkers & Prevention*, **21** (10), 1792–1798.

117 Yuan, Z.X., Wang, X.Y., Qin, Q.Y., Chen, D.F., Zhong, Q.H., Wang, L., and Wang, J.P. (2013) The prognostic role of BRAF mutation in metastatic colorectal cancer receiving anti-EGFR monoclonal antibodies: a meta-analysis. *PLoS One*, **8** (6), e65995.

118 Vasudevan, K.M., Barbie, D.A., Davies, M.A., Rabinovsky, R., McNear, C.J., Kim, J.J., Hennessy, B.T., Tseng, H., Pochanard, P., Kim, S.Y., Dunn, I.F., Schinzel, A.C., Sandy, P., Hoersch, S., Sheng, Q., Gupta, P.B., Boehm, J.S., Reiling, J.H., Silver, S., Lu, Y., Stemke-Hale, K., Dutta, B., Joy, C., Sahin, A.A., Gonzalez-Angulo, A.M., Lluch, A., Rameh, L.E., Jacks, T., Root, D.E., Lander, E.S., Mills, G.B., Hahn, W.C., Sellers, W.R., and Garraway, L.A. (2009) AKT-independent signaling downstream of oncogenic PIK3CA mutations in human cancer. *Cancer Cell*, **16** (1), 21–32.

119 Rosty, C., Young, J.P., Walsh, M.D., Clendenning, M., Sanderson, K., Walters, R.J., Parry, S., Jenkins, M.A., Win, A.K., Southey, M.C., Hopper, J.L., Giles, G.G., Williamson, E.J., English, D.R., and Buchanan, D.D. (2013) PIK3CA activating mutation in colorectal carcinoma: associations with molecular

features and survival. *PLoS One*, **8** (6), e65479.

120 Liao, X., Morikawa, T., Lochhead, P., Imamura, Y., Kuchiba, A., Yamauchi, M., Nosho, K., Qian, Z.R., Nishihara, R., Meyerhardt, J.A., Fuchs, C.S., and Ogino, S. (2012) Prognostic role of PIK3CA mutation in colorectal cancer: cohort study and literature review. *Clinical Cancer Research*, **18** (8), 2257–2268.

121 Sartore-Bianchi, A., Martini, M., Molinari, F., Veronese, S., Nichelatti, M., Artale, S., Di Nicolantonio, F., Saletti, P., De Dosso, S., Mazzucchelli, L., Frattini, M., Siena, S., and Bardelli, A. (2009) PIK3CA mutations in colorectal cancer are associated with clinical resistance to EGFR-targeted monoclonal antibodies. *Cancer Research*, **69** (5), 1851–1857.

122 Irahara, N., Baba, Y., Nosho, K., Shima, K., Yan, L., Dias-Santagata, D., Iafrate, A. J., Fuchs, C.S., Haigis, K.M., and Ogino, S. (2010) NRAS mutations are rare in colorectal cancer. *Diagnostic Molecular Pathology: The American Journal of Surgical Pathology, Part B*, **19** (3), 157–163.

123 Vaughn, C.P., Zobell, S.D., Furtado, L.V., Baker, C.L., and Samowitz, W.S. (2011) Frequency of KRAS, BRAF, and NRAS mutations in colorectal cancer. *Genes, Chromosomes & Cancer*, **50** (5), 307–312.

124 Nosho, K., Irahara, N., Shima, K., Kure, S., Kirkner, G.J., Schernhammer, E.S., Hazra, A., Hunter, D.J., Quackenbush, J., Spiegelman, D., Giovannucci, E.L., Fuchs, C.S., and Ogino, S. (2008) Comprehensive biostatistical analysis of CpG island methylator phenotype in colorectal cancer using a large population-based sample. *PLoS One*, **3** (11), e3698.

125 Tsiatis, A.C., Norris-Kirby, A., Rich, R.G., Hafez, M.J., Gocke, C.D., Eshleman, J.R., and Murphy, K.M. (2010) Comparison of Sanger sequencing, pyrosequencing, and melting curve analysis for the detection of KRAS mutations: diagnostic and clinical implications. *Journal of Molecular Diagnostics*, **12** (4), 425–432.

126 Lurkin, I., Stoehr, R., Hurst, C.D., van Tilborg, A.A., Knowles, M.A., Hartmann, A., and Zwarthoff, E.C. (2010) Two multiplex assays that simultaneously identify 22 possible mutation sites in the KRAS, BRAF, NRAS and PIK3CA genes. *PLoS One*, **5** (1), e8802.

127 Ogino, S., Kawasaki, T., Brahmandam, M., Yan, L., Cantor, M., Namgyal, C., Mino-Kenudson, M., Lauwers, G.Y., Loda, M., and Fuchs, C.S. (2005) Sensitive sequencing method for KRAS mutation detection by pyrosequencing. *Journal of Molecular Diagnostics*, **7** (3), 413–421.

128 Domingo, E., Church, D.N., Sieber, O., Ramamoorthy, R., Yanagisawa, Y., Johnstone, E., Davidson, B., Kerr, D.J., Tomlinson, I.P., and Midgley, R. (2013) Evaluation of PIK3CA mutation as a predictor of benefit from nonsteroidal anti-inflammatory drug therapy in colorectal cancer. *Journal of Clinical Oncology*, **31** (34), 4297–4305.

129 Esteller, M. (2007) Cancer epigenomics: DNA methylomes and histone-modification maps. *Nature Reviews Genetics*, **8** (4), 286–298.

130 Costello, J.F. and Plass, C. (2001) Methylation matters. *Journal of Medical Genetics*, **38** (5), 285–303.

131 Weisenberger, D.J., Campan, M., Long, T.I., Kim, M., Woods, C., Fiala, E., Ehrlich, M., and Laird, P.W. (2005) Analysis of repetitive element DNA methylation by MethyLight. *Nucleic Acids Research*, **33** (21), 6823–6836.

132 Feinberg, A.P. and Vogelstein, B. (1983) Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature*, **301** (5895), 89–92.

133 Goelz, S.E., Vogelstein, B., Hamilton, S.R., and Feinberg, A.P. (1985) Hypomethylation of DNA from benign and malignant human colon neoplasms. *Science*, **228** (4696), 187–190.

134 Hur, K., Cejas, P., Feliu, J., Moreno-Rubio, J., Burgos, E., Boland, C.R., and Goel, A. (2014) Hypomethylation of long interspersed nuclear element-1 (LINE-1) leads to activation of proto-oncogenes in human colorectal cancer metastasis. *Gut*, **63** (4), 635–646.

135 Yang, A.S., Estécio, M.R., Doshi, K., Kondo, Y., Tajara, E.H., and Issa, J.P. (2004) A simple method for estimating

global DNA methylation using bisulfite PCR of repetitive DNA elements. *Nucleic Acids Research*, **32** (3), e38.

136 Sunami, E., de Maat, M., Vu, A., Turner, R.R., and Hoon, D.S. (2011) LINE-1 hypomethylation during primary colon cancer progression. *PLoS One*, **6** (4), e18884.

137 Ogino, S., Kawasaki, T., Nosho, K., Ohnishi, M., Suemoto, Y., Kirkner, G.J., and Fuchs, C.S. (2008) LINE-1 hypomethylation is inversely associated with microsatellite instability and CpG island methylator phenotype in colorectal cancer. *International Journal of Cancer*, **122** (12), 2767–2773.

138 Kawakami, K., Matsunoki, A., Kaneko, M., Saito, K., Watanabe, G., and Minamoto, T. (2011) Long interspersed nuclear element-1 hypomethylation is a potential biomarker for the prediction of response to oral fluoropyrimidines in microsatellite stable and CpG island methylator phenotype-negative colorectal cancer. *Cancer Science*, **102** (1), 166–174.

139 Toyota, M., Ahuja, N., Ohe-Toyota, M., Herman, J.G., Baylin, S.B., and Issa, J.P. (1999) CpG island methylator phenotype in colorectal cancer. *Proceedings of the National Academy of Sciences of the United States of America*, **96** (15), 8681–8686.

140 Wong, J.J., Hawkins, N.J., Ward, R.L., and Hitchins, M.P. (2011) Methylation of the 3p22 region encompassing MLH1 is representative of the CpG island methylator phenotype in colorectal cancer. *Modern Pathology*, **24** (3), 396–411.

141 Karpinski, P., Ramsey, D., Grzebieniak, Z., Sasiadek, M.M., and Blin, N. (2008) The CpG island methylator phenotype correlates with long-range epigenetic silencing in colorectal cancer. *Molecular Cancer Research*, **6** (4), 585–591.

142 Jo, P., Jung, K., Grade, M., Conradi, L.C., Wolff, H.A., Kitz, J., Becker, H., Rüschoff, J., Hartmann, A., Beissbarth, T., Müller-Dornieden, A., Ghadimi, M., Schneider-Stock, R., and Gaedcke, J. (2012) CpG island methylator phenotype infers a poor disease-free survival in locally advanced rectal cancer. *Surgery*, **151** (4), 564–570.

143 Dahlin, A.M., Palmqvist, R., Henriksson, M.L., Jacobsson, M., Eklöf, V., Rutegård, J., Oberg, A., and Van Guelpen, B.R. (2010) The role of the CpG island methylator phenotype in colorectal cancer prognosis depends on microsatellite instability screening status. *Clinical Cancer Research*, **16** (6), 1845–1855.

144 Jover, R., Nguyen, T.P., Pérez-Carbonell, L., Zapater, P., Payá, A., Alenda, C., Rojas, E., Cubiella, J., Balaguer, F., Morillas, J.D., Clofent, J., Bujanda, L., Reñé, J.M., Bessa, X., Xicola, R.M., Nicolás-Pérez, D., Castells, A., Andreu, M., Llor, X., Boland, C.R., and Goel, A. (2011) 5-Fluorouracil adjuvant chemotherapy does not increase survival in patients with CpG island methylator phenotype colorectal cancer. *Gastroenterology*, **140** (4), 1174–1181.

145 Kaneda, A. and Yagi, K. (2011) Two groups of DNA methylation markers to classify colorectal cancer into three epigenotypes. *Cancer Science*, **102** (1), 18–24.

146 Shen, L., Toyota, M., Kondo, Y., Lin, E., Zhang, L., Guo, Y., Hernandez, N.S., Chen, X., Ahmed, S., Konishi, K., Hamilton, S.R., and Issa, J.P. (2007) Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. *Proceedings of the National Academy of Sciences of the United States of America*, **104** (47), 18654–18659.

147 Karpinski, P., Walter, M., Szmida, E., Ramsey, D., Misiak, B., Kozlowska, J., Bebenek, M., Grzebieniak, Z., Blin, N., Laczmanski, L., and Sasiadek, M.M. (2013) Intermediate- and low-methylation epigenotypes do not correspond to CpG island methylator phenotype (low and zero) in colorectal cancer. *Cancer Epidemiology, Biomarkers & Prevention*, **22** (2), 201–208.

148 Hughes, L.A., Melotte, V., de Schrijver, J., de Maat, M., Smit, V.T., Bovée, J.V., French, P.J., van den Brandt, P.A., Schouten, L.J., de Meyer, T., van Criekinge, W., Ahuja, N., Herman, J.G., Weijenberg, M.P., and van Engeland, M. (2013) The CpG island methylator

phenotype: what's in a name? *Cancer Research*, **73** (19), 5858–5868.

149 Hughes, L.A., Khalid-de Bakker, C.A., Smits, K.M., van den Brandt, P.A., Jonkers, D., Ahuja, N., Herman, J.G., Weijenberg, M.P., and van Engeland, M. (2012) The CpG island methylator phenotype in colorectal cancer: progress and problems. *Biochimica et Biophysica Acta*, **1825** (1), 77–85.

150 De Roock, W., De Vriendt, V., Normanno, N., Ciardiello, F., and Tejpar, S. (2011) KRAS, BRAF, PIK3CA, and PTEN mutations: implications for targeted therapies in metastatic colorectal cancer. *Lancet Oncology*, **12** (6), 594–603.

5
Nucleic Acid-Based Markers in Urologic Malignancies

Bernd Wullich, Peter J. Goebell, Helge Taubert, and Sven Wach

5.1
Introduction

The discovery, analysis, and validation of any biomarker, including nucleic acid-based biomarkers. remains a challenging task. Hundreds of reports on tumor markers have been published to date, addressing a variety of oncologic and clinical questions. Guidelines have now been established that support the discovery and validation of biomarkers in general [1]. However, the results of these studies are often inconsistent and sometimes contradictory. Recognized problems include different methods of performing assays, and the use of different subsets of patients (difference in stage or treatment) and endpoints (e.g., local versus distant recurrence versus survival), leading to incompatible data sets. This has impeded our understanding of the role of new markers. However, there has been a discussion about establishing general methodological principles and guidelines for design, conduct, analysis, and reporting of marker studies (analogous to those for clinical trials) [2–13].

For the case of prostate cancer biomarkers, a recent review [14] proposes the classification of biomarkers in six categories:

1) *Screening/detection.* These biomarkers are useful for assessing any individual's basic risk of developing cancer during their lifetime or for generating evidence that a tumor could be present. Ideally, these markers should be measurable in easy-to-collect biological material such as blood, urine, or buccal swabs.
2) *Diagnostic.* These biomarkers can support the classical histopathological evaluation of tissue samples such as biopsy specimens. Their main advantage is to provide additional information to detect the presence or to prove the absence of cancer.
3) *Prognostic.* Prognostic biomarkers provide information about the long-term outcome of the tumor disease. Depending on the specific endpoints, these markers provide a risk assessment for overall survival, tumor-specific

Nucleic Acids as Molecular Diagnostics, First Edition. Edited by Andreas Keller and Eckart Meese.

survival, or progression of disease, for example. When combined with information about specific treatment modalities, these biomarkers can be helpful for developing individual treatment strategies.

4) *Predictive.* These biomarkers are able to predict the individual patient's response to a specific treatment (drug or other therapy). These markers help in providing the most effective therapy to patients that are responsive while sparing side-effects for those patients that are unlikely to respond. Predictive biomarkers may also be useful in monitoring the effectiveness of the treatment.
5) *Therapeutic targets.* These biomarkers not only identify those patients that will benefit from a particular treatment regimen, they are also the direct molecular targets of novel therapies.
6) *Surrogate endpoint.* These markers represent a substitute for a clinical endpoint or for an indirect assessment of clinical benefit or harm, or the lack of it.

In this chapter, we will focus on the three most prominent urologic malignancies: bladder cancer, prostate cancer, and renal cell carcinoma. For every tumor entity we will provide an overview of the hereditary tumor syndromes and the underlying genetic aberrations followed by biomarkers in the setting of sporadic tumor disease.

5.2 Bladder Cancer

With an estimated number of 91 000 cases in Europe [15] and another estimated 61 000 newly diagnosed patients in the Unites States [16], bladder cancer is one of the most prevalent neoplasms. Bladder tumors comprise a heterogeneous group with respect to both histopathology and clinical behavior: A rough estimate of 30% of these patients present with muscle-invasive disease at initial diagnosis and another 10–15% of those cases with recurrent non-invasive lesions eventually progress during the course of their disease within a 5-year follow-up [17,18].

Histopathologic features, including tumor grade, depth of invasion, multiplicity, size, morphology, presence of vascular or lymphatic invasion, and presence or absence of carcinoma *in situ*, can be related to the risk of recurrence and progression [19]. Although these features represent statistically significant predictors within the bladder cancer population, they do not provide accurate predictions for individual patients. Large series of bladder tumors have been investigated to provide the current molecular information, including details on genetic and epigenetic alterations, and changes in gene expression at the RNA and protein levels. Information on post-translational changes in proteins is also available that may contribute directly or indirectly to the development of heterogeneity of bladder tumors. This information is summarized in recent reviews

[18,20,21] and these suggest that some subclassification can be achieved based on molecular characteristics.

Despite a number of identified environmental risk factors [22,23], there are genetic factors that are either inherited or acquired (sporadic) that contribute to the risk of development and progression of bladder cancer. Here, we will focus only on these hereditary and sporadic factors.

5.2.1 Hereditary Factors for Bladder Cancer

A large variety of inherent genetic changes may have an impact on the host's susceptibility for the development of bladder cancer caused by environmental factors; however, a "classic" linkage of genetic alterations to a higher risk to these inherent changes is still missing. This obstacle may be explained by using the example of smoking and the risk of bladder cancer: the most established risk factor is smoking, accounting for an estimated 50% of cases among men and 35% of cases among women [24]. Despite the fact that smoking cessation reduces the risk of developing bladder cancer immediately, the increased risk is still present even after 25 years. This is currently explained by the fact that a huge variety of carcinogens are present in tobacco smoke, which alter gene expression and damage the DNA. This complexity of carcinogens hampers the identification of a distinct and specific link between substance and effect [25].

Despite many years of research and the identification of several genes involved in bladder cancer pathogenesis, a large part of its heritability remains unknown [26]. This has led to the growing field of "molecular epidemiology" with the implementation of whole-genome sequencing or genome-wide association studies (GWAS) as part of the workflow to identify the possible features responsible for the onset of bladder cancer in susceptible populations [27].

In addition, the investigation of the functional role of newly identified susceptibility loci for bladder cancer remains challenging, as they mostly map to poorly understood, non-coding regions.

5.2.2 Single Nucleotide Polymorphisms

Despite many years of research and the identification of several genes involved in bladder cancer pathogenesis, a large part of its heritability remains unknown. DNA from different individuals is identical for most base positions; however, interindividual variants are observed approximately every 500 bases. If a variant occurs in more than 1% of the population, it is defined as a single nucleotide polymorphism (SNP).

Thus, the complexity of analyses of relevant genes for bladder cancer even increases with the well-known fact that SNPs contribute to the interindividual differences in bladder cancer susceptibility as well as in other cancer types [28–37]. The majority of SNPs relevant for bladder cancer that have been

validated in sufficiently large populations are located next to the following genes: *MYC*, *TP63*, *PSCA*, the *TERT–CLPTM1L* locus, *FGFR3*, *TACC3*, *NAT2*, *CBX6*, *APOBEC3A*, *CCNE1*, and *UGT1A* [33,37–43]. Despite many hypothesis-driven candidate gene studies, the majority of results have not been reproduced, with the exception of *GSTM1* and *NAT2* [44]. In summary, the functions of these genes are closely related to the detoxification of carcinogens, control of the cell cycle, and apoptosis, as well as maintenance of DNA integrity.

Despite their large numbers, the variants identified to date explain only approximately 5–10% of the overall inherited risk and that SNPs are associated with odds ratios smaller than 1.5. Thus, the remainder of currently undetected variants are likely to provide a risk below 1.1, making the justification for preventive measures not very reasonable.

5.2.3 RNA Alterations in Bladder Cancer

As in other malignancies, mRNAs and non-coding RNAs (ncRNAs) have also come to be the focus of more recent research in bladder cancer. As indicated above, ncRNAs comprise various subtypes of RNA including microRNAs (miRNAs). Of specific interest are miRNAs that have been shown to be dysregulated and are predicted to target key proteins in signal transduction relevant in bladder cancer. Some genomic alterations are strongly associated with specific histopathologic tumor features like grade or stage. These include fibroblast growth factor receptor 3 (FGFR3) mutations in papillary tumors and p53 and/or Rb pathway alterations in muscle-invasive lesions. These alterations mark or even define these particular tumors, and are compelling candidates for use in tumor subclassification. However, there are only few studies that reveal the impact of miRNA regulation in these signaling pathways in bladder cancer. In addition to the regulatory function of miRNAs in the process of cellular signaling, these RNAs themselves may be regulated through epigenetic silencing, which would add another level of regulation to the complex model of carcinogenesis of bladder cancer [45,46].

5.2.3.1 FGFR3 Pathway

FGFR3 and other growth factor receptors are susceptible to miRNA regulation during bladder cancer carcinogenesis. Some miRNAs that have been proven to be dysregulated in bladder cancer [47] are predicted by TargetScan 5.1 to have multiple target genes in the FGFR3 pathway. In particular, miR-129 has numerous predicted target genes: platelet-derived growth factor receptor, α polypeptide (*PDGFRA*); platelet-derived growth factor receptor, β polypeptide (*PDGFRB*); epidermal growth factor receptor (*EGFR*); phospholipase C, γ1 (*PLCG1*); SHC (Src homology 2 domain containing) transforming protein 4 (*SHC4*); protein kinase C, α (*PRKCA*); protein kinase C, β (*PRKCB*); protein kinase C, δ (*PRKCD*); protein kinase C, ε (*PRKCE*); protein kinase C, ι (*PRKCI*); growth factor receptor-bound protein 2 (*GRB2*); son of sevenless homolog 1 (*SOS1*);

SOS2; v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (*KRAS*); death-associated protein kinase 1 (*DAPK1*); *DAPK2*; *DAPK3*; and ribosomal protein S6 kinase, 90 kDa, polypeptide 5 (*RPS6KA5*) (also known as *MSK1*) (reviewed in [48]).

Some well-established dysregulated miRNAs in bladder cancer carcinogenesis are predicted to target the *FGFR3* receptor (e.g., miR-145 and miR-101). In addition, the regulation of *FGFR3* by miR-99a and miR-100 has been validated experimentally [49]. The miR-200 family is associated with an epithelial phenotype in bladder cancer cell lines. Ectopic expression of miR-200 in bladder cancer cell lines induces upregulation of epithelial markers and the downregulation of mesenchymal markers. The epithelial phenotype was accompanied by an increased sensitivity to EGFR inhibitors, which was explained by a direct regulation of *ERRFI1* by miR-200 [50]. Upregulation of miR-143 in bladder cancer specimens is accompanied by a downregulation of Ras. In addition, the forced overexpression of miR-143 in bladder cancer cell lines decreased Ras levels [51].

5.2.3.2 p53 Pathway

Several miRNAs that are dysregulated in bladder cancer are predicted by TargetScan to target p53 or its regulators. miR-10, for example, is predicted to target Mdm2 p53-binding protein homolog (mouse) (*MDM2*), Mdm2 p53-binding protein homolog (mouse) (*MDM4*), and ataxia telangiectasia mutated (*ATM*), whereas miR-129 potentially targets the genes *MDM4* and *ATM* in this pathway. Additionally, miR-125b, miR-143, miR-30a/c, and miR-223 are predicted to target *TP53* directly. Despite missing functional data, non-invasive and muscle-invasive carcinomas display specific miRNA patterns that enable investigators to pinpoint the roles of miRNAs in both pathways of bladder carcinogenesis [52].

5.2.3.3 Urine-Based Markers

To investigate diagnostic or prognostic miRNA markers in bladder cancer, most studies use urine samples.

In terms of diagnostic potential, miR-1224-3p showed a specificity of 83% and a sensitivity of 77% in bladder cancer detection. A combination of miR-15b, miR-135b, and miR-1224-3p could detect bladder cancer with a sensitivity of 94.1% and a specificity of 51% [53]. Interestingly, the ratio of miR-126 to miR-152 in pellet sediments of urine distinguished bladder cancer at a specificity of 82% and a sensitivity of 72% [54]. In addition, miR-96 and miR-183 have been demonstrated to be able to distinguish urothelial cancer patients from non-cancer patients with a good sensitivity and specificity (miR-96, 71.0 and 89.2%; miR-183, 74.0 and 77.3%) [55]. These investigators also combined the miR-96 detection data with the urinary cytology data and diagnosed 61 of 78 cases as urothelial cancers, increasing the sensitivity from 43.6 to 78.2%. In another study, patients with bladder cancer had a lower expression of the miR-200 family, miR-192, and miR-155 in the urinary sediment; and lower expression of miR-192 and higher expression of miR-155 in

the urinary supernatant. The expression of the miR-200 family, miR-205, and miR-192 in the urine sediment significantly correlated with urinary expression of epithelial–mesenchymal transition markers, including zinc-finger E-box-binding homeobox 1, vimentin, transforming growth factor-β1, and Ras homolog gene family, member A [56]. Others investigated the assessment of miR-145 levels to distinguish bladder cancer patients from non-cancer controls, and showed a sensitivity of 77.8% and a specificity of 61.1% for non-muscle invasive disease, and a sensitivity of 84.1% and specificity of 61.1% for invasive bladder cancer [57]. In addition, miR-200a was shown to be an independent predictor of recurrence of non-muscle invasive disease, with a higher risk of recurrence among patients with a lower miR-200a level compared with patients with higher miR-200a levels.

The differential expression of miR-143, miR-222, and miR-452 correlated with pathologic features, recurrence (miR-222 and miR-143), progression (miR-222 and miR-143), disease-specific survival (miR-222), and overall survival (miR-222) [58]. Protein expression patterns of potential miRNA targets correlated with recurrence (Erbb3), progression (Erbb4), disease-specific survival (Erbb3 and Erbb4), and overall survival (Erbb3 and Erbb4). In addition, quantitative real-time (qRT)-polymerase chain reaction (PCR)analysis of miR-452 and miR-222 in urine provided high accuracies for bladder cancer diagnosis [58]. Expression of another two miRNAs (miR-618 and miR-1255b-5p) in the urine of patients with invasive tumors was significantly ($P < 0.05$) increased, and urinary miR-1255b-5p expression reached a specificity of 68% and a sensitivity of 85% in diagnosing invasive bladder cancer cases [59].

Another interesting approach appears to be the comprehensive analysis of a set of 22 selected miRNAs with potentially diagnostic and prognostic values cross-validated in a separate cohort of samples yielding the identification of a six-miRNA diagnostic signature (miR-187, miR-18a*, miR-25, miR-142-3p, miR-140-5p, and miR-204) with a sensitivity of 84.8% and specificity of 86.5%, and a two-miRNA prognostic model (miR-92a and miR-125b) with a sensitivity of 84.95% and a specificity of 74.14% [60].

5.2.3.4 Serum-Based Markers

Some plasma miRNAs also showed diagnostic potential for bladder cancer [61]. For example, miR-148b, miR-200b, miR-487, miR-541, and miR-566 were upregulated in the plasma of bladder cancer patients, whereas expression of miR-25, miR-33b, miR-92a/b, and miR-302 was significantly downregulated. The predictive powers of these miRNAs were at an accuracy of 89% for discriminating bladder cancer from normal controls and 92% for distinguishing invasive bladder cancer from other cases [61]. Apparently, there is also data on another group of miRNAs (miR-26b-5p, miR-144-5p, and miR-374-5p) that have been shown to be significantly upregulated in invasive bladder tumor patients when compared with the control group, and miR-26b-5p detected the presence of invasive bladder tumors with 94% specificity and 65% sensitivity [59].

5.2.4 Sporadic Factors for Bladder Cancer

As indicated above, some genomic alterations are strongly associated with specific histopathologic tumor features, such as grade or stage. FGFR3 mutations in papillary tumors and p53 and/or Rb pathway alterations in muscle-invasive lesions may mark or even define these particular tumors and are implemented in tumor subclassification.

The current well-accepted "classic" model for bladder cancer implements at least two major molecular pathways based on the existence of two distinct groups of lesions: non-invasive papillary tumors and muscle-invasive tumors. Whereas the existence of these major pathogenesis pathways is supported by a large body of clinical, histopathologic, and molecular information, the origin of high-grade papillary Ta tumors remains uncertain. These lesions may arise from flat dysplasia or via an increase in grade from low-grade tumors. Similarly, the pathway (or pathways) to the development of T1 tumors remains unclear as they frequently present as T1 tumors with no prior history (or at least with no documented previous event or histology). Thus, there appears to be a third group of lesions with different clinicopathologic features that do not fall into either of these two "classic" categories. In addition, recent investigations support the possibility of two distinct types of invasive lesions with a different heritage.

The following aims to elucidate the current model for the pathogenesis of the major groups of bladder tumors describing the supportive evidence through information on genetic alterations based on these categories.

5.2.5 Genetic Changes in Non-Invasive Papillary Urothelial Carcinoma

Non-invasive papillary tumors (Ta) represent the largest group of bladder cancer cases at diagnosis. The genetic changes include alterations in oncogenes: activating mutation in *HRAS* ((11p15)/*NRAS* (1p13)/*KRAS2* (12p12)) and *FGFR3* (4p16) as well as amplification and/or overexpression of *CCND1* (11q13), activating mutation of *PIK3CA* (3q26) or overexpression of *MDM2* (12q13). In addition, alterations of tumor suppressor genes also play an important role in this subgroup of lesions. These include homozygous deletion or methylation or mutation of *CDKN2A* (9p21), deletion or mutation in *PTCH* (9q22), deletion or methylation of *DBC1* (9q32–33) as well as deletion or mutation of *TSC1* (9q34).

However, the most mature data on mutations in two protooncogenes, *FGFR3* and *PIK3CA*, may provide a substantial basis on which to study and subclassify this subgroup of tumors.

5.2.5.1 FGFR 3

The FGFR family consists of four active members of high-affinity cell surface-associated receptors (FGFR1–4) that are highly conserved both within the family

and throughout evolution [62]. These receptors have a common structure, consisting of an extracellular domain that includes an N-terminal hydrophobic signal peptide followed by three immunoglobulin (Ig)-like domains, a hydrophobic transmembrane domain, and an intracellular tyrosine kinase domain. Fibroblast growth factors (FGFs), the ligands for FGFRs, bind to extracellular Ig-like domains II and III, resulting in downstream signaling.

Since the initial analyses by Cappellen *et al.* [63], who identified mutations in nine of 26 (35%) bladder and three of 12 (25%) cervical carcinomas, there have been several studies that have identified 11 different mutations [64–75]. Mutation of FGFR3 in bladder cancer is strongly associated with low tumor grade and stage [64,73,76–79]. This strong association also led to investigations of urothelial papillomas [68], which are proposed precursor lesions for low-grade papillary bladder tumors. In this study, nine out of 12 urothelial papillomas contained mutations. This group also investigated 79 pTaG1 tumors, comprising 62 papillary urothelial neoplasms of low malignant potential and 17 low-grade papillary lesions according to the 2004 WHO classification [80], and found mutations in 86%. This supports the association of FGFR3 mutation with low-risk bladder cancer, and indicates that there is a strong molecular relationship between urothelial papilloma and low-grade bladder cancer.

Recently, FGFR3 protein expression was assessed in a population-based series of tumors of known mutation status [81]. Interestingly, many tumors, including 38% of muscle-invasive tumors with no detectable mutation, showed overexpression [81]. In these non-mutant tumors, FGFR3 signaling may contribute to the transformed phenotype via ligand-independent dimerization of the overexpressed protein, increased expression of ligand, or via differential splicing that generates a variant with altered ligand specificity. If both mutation and overexpression of this receptor contribute to disease pathogenesis, these findings implicate FGFR3 alterations in 80% of pTa and 51% of pT2 tumors.

One of the predicted effects of FGFR3 mutation is activation of the RAS–mitogen-activated protein kinase (MAPK) pathway [82]. Investigations on the relationship between *FGFR3* and *RAS* gene mutations in bladder cancer demonstrated that 68% of the tumors studied had mutation of either a *RAS* gene or *FGFR3* [76]. This accounted for 37 of 45 (82%) of grade 1 tumors and 32 of 39 (82%) of pTa tumors, but only 44% of grade 3 and 39% of pT2 tumors. Examination of the distribution of *RAS* and *FGFR3* mutations revealed that these events were mutually exclusive in both tumors and cell lines.

5.2.5.2 Changes in the Phosphatidylinositol 3-Kinase Pathway

Recent studies in a wide range of human cancers indicate that the phosphatidylinositol 3-kinase (PI3K) pathway is activated by a range of mechanisms. In bladder tumors, four mechanisms of pathway activation have been described: (i) mutational activation of the p110-α catalytic subunit of PI3K (*PIK3CA*), (ii) inactivation of *TSC1*, (iii) mutation of *AKT1* and (iv) inactivation of the tumor suppressor gene *PTEN*, which occurs most frequently in invasive tumors and thus will be discussed separately.

As genomic alterations in these genes are not mutually exclusive, combined mutations may have additive or synergistic effects and the non-canonical effects of these components may be important drivers of bladder cancer. Thus, a comprehensive approach investigating all alterations at the same time in the same set of lesions may help to identify their functional correlations.

1) Mutations of *PIK3CA* have been identified in a significant proportion of bladder cancer lesions, including many Ta tumors [83]. In this latter study, exons 9 and 20 that contain the mutation hotspot codons of *PIK3CA* were sequenced and mutations found in 13% of a panel of 87 tumors representing the whole spectrum of disease (65.5% Ta, 11.5% T1, 23% ≥T2). An association with low tumor grade and stage was found, but was less striking than that found for *FGFR3*, with mutations identified in approximately 20% of Ta tumors. In a second study, which examined a different distribution of tumor stages (35.4% Ta, 39.2% T1, 25.4% ≥T2), mutations were found in 25% analyzed for exon 9 and 20 mutations [84]. Interestingly, the *PIK3CA* mutation spectrum in bladder cancer differs significantly from that in other cancers [84]. Mutations E542K and E545K in the helical domain are most common (24% and 52%, respectively) and the kinase domain mutation H1047R, which is the most common mutation in other cancers, is less common in bladder tumors (46% compared with 13%). In total, nine different mutations have been found to date. Although FGFR3 is not known to activate the PI3K pathway in urothelial cells, an association of *FGFR3* mutation with *PIK3CA* mutation has been reported [83]. Eighteen of 69 (26%) *FGFR3* mutant tumors were also *PIK3CA* mutant compared with four of 58 (6.9%) *FGFR3*-(wild-type) tumors. A later study could not fully support this finding as the correlation did not reach statistical significance [84]. Despite this, the investigation of the functional relationship between these events remains an intriguing topic, and it will be of great interest to examine clinical outcome in tumors with both *PIK3CA* and *FGFR3* mutations. For the association of *RAS* and *PIK3CA* mutations, the current data is also inconclusive, although a small number of lesions harbor both alterations [84]. In the future, these potential interactions should be evaluated in larger series implementing recent developments of low-cost and high-throughput strategies [85].
2) *TSC1* is one of two genes that when mutated in the germline cause the syndrome tuberous sclerosis complex (TSC), characterized by the development of hamartomas in many organs [86]. *TSC1* (9q34) encodes hamartin [87] and *TSC2* (16p13.3) encodes tuberin [88]. Tumor development in TSC patients is thought to occur as the result of a somatic "second hit" in one of these genes. Although epithelial malignancy is not a common feature of TSC, loss of function of *TSC1* has been identified in bladder cancer [84,89,90]. No other human cancers have been shown to contain mutations in either of the TSC genes. More than 50% of bladder tumors of all grades and stages show loss of heterozygosity (LOH) for markers on

chromosome 9 [91] and the *TSC1* locus at 9q34 is a common critical region of deletion [89,92–94]. Studies to date indicate an overall *TSC1* mutation frequency of 14.5% in bladder cancers, including nonsense (35%), missense (26%), frameshift (26%), in-frame deletions (3%), and splicing (10%) mutations [84,95]. Missense mutations have been shown to cause loss of function through aberrant splicing, protein instability, or protein mislocalization [96]. Immunohistochemical analysis of hamartin expression in bladder cancer tissues reveals a range of levels of protein expression in tumors that do not carry a mutation, indicating that other mechanisms of regulation of expression may play a role in reducing TSC1 function in urothelial cancer [84]. The widely studied functions of TSC1 and TSC2 are attributed to the TSC1/TSC2 complex that regulates mammalian target of rapamycin (mTOR) activity via Rheb. Although some independent functions have been ascribed to TSC2, independent functions for TSC1 are not clear. The finding of mutations in bladder but not other cancers and the lack of mutual exclusivity of *TSC1* mutations with either *PIK3CA* or *PTEN* alterations may indicate that TSC1 has an independent function in the urothelium [84].

3) AKT1 is one of the three family members of the serine/threonine kinases that are activated by PDK1 and TORC2. Recently, an activating mutation in the pleckstrin homology domain of AKT1 was identified in breast, colorectal, and ovarian cancers [97]. Glu17 is located in the phosphoinositide-binding pocket, and plays a role in activation of AKT1. The change to lysine introduces a positive change that results in aberrant translocation to the membrane, resulting in constitutive downstream signaling. The same mutation has been identified in bladder cancer [98].

5.2.6
Genetic Changes in Muscle-Invasive Urothelial Carcinoma

Of the many genomic alterations that have been reported in muscle-invasive bladder tumors, some involve known genes and others are non-random copy number alterations. These include alterations in oncogenes: activating mutation in *HRAS* (11p15)/*NRAS* (1p13) and *KRAS2* (12p12), activating mutations/overexpression of *FGFR3* (4p16), amplification and/or overexpression of *ERBB2* (17q), amplification and/or overexpression of *CCND1* (11q13), amplification and/or overexpression of *MDM2* (12q13) as well as amplification and overexpression of *E2F3* (6p22). In addition, alterations of tumor suppressor genes also play an important role in this subgroup of lesions. These include homozygous deletion or methylation or mutation of *CDKN2A* (9p21), deletion or mutation in *PTCH* (9q22), deletion or methylation of *DBC1* (9q32–33), deletion or methylation of *TSC1* (9q34), homozygous deletion or mutation of *PTEN*, deletion of *RB1* (13q14) as well as deletion or mutation of *TP53* (17p13).

Apart from studies of key genes such as *TP53*, *RB1*, *CDKN2A*, and *PTEN*, most of the alterations identified in this group cannot be related to outcome or

response to treatment and their role needs further refinement. Despite the fact that some of these alterations can also be found in Ta tumors, the majority of these alterations are more frequent in invasive tumors.

5.2.6.1 TP53, RB, and Cell Cycle Control Genes

One of the most striking genetic differences between Ta and T2 bladder tumors is in the frequency at which the p53 and Rb pathways are inactivated. *TP53* is inactivated by mutation in many T2 tumors (≥40%) [99–101]. In human tumors, somatic mutations occur in a wide variety of codons of the gene [102,103]. Most are missense point mutations (80% or greater). A fraction of mutations lead to a premature stop codon and a truncated non-functional protein. The most common missense mutations cluster in exons 5–8, coding for the DNA-binding domain of the protein, required for its transcriptional regulatory activity. Among other sources, a detailed compilation of mutants in relationship to tumor type and clinico-pathologic characteristics has been established at the International Agency for Research on Cancer (p53.iarc.fr) [104] and at Institute Curie (p53.free.fr) in France.

As p53 nuclear overexpression is often associated with gene mutations, it has long been used as a surrogate for mutational inactivation, but it is clear that some mutant forms do not show an increased half-life and that stabilization can occur via other mechanisms. Despite the limited number of hotspot codons accounting for 30% of the mutations, the remainder displays a wide distribution with an overall reported number of over 1000 different mutations. Together with the notion that mutant p53 proteins may acquire additional functions [105–107], the interpretation of *TP53* mutation (and p53 overexpression) becomes extremely complex since different mutations harbor distinct functional properties. In addition, it must be noted that most mutations in the DNA-binding domain lead to the synthesis of dominant-negative proteins. Thus, a simple classification of human tumors as either wild-type or mutant is problematic [108].

A more recent analysis on the prognostic impact in muscle-invasive tumors of both mutation (including all coding sequences) and stabilized protein status (assessed by immunohistochemistry) merits attention [109]. This is one of the largest reports on p53 mutation and protein status, and unlike most studies examining *TP53* gene mutations, all coding exons were included. This group demonstrated that 51% of *TP53* mutant and 27% of *TP53* wild-type tumors showed protein accumulation, indicating that the concordance of both analyses is limited, as has been reported earlier [110,111]. In this study, three prognostic groups were defined within a cohort of 150 cystectomy patients: patients with abnormal findings in both assays had the worst prognosis, those with normal findings in both assays had the best prognosis, and those with one abnormal and one normal finding had an intermediate probability of recurrence and survival. p53 molecular information was an independent predictor only among patients without lymph node involvement. Significantly, patients with stabilized but non-mutated protein showed a higher risk of progression than those with wild-type p53 and normal staining pattern. The finding that certain mutations do not

appear to have any effect on clinical outcome suggests that these mutations may not result in functional inactivation of the p53 protein, at least concerning tumor progression. Other alterations that affect function of the p53 pathway include overexpression of *HDM2* [112–115], perturbations of *INK4A* regulation [116–120], and loss of p21 expression [121–123].

The Rb pathway also shows frequent alteration in muscle-invasive tumors. Although mutation screening of *RB1* has not been carried out, LOH of the gene region is common [124]. Altered Rb protein expression including upregulation compared with normal, which is associated with loss of the p16 tumor suppressor [125], has been found in about 54% of cystectomy tumor samples [126]. Similar data were reported by Chatterjee *et al.* [122]. Such alterations are infrequent in Ta tumors [127,128]. Undoubtedly, the p53 and Rb pathways are major determinants of the phenotype of invasive bladder tumors, and the potential for subclassification according to mechanism and genes implicated has already been shown. Inactivation of the Rb pathway and overexpression of a gene or genes on 6p22 appear to be cooperating events in a subgroup of invasive tumors: high-level amplification of a region on 6p22 has been identified in around 9% of invasive tumors [129]. Of these, all had lost Rb expression apart from one, which had lost p16 expression. All tumor cell lines with amplification also showed loss of Rb expression. The minimum region of amplification contains two candidate genes, *E2F3* and *CDKAL1* (a novel gene of unknown function) [129–132]. 6p22 amplification is associated with a higher proliferative index [131] and recent studies implicate *E2F3* as the critical gene in this amplicon that affects cell proliferation [129,133,134]. As more than 30% of muscle-invasive tumors show loss of expression of Rb [127,128,135], these data suggest that there may be two distinct molecular subgroups of Rb-null tumors, one with 6p22 amplification and the other without.

5.2.6.2 Other Genomic Alterations

Another genomic region of particular interest in the invasive tumor group is 8p. Allelic losses, amplification, and rearrangement of this chromosome arm are common in aggressive tumors of several tissue types including bladder [136–139], although to date no genes have been adequately validated as either oncogenes or tumor suppressor genes [140–143]. The high frequency at which 8p alterations are seen in invasive bladder tumors implies the acquisition of a selective advantage and it will be of great interest to identify genes of functional significance in this region. Changes in 8q have also been related to bladder cancer [144,145].

Deletions on 10q are found at a higher frequency (24%–58%) in invasive than in non-invasive tumors [146–148]. The deleted region includes the tumor suppressor gene *PTEN* – a key negative regulator of the PI3K pathway. Mutations are not detected frequently on the retained allele of *PTEN* [149,150], although occasional cases with homozygous deletion of the entire gene or part of it are reported [148]. It is not clear what the consequences of the loss of one allele in urothelial cells are, but the finding of LOH only in invasive tumors and the known haploinsufficiency in other tissues [151,152] suggests that it may play an

important role. Functional studies indicate that *PTEN* has a profound effect on the invasive characteristics of bladder tumor cells [153] and in a urothelium-specific Pten-null mouse model, all mice showed urothelial hyperplasia and some developed bladder cancer [154]. As described above, *PIK3CA* and *TSC1* mutations are also detected in some invasive lesions [83,84,95]. The relationship of these alterations remains unclear and it could be speculated that different mechanisms of pathway dysregulation have different phenotypic consequences. Recent results show that these three events are not mutually exclusive, suggesting possible distinct effects roles of different mechanisms of pathway activation [84].

The receptor tyrosine kinase gene *ERBB2* is amplified and/or overexpressed at the protein level in some bladder tumors. There is a large amount of literature on this, but controversy remains concerning the clinical associations of these alterations. Although DNA amplification is not common, the consensus is that this is associated with higher tumor stage and/or grade (e.g., [155–159]). While it is clear that a larger number of tumors show protein overexpression rather than gene amplification, there are conflicting results about the relationship of expression to clinical parameters. As ERBB2 is known to heterodimerize with other members of the ERBB family, particularly ERBB3, it is likely that studies that examine other family members and their ligands may lead to improved predictive power.

Muscle-invasive bladder tumors commonly display extensive genomic instability. Karyotypes and comparative genomic hybridization (CGH) profiles are complex [145,160–162] and multifocal tumors from the same patient may be highly genetically divergent, indicating rapid molecular evolution [163]. CGH and array CGH studies of large tumor panels have provided much information [144,145,162,164,165]. Although many alterations have been described and candidate genes have been listed, functional validation for the majority of these is pending. In addition, it is not yet clear whether some of these events may be so-called "passenger" events that are a manifestation of the inherent instability rather than "driver" events that confer selective advantage.

5.2.7
Genetic Alterations with Unrecognized Associations to Tumor Stage and Grade

Although there is some understanding of what combinations occur together most frequently and in which type of tumor, some of the genetic and epigenetic alterations, that have been identified in bladder tumors are difficult to correlate with specific subtypes. These include alterations to chromosome 9 and the *RAS* genes, which have been found at similar frequency in tumors of all grades and stages.

5.2.7.1 Alterations of Chromosome 9

More than half of all bladder tumors show chromosome 9 alterations [91,166, 167]. In addition, deletion of chromosome 9 was one of the first alterations identified cytogenetically in bladder tumors. Loss of an entire copy of the

chromosome was described as the only karyotypic alteration in near-diploid tumors [168,169]. Chromosome 9 is the most-studied genomic region in bladder tumors, as much effort has been made to identify the tumor suppressor gene or genes that drive this common loss. This has proven extremely difficult, as many tumors lose an entire homolog and thus information on the location of relevant genes is lost. However, subchromosomal deletions found in some cases have allowed mapping of common regions of loss.

On the short arm (9p), a single region encompassing the known tumor suppressor genes and cell cycle regulators *CDKN2A* (encoding p16 and $p14^{ARF}$) and *CDKN2B* (encoding p15) has been defined [117,170–173]. The majority of deletions focus on *CDKN2A*. p16 is a negative regulator of the Rb pathway and $p14^{ARF}$ is a negative regulator of the p53 pathway. The locus is commonly homozygously deleted in bladder tumors. An association of homozygous deletion with high grade and stage has been found [174], and loss of p16 expression has been associated with increased risk of recurrence in Ta and T1 tumors [175]. However, there is not complete concordance between studies, as a study of homozygous deletion of exon 1B of *CDKN2A* found no association with tumor grade and stage [176]. As knockout mouse studies and *in vitro* experiments on mouse cells indicate that p16 and/or $p14^{ARF}$ may be haplo-insufficient [174], this may be phenotypically important in all bladder tumors with LOH.

On 9q, the effect of LOH is less clear, as several regions have been mapped. In three regions, candidate genes are implicated: *PTCH* (Gorlin syndrome gene) at 9q22 [177,178], *DBC1* at 9q33 [179–181], and *TSC1* at 9q34 [89,90,95]. Mutations of *PTCH* are infrequent [177], but reduction in mRNA expression has been found in tumors with 9q22 LOH [178]. Small sequence alterations of *DBC1* have not been identified in the retained allele in tumors with 9q LOH [182,183], but homozygous deletion of the region has been identified in a few cases [181,183,184] and the gene is commonly silenced by methylation [179,185,186]. Functional studies indicate a negative effect on cell cycle progression and overexpression of the gene in non-expressing bladder tumor cells induces cell death [187].

Although further validation of a tumor suppressor role is desirable for both *PTCH* and *DBC1*, there is compelling evidence for a tumor suppressor role for the tuberous sclerosis gene *TSC1* (9q34) in bladder cancer, as indicated above. Mutations of *TSC1* show no apparent association with tumor grade or stage. However, in a single study, LOH at D9S149 (close to *TSC1*) has been reported to be associated with tumor invasion [188]. Although the pivotal importance of chromosome 9 alterations in bladder tumors is clear, the potential impact of inactivation of multiple genes via loss of an entire chromosome and the possible effects of inactivation of different combinations of chromosome 9 genes in tumors of different grade and stage will be difficult to unravel.

5.2.7.2 RAS Gene Mutations

As indicated above, and in contrast to the relationship between *TP53* and *FGFR3* mutations, the absolute mutual exclusivity of *FGFR3* and *RAS* gene mutations is

thought to reflect activation of the same pathway by either event. *RAS* mutations are found in approximately 13% of bladder tumors (all stages and grades) and in total 85% of low-grade Ta tumors were found to have a *RAS* or a *FGFR3* mutation, suggesting that MAPK pathway activation may be an obligate event in most of these tumors. Interestingly, unlike *FGFR3* mutation, no obvious relationship of mutation of a *RAS* gene with tumor grade of stage has been found. This implies that although both events may fulfill at least one function that precludes selection for both, there is a difference, possibly in the strength and/or duration of signals generated, that allows *RAS* mutation to contribute equally well to the development of both major tumor groups.

5.3
Prostate Cancer

Prostate cancer is the second most often occurring malignant tumor in men worldwide, with about 910 000 new cases diagnosed annually. However, it is the sixth commonest cause of male cancer-related death with more than 260 000 deaths per year worldwide [189]. The lifetime risk of prostate cancer diagnosis is estimated at about 18%, whereas that for death from prostate cancer is about 3% [14]. The incidence rates of prostate cancer vary due not only to genetic factors, but also to ethnicity, geographic residence area, age, socio-economic status, and other environmental factors, such as lifestyle, place of work, and dietary habits [190]. Looking to the regional distribution, Australia/New Zealand, North America, and Western and Northern Europe have the highest rates of prostate cancer, whereas countries in South-Central Asia have the lowest incidence [189]. Screening for prostate-specific antigen (PSA) may be the reason for a higher detection rate in the former regions. However, recent studies on migrant populations from South Korea to North America compared with native South Koreans show an increased risk for the migrants what points to lifestyle as a risk factor for prostate cancer [191]. An important part of the risk factors are, however, genetic and epigenetic factors either inherited or acquired (sporadic). Genetic and/or epigenetic factors have been extensively reviewed previously [192,193].

5.3.1
Hereditary Factors for Prostate Cancer

There are some hereditary factors for prostate cancer that have been identified and, importantly, the distribution of gene alterations can differ considerably in different ethnic backgrounds (reviewed in [190,194,195]). Several germline mutations and SNPs are present in prostate cancer susceptibility genes [193,196–198]. Germline mutations in the breast cancer predisposition gene 2 (*BRCA2*) are the genetic events known to date that confer the highest risk of prostate cancer (8.6-fold in men aged 65 years or less) [196]. Carriers of *BRCA2* germline mutations showed a significantly shorter disease-specific and metastasis-free survival

compared with non-carriers ($P < 0.001$) [199]. However, germline mutations in the *BRCA2* gene are rare and occur only in 1–2% of sporadic prostate cancer cases [196].

Most recently, a novel variant of the homeobox B gene 13 (*HOXB13* G84E; a substitution of glutamic acid for glycine) could be shown to be associated with a significantly increased risk of hereditary prostate cancer (HPC) [200]. Hoxb13 is required for normal differentiation and secretory function of the ventral prostate in mice [201], and the human Hoxb13 protein can physically interact with androgen receptor [202,203], but the function of the G84E mutation has not yet been identified. However, this mutation was mostly present in only four families with HPC. Beyond these cases, only 1.4% (72/5083) of patients with prostate cancer and 0.1% (1/1401) of subjects without prostate cancer carried this mutation. Altogether, this mutation was significantly more common in men with early-onset, familial prostate cancer (3.1%) than in those with late-onset, non-familial prostate cancer (0.6%) ($P = 2.0 \times 10^{-6}$).

Genetic polymorphisms in c-*MYC* (see below), a protooncogene located on chromosome 8, have consistently been shown to be associated with prostate cancer risk across multiple ethnic groups (reviewed in [204]).

In addition, there are a few HPC genes. Most relevant candidate genes for HPC are RNase L (*HPC1*), located at chromosome 1q22; *MSR1* (macrophage scavenger protein) with a linkage region on chromosome 8p; and *ELAC2* (*HPC2*; elaC homolog 2, *Escherichia coli*), on chromosome 17p11 [194]. RNase L regulates cell proliferation and apoptosis through the interferon-regulated 2-5A pathway [205], Msr1 mediates the binding, internalization, and processing of a wide range of negatively charged macromolecules [206], and Elac2 has similarity to a family of DNA interstrand cross-link repair proteins [207]. What could be the functions of these genes? RNase L is involved in cellular proliferation and apoptosis in response to viral infections and other external stimuli. Msr1, as a macrophage pattern recognition receptor, affects, together with Toll-like receptor 4, the balance between life and death of macrophages [208], Elac2 can bind to γ-tubulin, and an overexpression of Elac2 in tumor cells causes a delay in G_2–M cell cycle progression that is characterized by accumulation of cyclin B [209]. Altogether, these three genes have been identified as hereditary tumor suppressor genes in prostate cancer [210] as their normal function is abrogated by mutations and/or SNPs.

In addition, several other HPC genes, such as *HPC2* (chromosome 17p12), *HPC8*/*PCAP* (prostate cancer hereditary 8, chromosome 1q42–43), *HPC20* (chromosome 20q13), *HPCX* (chromosome Xq27–28), and *CAPB*/*PCBC* (prostate cancer/brain cancer susceptibility gene, chromosome 1p36), are suggested as prostate cancer susceptibility genes [190,195]. Further gene mutations and polymorphisms involved in hormone metabolism (*AR*, *SRD5A1*), detoxification (*GSTM1*), regulation of growth factor activity (*IGFBP3*), and folate metabolism (*MTHFR*) have been summarized previously [192].

It has been suggested that there is an inherited susceptibility to develop a *TMPRSS2–ERG* fusion. Such a fusion is characteristic for about 50% of non-

HPCs (see below). The fusion consists mostly of the promoter of the androgen-regulated gene *TMPRSS2* (transmembrane serine protease 2) and the coding sequence of the *ETS* family transcription factor *ERG* resulting in an overexpression of Erg protein in prostate tissue. Of 75 patients from 36 families with an established history of prostate cancer, the *TMPRSS2–ERG* fusion could be detected in 44 (59%) patients. Whereas almost three quarters (73%) of fusion-positive patients accumulated within 16 specific families, 27% were single fusion-positive cases within one family. An association between *TMPRSS2–ERG* fusion and higher Gleason scores has been reported ($P = 0.027$). A genome-wide linkage scan at 500 markers revealed several loci located on chromosomes 9, 18, and X that were suggestive of linkage to the *TMPRSS2–ERG* fusion-positive prostate cancer phenotype. Any of these loci may include a DNA repair gene.

A recent study from the international PRACTICAL Consortium genotyped 211 155 SNPs in blood DNA from 25 074 prostate cancer cases and 24 272 controls to identify prostate cancer susceptibility alleles [197]. This study suggests that more than 70 of these loci identified can explain around 30% of the familial risk for the disease. On the basis of combined risks conferred by the new and previously known risk loci, the top 1% of the risk distribution has a 4.7-fold higher risk than the average of the population being profiled [197].

Recently, when studying a 1536 SNP panel in African-American and European-American men, 11 SNP could be associated with prostate cancer aggressiveness ($P < 0.05$). They were located in three genomic regions: one at 3p12 (European-American), seven at 8q24 (five African-American, two European-American), and three at 19q13 at the kallilkrein-related peptidase 3 (*KLK3*) locus (two African-American, one European-American) [211]. Furthermore, a meta-analysis of GWAS was performed for 5953 cases of aggressive prostate cancer and 11 463 controls (men without prostate cancer) to identify prostate cancer susceptibility loci [198]. This study included approximately 2.6 million SNPs associated with aggressive prostate cancer. The prostate cancer susceptibility locus rs11672691 (chromosome 19) showed clearly an association with aggressive prostate cancer (odds ratio = 1.12; $P = 1.4 \times 10^{-8}$). The rs11672691 lies between the gene loci *ATP5SL* and *CEACAM21*, and interestingly within a hypothetical locus, LOC100505495, of a ncRNA. *ATP5SL* codes for an ATP synthase-like protein but its function is not known. The *CEA* (carcinoembryonic antigen) gene family belongs to the Ig superfamily of genes. Two members of this family have a role in cell adhesion, invasion, and metastasis, but a role for *CEACAM21* in cancer is unknown. However, the role of the ncRNA (LOC100505495) has to be determined as well.

Among the ncRNAs, the miRNAs have recently gained much interest in cancer research. Variant alleles of SNP within miRNA genes and miRNA target genes can affect susceptibility for cancer (reviewed in [212]). SNP variants in miRNAs can protect against the pathogenesis or increase the risk for cancer, including prostate cancer [213]. In particular, SNPs in miR-146a, miR-196a2, and miR-499a have been reported to play a role for prostate cancer susceptibility (reviewed in [213]).

5.3.2 Sporadic Factors for Prostate Cancer

Most of the biomarkers described for prostate cancer have been studied on the protein rather than on the nucleic acid (DNA/RNA) level. Protein biomarkers have been reviewed extensively and are not the main subject of this chapter [14,204,214]. However, since the only biomarker that is routinely used by urologists is the total PSA (human kallikrein 3, human kallikrein-related peptidase 3, hK3) and there is actually a great debate about the advantages and disadvantages of PSA screening, we will add a short paragraph concerning PSA and list briefly other protein biomarkers.

5.3.2.1 PSA and Other Protein Markers

PSA screening is an organ-specific rather than a cancer-specific test. A total PSA >4 ng/ml is considered as abnormal. However, abnormal PSA levels can be caused, in addition to prostate cancer, by prostatitis or cystitis, benign prostatic hyperplasia, ejaculation, prostatic intraepithelial neoplasia, perineal trauma, or the recent use of instruments for testing or surgery [215,216]. However, a normal PSA value does not rule out the existence of a prostate cancer [216]. The specificity of the PSA test is very weak and sensitivity is not as good as desired. The diagnostic sensitivity (i.e., the percentage of prostate cancer patients identified by the test among all prostate cancer patients) is about 85% and the specificity (i.e., the percentage of healthy patients identified by the test among all healthy patients) is about 23% [217].

A problem with the test is that overdiagnosis and overtreatment with a risk of biopsy-related hematuria and urosepsis have to be considered [218]. However, the European Randomized Study of Screening for Prostate Cancer, the largest randomized study with 162 388 participants, revealed at a median follow-up of 9 years, and a prostate cancer mortality reduction of 20% ($P=0.04$) and, after 11 years, 21% [218,219].

To reduce the harm, it is important to consider the age of patients when the PSA test is suggested. In a 10-year competing risk analysis of 19 639 men aged 66 years and older, relatively few men diagnosed with localized prostate cancer will die as a result of prostate cancer within 10 years of diagnosis. Recently, a treatment recommendation based on PSA screening when considering prognostic factors has been suggested [220]. Men aged 55–59 years with moderate-risk prostate cancer are predicted to derive the greatest benefit from immediate curative treatment. Immediate treatment is least favorable for men aged 70–74 years with either low-risk or high-risk prostate cancer [220]. Thus, PSA screening may gain importance for men in the age range 40–70 years for treatment decisions, but might have less impact on survival and treatment options for men older than 70 years.

In addition to screening for total PSA, different PSA molecular forms and derivatives have been discussed in detail for their impact to add some value to the screening, diagnostic, and/or predictive value of total PSA (reviewed in [14]).

Furthermore, several other proteins, such as growth factors (e.g., (transforming growth factor-β and its target genes), chemokines/receptors (e.g., interleukin-6 and its receptor), and peptidases/receptors (e.g., plasminogen activator, urinary, synonym/urokinase-type plasminogen activator and its receptor; KLK2/hK2 (kallikrein-related peptidase 2, synonym: kallikrein 2, glandular kallikrein, prostatic kallikrein)) have been described as biomarkers in prostate cancer [221–234] (reviewed in [235,236]).

More promising than characterization of one protein marker is the analysis of a marker panel. This can include kallikrein-related markers as well as different biomarkers. A statistical prediction model for prostate cancer based on four kallikrein markers in blood (total, free, and intact PSA, and KLK2/hK2) had a much higher predictive accuracy than total PSA alone (area under the curve (AUC) of 0.711 versus 0.585) [237]. Recently, a simultaneous and combined detection of multiple tumor biomarkers for prostate cancer in human serum by suspension array technology has been described. The multiparametric detection of four prostate cancer biomarkers (i.e., levels of PSA, prostate stem cell antigen, prostate-specific membrane antigen, and prostatic acid phosphatase) may allow large-scale screening of high-risk groups of prostate cancer patients in the future [238].

5.3.2.2 Nucleic Acid Biomarkers

At the nucleic acid level, biomarkers can be separated concerning RNA or DNA. In addition to alterations in the nucleic acid sequence in RNA or DNA (mutations, SNPs, RNA splice variants), alterations in the RNA levels and modifications of RNA/DNA (methylation, acetylation) can lead to biomarkers.

RNA Biomarkers

An RNA biomarker in prostate cancer is the *TMPRSS2–ERG* fusion transcript (see below) as it can be found in about 54% of primary prostate cancer [239]. *TMPRSS2–ERG* fusion transcripts can be detected non-invasively in the urine of prostate cancer patients after digital rectal examination, and their presence has been associated with tumor size and high Gleason score at prostatectomy [240]. The diagnostic performance of *TMPRSS2–ERG* fusion transcripts was 84–93% specificity and 31–37% sensitivity [241,242]. It is suggested that detection of these transcripts could help to detect prostate cancer in 15–20% of patients with PSA < 4 ng/ml [243]. In the patient cohort described by Hessels *et al.* [241], prostate cancer antigen 3 (*PCA3*) transcripts were in addition detected with a sensitivity of 62%. When both markers (*TMPRSS2–ERG* and *PCA3*) were combined, the sensitivity increased to 73% [241]. However, the association of *TMPRSS2–ERG* fusion transcripts with biochemical recurrence, progression-free survival or poor outcome is still controversial (reviewed in [214,243]). Altogether, detection of *TMPRSS2–ERG* fusion transcripts in the urine can have potential as a diagnostic biomarker for prostate cancer, but this has to be evaluated in clinical tests.

There are several studies suggesting that the expression of sets of genes can be used as diagnostic or prognostic biomarkers, identifying mainly the genes

α-methylacyl CoA racemase (*AMACR*), enhancer of zeste homolog 2 (*EZH2*), α-2-glycoprotein 1 (*AZGP1*), activated leukocyte cell adhesion molecule (*ALCAM*/*MEMD*/*CD166*), and cluster of differentiation 24 (*CD24*) (reviewed in [244]). These studies were continued to distinguish between different Gleason scores, and to identify markers for stem cell-like states or increased cell cycle progression that is characteristic of more aggressive prostate cancer (reviewed in [193]), but there is still no consensus which marker combination is most useful for the diagnosis or prognosis of prostate cancer. Despite this, several companies are trying to establish gene expression assays for diagnostic tests. One company is Myriad Genetics (Salt Lake City, UT, USA) that offers a 46-gene set involved in cell cycle progression for diagnostic purposes. However, further clinical testing is necessary to evaluate this and other tests for their applicability.

Recently, a molecular signature predictive of indolent prostate cancer has been reported [245]. Irshad *et al.* showed that prostate cancer with a low Gleason score can be separated into indolent and aggressive subgroups by the expression of the three genes *FGFR1*, peripheral myelin protein 22 (*PMP22/GAS3*) and cyclin-dependent kinase 1a (*CDKN1a/P21*) on the RNA and on the protein level. A high expression of all three genes/proteins was predictive for a good clinical outcome of low Gleason prostate cancer, whereas a low expression was associated with a poor outcome of the low Gleason patient cohort [245].

In addition to mRNAs, ncRNAs have come into the scientific focus of cancer research during the last 15 years. ncRNAs comprise mainly ribosomal RNA (rRNAs), transfer RNAs (tRNAs), the group of small ncRNAs with miRNAs (miRNAs), small interfering RNAs (siRNAs), PIWI-interacting RNAs (piRNAs), small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), long ncRNAs (lncRNAs) and antisense RNAs (AS-RNAs) [246,247].

Several ncRNAs, including prostate-specific gene *PCGEM1*, prostate *PCA3* (prostate specific gene *DD3*), prostate cancer ncRNA 1 (*PRNCR1*), and prostate cancer-associated ncRNA transcript 1 (*PCAT-1*) have been previously reported in prostate cancer [248–251]. However, out of these, only measurement of the transcript of the *PCA3* gene in urine has shown some relevance in clinical application until now. PCA3 RNA expression is restricted to the prostate (i.e., it is not expressed in any other normal human tissue or tumor) [252]. In an assay (PROGENSA PCA3 assay; Gen-Probe) that was approved in February 2012 by the US Food and Drug Administration [253], the PSA RNA and PCA3 RNA molecules are measured and the PCA3 score is calculated (a ratio of PCA3 RNA to PSA RNA). This test can be used in men (aged 50 years or older) who have had one or more previous negative prostate biopsies and for whom a repeat biopsy would be recommended by a urologist based on current standard of care. A PCA3 score <25 is associated with a decreased likelihood of a positive biopsy. As a result, some patients who would otherwise be recommended for a biopsy may be saved from unnecessary biopsies and related complications. When summarizing several studies, Bradley *et al.* concluded that PCA3 has a higher diagnostic accuracy for prostate cancer than (total) PSA, but the strength of the evidence is low [254].

Takayama *et al.* recently reported the identification of CTBP1-AS – a novel androgen-regulated long ncRNA that promotes prostate cancer growth through sense–antisense repression of the transcriptional coregulator CTBP1, as well as through the global epigenetic regulation of tumor suppressor genes [255,256].

Special attention in cancer research has been given to a group of ncRNAs – the miRNAs (reviewed in [257]). miRNAs are important players in tumorigenesis, tumor progression, and therapy response, and function as onco-miRNAs or tumor suppressor miRNAs [258], as has also been shown for prostate cancer (reviewed in [48,259–263]). miRNAs are attractive as biomarkers as they (i) can be readily detected in small-volume samples using specific and sensitive qRT-PCR, (ii) have been isolated from most body fluids, including serum, plasma, urine, saliva, breast milk, tears, and semen, and (iii) are known to circulate in a highly stable, cell-free form [264,265].

Circulating miRNAs detected in plasma or serum have been suggested as new biomarkers for prostate cancer [266–269] (reviewed in [265,270,271]). They may function as diagnostic, predictive, prognostic biomarkers and as treatment targets (reviewed in [265]). Mitchell *et al.* first described that the miR-141 level was 46-fold increased in the serum of metastatic prostate cancer patients compared with healthy individuals, and therefore could identify advanced prostate cancer patients with 60% sensitivity and 100% specificity [266]. In the plasma of 28 healthy controls compared with 78 prostate cancer cases, Bryant *et al.* identified 11 miRNAs significantly upregulated and one miRNA downregulated in prostate cancer, with miR-107 the most upregulated [268]. Selth *et al.*, using a prostate cancer mouse model (TRAMP), identified significantly expressed miRNAs and could validate them in human serum samples when comparing healthy controls with metastatic castration-resistant prostate cancer (upregulated: miR-141, miR-298, miR-346, and miR-375) [269]. Bryant *et al.* detected 16 miRNAs (15 upregulated: miR-20a*, miR-17*, miR-23a*, miR-130b, miR-198, miR-200b, miR-375, miR-379, miR-513a-5p, miR-577, miR-582-3p, miR-609, miR-619, miR-624*, and miR-1236; one downregulated: miR-572) that differed significantly in their expression levels between localized prostate cancer and metastatic prostate cancer [268]. Nguyen *et al.* could also show that serum samples from patients with a localized low-risk prostate cancer and from metastatic castration-resistant prostate cancer patients exhibited distinct circulating miRNA signatures (upregulated: miR-141, miR-375, and miR-378; downregulated: miR-409-3p) [271]. Although the miRNAs detected as diagnostic biomarkers in those studies differed somewhat, miR-141 and miR-375 were mostly overexpressed in blood-derived samples of metastatic prostate cancer patients and appear as promising diagnostic candidates.

miRNAs were also identified as prognostic biomarkers in prostate cancer. Significantly different expression levels of miRNAs in serum or plasma were detected between tumor stages, Gleason scores, and Cancer of the Prostate Risk Assessment (CAPRA) scores (reviewed in [265,270]). Brase *et al.* compared sera from primary prostate cancer with those from metastatic prostate cancer. After finding 69 miRNAs increased in metastatic prostate cancer patients, a subset of

the miRNAs was analyzed in the sera of localized prostate cancer patients and three miRNAs (miR-141, miR-200b, and miR-375) showed an elevated expression at increasing tumor stage and Gleason score [267]. In the above-mentioned study by Bryant *et al.*, 16 miRNAs were significantly higher in sera of metastatic compared with localized prostate cancer patients [268]. When comparing miRNA levels in serum samples from 51 prostate cancer patients with different tumor stages (18 localized, eight locally advanced, and 25 metastatic), miR-141, miR-21, and miR-221 were more highly expressed in advanced compared with localized prostate cancer serum samples [272]. Two studies investigated miRNA levels in serum samples of prostate cancer patients with different CAPRA scores. Moltzahn *et al.* found miR-24 decreased and miR-93, miR-106a, and miR-451 were elevated with increasing CAPRA score [273], whereas Shen *et al.* detected miR-20a and miR-21 elevated at a high-risk CAPRA score [274]. In summary of the results in blood-based components, miR-141, miR-200b, and miR-375 are suggested as significant disease correlates that could be used in a test at the time of diagnosis to identify those patients with previously undetectable micrometastases [265].

miRNAs may have also the potential as predictive biomarkers, that is, to predict the response of prostate cancer patients to different treatment regimens and/or to monitor the effectiveness of the treatment. Zhang *et al.* measured expression levels of miR-21 in serum of patients with either localized, androgen-dependent or castrate-resistant prostate cancer. They detected that miR-21 levels were significantly higher in androgen- and castrate-resistant prostate cancer patients with PSA > 4 ng/ml, but not in those with PSA < 4 ng/ml. Four out of 10 castrate-resistant prostate cancer patients, who were resistant to docetaxel-based chemotherapy, showed significantly higher miR-21 levels than the six patients with partial remission [275]. Gonzales *et al.* compared miR-141 in plasma with PSA levels and occurrence of circulating tumor cells in 21 metastatic prostate cancer patients in a longitudinal study. Levels of miR-141 and PSA were significantly correlated. Changes (increasing or decreasing) in PSA, circulating tumor cells, and miR-141 had sensitivity in predicting clinical outcome (progression versus non-progressing) of 78.9% [276].

miRNAs detected in whole-blood samples can also be informative as biomarkers. Detection of miRNAs in 454 whole-blood samples from human individuals with different cancers or non-cancerous diseases by array analysis and validation by qRT-PCR gives a "miRNome" – a deregulated profile for all tested diseases including prostate cancer [277]. Successful pathway analysis confirmed disease association of the respective miRNAs [277].

Detection of miRNAs in prostate cancer tissues can be used as biomarker as well although it is connected with an invasive procedure such as a biopsy or prostatectomy. Others and ourselves could show that miRNAs detected in prostate cancer can be used as a diagnostic and/or prognostic tool. When comparing malignant and matched normal tissue of 40 prostate cancer patients, Tong *et al.* identified five miRNAs (miR-23b, miR-100, miR-145, miR-221, and miR-222) to be significantly downregulated [278]. Schaefer *et al.* studied miRNA expression

in matched tumor and adjacent normal tissues from 76 patients, and identified 15 differentially expressed miRNAs (downregulated: miR-16, miR-31, miR-125b, miR-145, miR-149, miR-181b, miR-184, miR-205, miR-221, and miR-222; upregulated: miR-96, miR-182, miR-182*, miR-183, and miR-375). miR-205 and miR-183 were the best discriminating miRNAs with a correct overall classification of 82% [279]. In our study, 25 miRNAs were significantly deregulated and a combination of three miRNAs (miR-375, miR-143, and miR-145) correctly classified prostate cancer tissue samples with an accuracy of 77.6% [280]. In the above-mentioned study of Schaefer *et al.*, they showed that the expression of a few miRNAs correlated with Gleason score (miR-31, miR-96, and miR-205) or pathological tumor stage (miR-125b, miR-205, and miR-222). Another study by Spahn *et al.* in primary carcinoma tissue and in corresponding metastatic tissue revealed that miR-221 was progressively downregulated in aggressive forms of prostate cancer, which was associated with Gleason score and clinical recurrence during follow-up [281].

DNA Alterations

The role of the tumor suppressor phosphatase and tensin homolog (PTEN; tumor suppressor phosphatase and tensin homolog deleted on chromosome 10, PTEN1) in human cancer has been reviewed excellently by the group of Pandolfi [282,283]. *PTEN* encodes a lipid phosphatase that is mainly involved in the control of the cell cycle and DNA damage repair. The *PTEN* gene is altered or lost in about 70–80% of all primary prostate cancers [282]. However, the rate of *PTEN* deletions is less common in localized prostate cancer (37%) compared with castration-resistant prostate cancer (63%) [284]. *PTEN* loss in circulating tumor cells could be found in 26% of patients with circulating tumor cells [285]. Patients with a *PTEN* loss in their tumors had a significantly higher risk of dying from prostate cancer (hazard ratio: 4.87; 95% confidence interval: 2.28–10.41; $P = 0.001$) [286]. Loss of the *PTEN* gene or PTEN expression was significantly associated with risk of biochemical and/or clinical recurrence in prostate cancer patients after prostatectomy [287–289]. *PTEN* losses exhibit heterogeneity in multifocal prostatic adenocarcinoma and are associated with higher Gleason grade [290]. *PTEN* loss plays a major role in the activation of the oncogenic PI3K/AKT pathway [291,292] (reviewed in [293]). Altogether, *PTEN* may potentially be a useful biomarker in active surveillance protocols, given it appears to be an early biomarker of disease progression [204].

The v-*myc* avian myelocytomatosis viral oncogene homolog (oncogene *MYC*) is a protooncogene identified three decades ago that encodes a transcription factor that plays a role in many human cancers (reviewed in [294]). Myc is a sequence-specific transcription factor, which binds DNA only when dimerized with its obligate partner Max. In this way it can regulate about 15% of all human genes by binding to enhancer box sequences (E-boxes) and by histone post-translational modifications [295]. However, Myc is not an on/off specifier of a particular transcriptional program(s), but is a universal amplifier of gene expression, increasing output at all active promoters [294]. In prostate carcinogenesis,

alterations in c-Myc expression might be an early event. c-Myc is already overexpressed in prostatic intraepithelial neoplasia, but this is not correlated with *MYC* gene amplifications [296]. However, genomic alterations affecting the 8q24 region that includes the *MYC* locus are frequent in prostate cancer. Amplification at the *MYC* locus was observed in 29% (64/221) of prostate cancer cases, and was closely associated with both disease progression (from 15% in pT2 tumors to 53% in metastasis) and Gleason score (from less than 3% in Gleason 6 tumors to 66% in Gleason 8 or higher tumors [297]. *MYC* amplification status but not c-Myc protein expression was significantly predictive of biochemical recurrence after prostatectomy and could be a prognostic tool for prostate cancer patients [297]. Comparable results were reported by Zafarana *et al.* when they found that copy number alterations of *MYC* (gain) alone and in combination with *PTEN* (deletion) are prognostic factors for relapse after prostate cancer radiotherapy [298]. Altogether, *MYC* represents a promising candidate for early detection of prostatic intraepithelial neoplasia and prostate cancer progression.

The NK3 homeobox 1 gene (*NKX3.1*, homeobox, family 3 member A) is expressed in a largely prostate-specific and androgen-regulated manner [299]. It marks a luminal stem cell population that functions during prostate regeneration [300]. Targeted deletion of the *PTEN* tumor suppressor gene in castration-resistant *NKX3.1*-expressing cells in a mouse model results in rapid formation of carcinoma following androgen-mediated regeneration. However, loss (haploinsufficiency) of the *NKX3.1* gene is prognostic for prostate cancer relapse following surgery or image-guided radiotherapy [301]. Although a role of *NKX3.1* in prostate cancer genesis has been clearly shown, its function as an oncogene or tumor suppressor gene still has to be resolved.

Another way of regulating gene activation is DNA methylation. Alterations of DNA methylation are unusually prevalent in prostate cancer [192]. One of the most affected genes is the *GSTP1* gene encoding an enzyme protective against electrophilic compounds from exogenous and endogenous sources whose expression is lost by DNA methylation early in prostate cancer development [192]. Among miRNA genes, the miR-145 gene in particular has been reported to be inactivated by DNA promoter methylation in prostate cancer cell lines and patient samples [302,303].

Of special interest as biomarkers are tumor-specific genetic alterations. The fusion of the androgen-regulated transmembrane protease, serine 2 gene (*TMPRSS2*) and an *ETS* transcription factor family gene (mostly with *ERG* or *ETV1*) could be specifically identified in prostate cancer patients [304]. The most prominent fusion partner from the *ETS* gene family is the v-*ETS* avian erythroblastosis virus E26 oncogene homolog (*ERG*, *ETS*-related gene) [305]; a fusion between *TMPRSS2* and *ERG* is observed in about 50% of patients diagnosed with prostate cancer [306]. Erg is an oncogene with mitogenic and transforming activity [305,307]. Mani *et al.* could show that *TMPRSS2–ERG* mediates a feed-forward regulation of wild-type *ERG* in prostate cancer [308]. *TMPRSS2–ERG* fusion transcripts have potential as a diagnostic biomarker in urine of prostate cancer patients (see above). Another study performed in mice indicates that

a combination of biomarkers may have more impact for the detection of prostate cancer. Overexpression of Erg was not sufficient to initiate prostate cancer, but it can result in an invasive phenotype when, in addition, one allele of the *PTEN* gene was missing [309]. The prognostic impact of the *TMPRSS2–ERG* fusion is controversial [214]. An increased copy number of wild-type *ERG* has recently been shown to be prognostic for prostate cancer recurrence, but not *TMPRSS2–ERG* fusion [310]. In a meta-analysis, Pettersson *et al.* suggest that *TMPRSS2–ERG*, or Erg overexpression, is associated with tumor stage but does not strongly predict recurrence or mortality among men treated with radical prostatectomy [311].

5.3.3
Prostate Cancer: Summary

Briefly summarizing, there are several pathways that are important for prostate cancer tumorigenesis and progression, especially the *PTEN* (and its targets PI3K/AKT), c-Myc, and Erg pathways, and therefore a combination of biomarkers in all three pathways may have the greatest impact.

5.4
Renal Cell Carcinoma

With a total of more than 270 000 new cases worldwide, kidney cancer is a significant contributor to the global tumor burden [189]. In Europe, the age-standardized incidence rates per 100 000 individuals are 15.8 for males and 7.1 for females, making it the 10th most common tumor in this region [312]. RCC is not a uniform disease, but rather a multitude of different tumor entities [80], each with its own etiology. The most common RCC entity is the conventional, clear cell RCC with an incidence of 75%. This is followed by papillary RCCs (10–15%), chromophobe RCCs (5%), and Ductus–Bellini-like tumors (1%) [313]. There are only a few well-studied risk factors associated with the development of RCC. For example, the incidence of RCC is highest in African-Americans in the United States [314]. Other known risk factors include lifestyle factors such as cigarette smoking [315], obesity [316], and hypertension [317].

5.4.1
Hereditary Factors for RCC

RCC is unique in the way that there exist several well-known hereditary syndromes associated with the development of distinct entities of RCC (reviewed in [318–321]). The most common hereditary syndrome is the von Hippel–Lindau (VHL) disease (OMIM: 193300). VHL affects 1: 36 000 individuals [322] and is inherited in an autosomal dominant fashion. The VHL gene is located on chromosome 3p25–26 and is highly conserved throughout all multicellular

organisms. By alternative translation initiation, two protein products are encoded: the 30-kDa full-length protein consisting of 213 amino acids and the smaller 19-kDa protein with 160 amino acids [323]. Both protein variants are able to regulate the hypoxia-inducible factor (HIF)-α and therefore exhibit tumor-suppressive function. VHL disease is caused by inactivating germline mutations of one VHL allele [324–326], with a subsequent LOH of the wild-type allele by second-hit mutations [327,328] or epigenetic silencing [328,329]. Tumors associated with VHL syndrome can occur in several organ systems. These are mainly solid, vascular tumors of the retina, the central nervous system, kidney (clear cell RCC), adrenal gland, or the endolymphatic sac [330]. These tumors occur with a high penetrance of 80–90% and develop in the second to fourth decade of life. The patients are at risk of developing hundreds of tumors per kidney [331]. There exists a very interesting genotype/phenotype correlation in VHL syndrome patients. Mutations such as missense mutations or premature termination mutations result in a complete absence of VHL function. These type 1 VHL mutations predispose to the complete spectrum of VHL phenotypes except phaeochromocytomas. Type 2 VHL mutations such as missense amino acid changes that drastically reduce VHL function predispose to the complete spectrum of VHL-associated phenotypes including (type 2B) or excluding (type2A) clear cell RCC [332,333]. Interestingly, the types of VHL mutations that predispose to the development of clear cell RCC (type 1 and type 2B) all result in a complete absence of HIF-1α ubiquitinylation [334]. Germline mutations affect the whole VHL gene. There are only 26 codons in the VHL gene for which no mutations have been reported [335]. Therefore, a direct sequencing of the entire coding region [336] of the VHL gene is the method of choice for detecting VHL mutations.

Hereditary papillary renal carcinoma (HPRC) syndrome (OMIM: 164860) represents a rare, autosomal dominant syndrome. The kidney is the only organ affected in this disease. Patients have a high risk (90% penetrance by the age of 80) of developing renal tumors [337]. Hereby, HPRC tumors are found bilateral with hundreds or even thousands of tumors per kidney [338]. The genetic basis of the HPRC syndrome has been identified as germline missense mutations of the c-*MET* (hepatocyte growth factor receptor) oncogene [339–341]. The c-Met oncogene is a membrane tyrosine kinase and shares a striking homology to other well-characterized protooncogenes such as c-Kit, c-Erb, c-Src, or Ret [342]. Germline and sporadic mutations target the tyrosine kinase domain of the c-*MET* gene [339], which leads to ligand-independent *Met* phosphyorylation and downstream signaling [343,344]. However, the frequency of somatic mutations in the c-*MET* gene in sporadic papillary RCC cases was reported as only 13% [340]. Interestingly, there is a strict genotype/phenotype correlation concerning germline mutations of the c-*MET* gene. All papillary RCCs harboring an activating c-*MET* mutation share the phenotype of a type 1 papillary RCC [345] as defined in the 2004 WHO classification guidelines [80]. For diagnostic purposes there exists the possibility of sequencing of the tyrosine kinase domain (exons 16–19) or the targeted mutation analysis of known missense mutations in the tyrosine kinase domain. Additionally, amplifications of chromosome 7q, harboring the

mutated c-*MET* allele, have been reported with high frequency [346]. For diagnostic purposes, the amplification of chromosome 7 is detected by CGH.

Hereditary leiomyomatosis and RCC (HLRCC) syndrome (OMIM: 150800) is another rare hereditary disease syndrome associated with the occurrence of renal neoplasms. The genetic bases of this disease are mutations of the fumarate hydratase gene located on chromosome 1q24 [347]. Although the gene is located at chromosome 1, it encodes a mitochondrial enzyme, which converts fumarate to malate [348]. The most common mutations of the fumarate hydratase gene include missense mutations, insertions, deletions, or nonsense mutations, and only rarely whole-gene deletions. Affected individuals display a reduction of fumarate hydratase activity by up to 50% in lymphoblastoid cells or skin fibroblasts [349], with a complete loss of fumarate hydratase enzyme activity in cutaneous leiomyomas. Interestingly, in contrast to VHL or HPRC syndrome, HLRCC-associated tumors are typically solitary and unilateral tumors [350]. Although most renal tumors associated with HLRCC are aggressive type 2 papillary RCCs with high nuclear grade [351], the penetrance for developing papillary RCC over a lifetime is only 20–35% [352]. In sporadic cases of type 2 papillary RCC, only very few somatic mutations of the fumarate hydratase gene have been identified so far [353]. Mutations are found throughout the entire gene without any particular genotype/phenotype correlations [350].

Birt–Hogg–Dubé (BHD) syndrome (OMIM: 135150) is a rare, autosomal dominant syndrome associated with the occurrence of renal neoplasms. Originally described as a genodermatosis syndrome with dermatological symptoms such as fibrofolliculomas [354], it became evident that renal neoplasms were also associated with this disease. Renal tumors affect about 15–30% of BHD patients, and they can be bilateral and multifocal as well as unilateral and solitary [355]. The morphological spectrum of renal tumors varies between chromophobe RCCs, oncocytomas [356], or oncocytic hybrid tumors that exhibit morphological features of both chromophobe RCC and oncocytoma [356,357]. By linkage analysis, the genomic location of the affected gene was identified as chromosome 17p11.2 [358,359] and, subsequently, mutations in the BHD gene (folliculin) were identified [360]. Most of the mutations found in the BHD gene are insertions, deletions, and nonsense mutations that truncate or prematurely terminate the folliculin protein [359,360], and the mutation hotspot is a mononucleotide stretch of cytosines [361]. In sporadic RCC, the BHD gene is only infrequently mutated with a frequency of less than 10% in sporadic RCC cases [362,363]. Diagnostic tests for BHD syndrome include analysis of the BHD gene by sequencing of the entire coding region or targeted mutation analysis.

There are several other, rare hereditary syndromes associated with the occurrence of RCC (reviewed in [320]). The affected genes in these rare hereditary syndromes include regulators of mTOR signaling *TSC1* and *TSC2* [364], inhibitors of the c-*MYC* protooncogene, parafibromin [365], or the succinate dehydrogenase gene B, *SDHB* [366]. The fact that many of the genes associated with the development of RCC play a vital role in the sensing of oxygen, energy, iron, and nutrition has led to the hypothesis that RCC can be regarded as a metabolic disease [367].

5.4.2 Sporadic Factors for RCC

5.4.2.1 The Old

Although the causative genomic loci for hereditary RCC syndromes are well known, the majority of RCC cases represent sporadic tumor events. Accordingly, the spectrum of genetic and genomic alterations shows more variability. Remarkably, in clear cell RCC, the most frequent aberration in sporadic tumors is the inactivation of VHL, further demonstrating the vital impact of this tumor suppressor gene in the pathogenesis of clear cell RCC. VHL is very frequently lost by deletions of chromosome 3 [368]. Up to two-thirds of all sporadic clear cell RCC cases have biallelic loss or inactivation of VHL by genomic deletion, somatic mutation [369], or promoter hypermethylation [370]. In addition to VHL, there are other tumor suppressor genes that are frequently deleted in clear cell RCC cases. These include *TP53* on chromosome 17q, *RB* on chromosome 13q, $p16^{INK4a}$/$p16^{ARF}$ on chromosome 9p, or *PTEN* on chromosome 10q [369]. Interestingly, a specific genomic deletion of chromosome 14q affects the expression of HIF1-α. The deletion of the chromosome 14q genomic region decreased the expression of HIF1-α while the expression of other HIF transcription factors remained unaffected. The loss of genomic material on chromosome 14q was furthermore associated with a reduction in overall patient survival of clear cell RCC patients [371]. Chromosomal amplifications affecting the expression of known protooncogenes is a common event in clear cell RCC. Up to 12% of the cases harbor a chromosomal amplification of chromosome 8q,which is in close proximity of the c-*MYC* protooncogene, and about 30% of all clear cell RCC cases harbor an amplification of the 7q22 genomic interval that does not include the EGFR [372].

The well-known mutations of the c-*MET* protooncogene in hereditary papillary RCCs are not very common in sporadic cases. The frequency of c-*MET* mutations varies between 5% [373] and 13% [340]. Instead of mutations in the c-*MET* oncogene, gene amplification due to trisomy of chromosome 7 is more frequently found in papillary RCCs [340]. In sporadic cases of the highly aggressive papillary RCC type 2, mutations in the fumarate hydratase gene are extremely rare [374]. Characteristic chromosomal aberrations found in a substantial fraction of papillary RCCs hold the potential of being diagnostic biomarkers. The characteristic aberrations are trisomies of chromosomes 7 and 14 as well as loss of the Y chromosome [375]. These changes are specific for papillary RCC, and could be used in a diagnostic setting for the differentiation of papillary RCC and other RCC entities [376]. The chromophobe RCC is unique in the way that this RCC entity is characterized by extensive chromosomal losses. Very frequently, monosomies of chromosomes 1, 2, 6, 10, 13, 17, and 21 have been detected [377], and a relatively high proportion of chromophobe RCCs (up to 30%) additionally harbor mutations in the *TP53* tumor suppressor gene [378]. A distinct molecular entity of RCCs are the so-called translocation RCCs. This RCC entity was initially incorporated in the 2004 WHO classification of tumors

of the urinary tract [80]. The characteristic feature of translocation RCCs is a translocation of the *TFE3* gene, which is located at chromosome Xp11 and can be fused to a variety of translocation partners [379]. Translocation RCCs are very common among pediatric RCCs (up to one-third of pediatric RCCs) and they still represent up to 15% of all RCCs in individuals below 45 years of age [380]. Interestingly, there are reports that translocation RCCs exhibit a very aggressive behavior in male adults [381,382].

5.4.2.2 The New

Recent advances have enabled the application of nucleic acid-based biomarkers for diagnostic and prognostic purposes. SNPs may prove useful for assessing the predisposition for the development of RCC. GWAS have identified several loci that are associated with the risk for developing RCC. The minor alleles of SNPs, rs11894252 and rs7579899, were correlated with an increase in the prevalence of RCC (odds ratios 1.14 and 1.15, respectively) while the minor allele of SNP, rs7105934 (odds ratio 0.69), was associated with a reduction in the risk for developing RCC [383]. Another GWAS identified two SNPs in linkage disequilibrium on chromosome 12p11.23 as genetic variants associated with an increased risk of developing RCC, rs718314 (odds ratio 1.19) and rs1049380 (odds ratio 1.18) [384]. A fine mapping analysis of a previously reported RCC susceptibility locus on chromosome 2p21 identified, in addition to the previously reported SNP rs11894252, four distinct risk alleles within a 120-kb region of chromosome 2p21. An additional meta-analysis of five studies discovered odds ratios of 1.28 for rs12617313, 1.24 for rs4953346, 1.23 for rs4953348, and 1.22 for rs10208823 [385]. In the same way, the RCC susceptibility locus on chromosome 2q22.3 was identified by performing GWAS with a subsequent meta-analysis. Thereby, the SNP rs12105918 was identified that conferred to an elevated disease risk (odds ratio 1.29) [386]. SNPs in genes coding for miRNA processing proteins have likewise been demonstrated to modify the risk for developing RCC [387]. Specifically, SNPs in the *GEMIN4* gene seem to influence the disease risk, whereby the favorable genotypes were associated with a 0.66-fold risk of developing RCC and the unfavorable genotypes with an increased risk of disease of 2.49-fold. In addition to miRNA processing proteins, miRNAs themselves have been proposed as diagnostic markers. The most prominent miRNA associated with RCC is miR-210. When quantified as a potential diagnostic marker in serum samples, miR-210 was able to identify RCC patients from healthy control individuals with a sensitivity of 81% and a specificity of 79% [388]. The additional finding of Zhao *et al.* that serum miR-210 expression significantly decreases within 7 days after surgical removal of the tumor suggests that this miRNA may also be useful as a marker for therapy monitoring or patient follow-up. Using defined sets of different miRNAs, it is possible not only to distinguish between tumor and normal tissue, but also to perform a molecular classification of RCC entities or subtypes [389,390]. It is furthermore possible to assess the quantity as well as the methylation status of cell-free DNA extracted from patient sera. Thereby, the overall quantity of DNA (qPCR quantification of ring finger protein 185, *RNF185* DNA)

and the methylation status of Ras association domain family member 1A (*RASSF1A*) was shown to be diagnostic for the presence of RCC with AUCs of 0.755 and 0.705, respectively [391]. Another study measured the quantity of mitochondrial DNA in cell-free DNA preparations from patient serum. Two independent amplicons of the gene coding for the mitochondrial 16 S RNA were significantly elevated in patients with urologic malignancies, including RCC, bladder cancer, and prostate cancer. Combined, the quantification of mitochondrial DNA allowed the differentiation between serum from cancer patients and control serum with 97% specificity and 84% sensitivity [392]. Likewise, the methylation status of Wnt antagonist family member genes has been suggested as a diagnostic marker. Six members of the family of Wnt antagonists (*sFRP-1*, *sFRP-2*, *sFRP-4*, *sFRP-5*, *WIF*, and *DKK-3*) were analyzed by methylation-specific PCR. In tissue samples, the overall methylation score was a good predictor of tissue dignity with a sensitivity of 79% and a specificity of 76% for the discrimination of tumor and normal tissue. Furthermore, the pattern of aberrant gene methylation could be confirmed for a high fraction (73%) of RCC cases in cell-free serum DNA. Thereby, none of a cohort of healthy control individuals displayed any aberrant methylation patterns in any gene tested [393].

When comparing RCC tumor tissue with normal control tissue, several genes have been described as differentially expressed. Examples for this deregulation include the specific expression of angiopoietin-like 4 (*ANGPTL4*) mRNA specifically in clear cell RCC [394], elevated expression of hTERT mRNA levels in RCC tumor tissue [395], or the tumor-specific expression of CD70 [396]. A very common finding is the deregulation of miRNA expression in RCC tumor tissue. Very prominent miRNAs that are deregulated include the downregulation of miR-200c and miR-141, as well as the upregulation of miR-210 [390,397–399].

In addition to markers that are able to distinguish between malignant and non-malignant tissue, there exist several reports about the prognostic value of nucleic acid-based markers. For example, the elevated expression of Fork head box M1 (*FOXM1*) mRNA in clear cell RCC tumor tissue is an independent prognostic factor for overall patient survival, associated with a threefold increased risk of death [400]. Another interesting marker gene is the gene for the vascular cell adhesion molecule VCAM1. Interestingly, VCAM1 was found to be upregulated in clear cell RCC and papillary RCC while it was downregulated in chromophobe RCC and oncocytomas compared with normal kidney tissue. A high RNA expression level of VCAM1 specifically in clear cell RCC was an independent predictor of longer disease-free survival, while patients with a lower expression of VCAM1 had a 2.5-fold increased risk of disease recurrence [401].

5.5 Summary

Table 5.1 provides a summary of the classification of biomarkers (see Section 5.1) for bladder cancer, prostate cancer, and RCC.

Table 5.1 Summary of the classification of biomarkers for bladder cancer, prostate cancer, and RCC.

Marker	Detection and screening	Diagnostic	Prognostic	Predictive	Therapeutic target	Surrogate endpoint	Reference
Bladder cancer							
SNPs		×	(×)				[42]
GSTM1			(×)				[402]
NAT2			(×)				[403]
miRNA targeting FGFR3 (i.e., miR-145, miR-101, miR-99a, and miR-100)		×	(×)				[49]
miR-200 family		(×)	(×)				[50]
miR-143		(×)	(×)				[51]
miRNA involved in the p53 pathway		(×)	(×)				[52]
HRAS			×		(×)		[404,405]
FGFR3			×		(×)		[21]
p53			×				[21]
Rb			×				[21]
TSC1 and TSC2					(×)		[84]
AKT1					(×)		[97,98]
Prostate cancer							
BRCA2 germline mutation	×		×				[196]
HOXB13 germline variant	×						[200]
MYC polymorphisms	×						[204]
RNase L (HPC1); MSR1; ELAC2 (HPC2)	×						[194]
TMPRSS–ERG fusion transcripts	×						[240]
SNP rs11672691	×					×	[198]
SNP combinations			×				[211]
PCA3 ncRNA	×	(×)					[254]

(*continued*)

Table 5.1 (*Continued*)

Marker	Detection and screening	Diagnostic	Prognostic	Predictive	Therapeutic target	Surrogate endpoint	Reference
PCGEM1		×					[248]
PCNR1 ncRNA		×					[249]
PCAT1 ncRNA		×					[250]
miR-141 in serum	×	×	×				[266,276]
miR-375 in serum	×	×					[268]
miR-21 in serum				×			[275]
miRNAs in tissue samples		×					[278–280]
miR-221 in prostate cancer tissue			×				[281]
PTEN genomic deletion		×	×				[282,286,288]
TMPRSS2–ERG gene fusion	×	×	(×)				[304,310]
MYC amplification		×	×				[297]
NKX3.1 genomic deletion			×				[301]
RCC							
VHL germline inactivation	×	×					[324,325]
c-MET germline activating mutations	×	×					[339,340]
Fumarate hydratase germline missense mutations	×	×					[348,406]
Folliculin germline inactivating mutations	×	×					[359,360]
Chromosome 3 deletions		×					[368]
Chromosome 14q deletion			×				[371]
Trisomies of chromosomes 7, 14 and Y		×					[340,376]
Monosomies of chromosomes 1, 2, 6, 10, 13, 17, and 21		×					[377]
TFE3 translocations		×	×				[381,382]

SNPs rs11894252 and rs7579899	×				×	[383]
SNPs rs718314 and rs1049380	×				×	[384]
SNP rs12105918	×				×	[386]
miR-210 in serum	×	×		×		[388]
Cell-free DNA in serum		×				[391]
Mitochondrial DNA in serum		×				[392]
Methylation of DNA in serum		×				[393]
Angiopoietin-4 tumor specific expression		×				[394]
hTERT elevated expression		×				[395]
CD70 tumor-specific expression		×				[396]
miRNAs miR-210, miR-200c, and miR-141 in tumor tissue		×				[390,397–399]
FoxM1 expression			×			[400]
VCAM1 expression			×			[401]

Key: ×, applicable; (×), conditionally applicable.

References

1 Ilyin, S.E., Belkowski, S.M., and Plata-Salaman, C.R. (2004) Biomarker discovery and validation: technologies and integrative approaches. *Trends in Biotechnology*, **22** (8), 411–416.

2 McGuire, W.L. (1991) Breast cancer prognostic factors: evaluation guidelines. *Journal of the National Cancer Institute*, **83** (3), 154–155.

3 Altman, D.G., Lausen, B., Sauerbrei, W., and Schumacher, M. (1994) Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. *Journal of the National Cancer Institute*, **86** (11), 829–835.

4 Simon, R. and Altman, D.G. (1994) Statistical aspects of prognostic factor studies in oncology. *British Journal of Cancer*, **69** (6), 979–985.

5 Drew, P.J., Ilstrup, D.M., Kerin, M.J., and Monson, J.R. (1998) Prognostic factors: guidelines for investigation design and state of the art analytical methods. *Surgical Oncology*, **7** (1–2), 71–76.

6 Altman, D.G. and Royston, P. (2000) What do we mean by validating a prognostic model? *Statistics in Medicine*, **19** (4), 453–473.

7 Begg, C.B., Cramer, L.D., Venkatraman, E.S., and Rosai, J. (2000) Comparing tumour staging and grading systems: a case study and a review of the issues, using thymoma as a model. *Statistics in Medicine*, **19** (15), 1997–2014.

8 Pajak, T.F., Clark, G.M., Sargent, D.J., McShane, L.M., and Hammond, M.E. (2000) Statistical issues in tumor marker studies. *Archives of Pathology & Laboratory Medicine*, **124** (7), 1011–1015.

9 Schmoor, C., Sauerbrei, W., and Schumacher, M. (2000) Sample size considerations for the evaluation of prognostic factors in survival analysis. *Statistics in Medicine*, **19** (4), 441–452.

10 Pepe, M.S., Etzioni, R., Feng, Z., Potter, J.D., Thompson, M.L., Thornquist, M., Winget, M., and Yasui, Y. (2001) Phases of biomarker development for early detection of cancer. *Journal of the National Cancer Institute*, **93** (14), 1054–1061.

11 McShane, L.M., Altman, D.G., Sauerbrei, W., Taube, S.E., Gion, M., Clark, G.M., and Statistics Subcommittee of the NCI–EORTC Working Group on Cancer Diagnostics (2006) REporting recommendations for tumor MARKer prognostic studies (REMARK). *Breast Cancer Research and Treatment*, **100** (2), 229–235.

12 Goebell, P.J., Groshen, S.L., and Schmitz-Drager, B.J. (2008) Guidelines for development of diagnostic markers in bladder cancer. *World Journal of Urology*, **26** (1), 5–11.

13 Shariat, S.F., Lotan, Y., Vickers, A., Karakiewicz, P.I., Schmitz-Drager, B.J., Goebell, P.J., and Malats, N. (2010) Statistical consideration for clinical biomarker research in bladder cancer. *Urologic Oncology*, **28** (4), 389–400.

14 Shariat, S.F., Semjonow, A., Lilja, H., Savage, C., Vickers, A.J., and Bjartell, A. (2011) Tumor markers in prostate cancer I: blood-based markers. *Acta Oncologica*, **16** (Suppl.), 161–175.

15 Boyle, P. and Ferlay, J. (2005) Cancer incidence and mortality in Europe, 2004. *Annals of Oncology*, **16** (3), 481–488.

16 Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, J., Murray, T., and Thun, M.J. (2008) Cancer statistics, 2008. *CA: A Cancer Journal for Clinicians*, **58** (2), 71–96.

17 Prout, G.R. Jr., Barton, B.A., Griffin, P.P., and Friedell, G.H. (1992) Treated history of noninvasive grade 1 transitional cell carcinoma. The National Bladder Cancer Group. *Journal of Urology*, **148** (5), 1413–1419.

18 Wu, X.R. (2005) Urothelial tumorigenesis: a tale of divergent pathways. *Nature Reviews Cancer*, **5** (9), 713–725.

19 Lopez-Beltran, A. (2008) Bladder cancer: clinical and pathological profile. *Scandinavian Journal of Urology and Nephrology Supplementum*, (218), 95–109.

20 Knowles, M.A. (2006) Molecular subtypes of bladder cancer: Jekyll and

Hyde or chalk and cheese? *Carcinogenesis*, **27** (3), 361–373.
21 Goebell, P.J. and Knowles, M.A. (2010) Bladder cancer or bladder cancers? Genetically distinct malignant conditions of the urothelium. *Urologic Oncology*, **28** (4), 409–428.
22 Goebell, P.J., Villanueva, C.M., Rettenmeier, A.W., Rubben, H., and Kogevinas, M. (2004) Environmental exposure, chlorinated drinking water, and bladder cancer. *World Journal of Urology*, **21** (6), 424–432.
23 Kiriluk, K.J., Prasad, S.M., Patel, A.R., Steinberg, G.D., and Smith, N.D. (2012) Bladder cancer risk from occupational and environmental exposures. *Urologic Oncology*, **30** (2), 199–211.
24 Murta-Nascimento, C., Schmitz-Drager, B.J., Zeegers, M.P., Steineck, G., Kogevinas, M., Real, F.X., and Malats, N. (2007) Epidemiology of urinary bladder cancer: from tumor development to patient's death. *World Journal of Urology*, **25** (3), 285–295.
25 Volanis, D., Kadiyska, T., Galanis, A., Delakas, D., Logotheti, S., and Zoumpourlis, V. (2010) Environmental factors and genetic susceptibility promote urinary bladder cancer. *Toxicology Letters*, **193** (2), 131–137.
26 Kiemeney, L.A., Grotenhuis, A.J., Vermeulen, S.H., and Wu, X. (2009) Genome-wide association studies in bladder cancer: first results and potential relevance. *Current Opinion in Urology*, **19** (5), 540–546.
27 Wu, X., Hildebrandt, M.A., and Chang, D.W. (2009) Genome-wide association studies of bladder cancer risk: a field synopsis of progress and potential applications. *Cancer Metastasis Reviews*, **28** (3–4), 269–280.
28 Arand, M., Muhlbauer, R., Hengstler, J., Jager, E., Fuchs, J., Winkler, L., and Oesch, F. (1996) A multiplex polymerase chain reaction protocol for the simultaneous analysis of the glutathione S-transferase GSTM1 and GSTT1 polymorphisms. *Analytical Biochemistry*, **236** (1), 184–186.
29 Hengstler, J.G., Arand, M., Herrero, M.E., and Oesch, F. (1998) Polymorphisms of *N*-acetyltransferases, glutathione *S*-transferases, microsomal epoxide hydrolase and sulfotransferases: influence on cancer susceptibility. *Recent Results in Cancer Research*, **154**, 47–85.
30 Carmo, H., Brulport, M., Hermes, M., Oesch, F., Silva, R., Ferreira, L.M., Branco, P.S., Boer, D., Remiao, F., Carvalho, F., Schon, M.R., Krebsfaenger, N., Doehmer, J., Bastos Mde, L., and Hengstler, J.G. (2006) Influence of CYP2D6 polymorphism on 3,4-methylenedioxymethamphetamine ('Ecstasy') cytotoxicity. *Pharmacogenetics and Genomics*, **16** (11), 789–799.
31 Hewitt, N.J., Lechon, M.J., Houston, J.B., Hallifax, D., Brown, H.S., Maurel, P., Kenna, J.G., Gustavsson, L., Lohmann, C., Skonberg, C., Guillouzo, A., Tuschl, G., Li, A.P., LeCluyse, E., Groothuis, G.M., and Hengstler, J.G. (2007) Primary hepatocytes: current understanding of the regulation of metabolic enzymes and transporter proteins, and pharmaceutical practice for the use of hepatocytes in metabolism, enzyme induction, transporter, clearance, and hepatotoxicity studies. *Drug Metabolism Reviews*, **39** (1), 159–234.
32 Gehrmann, M., Schmidt, M., Brase, J.C., Roos, P., and Hengstler, J.G. (2008) Prediction of paclitaxel resistance in breast cancer: is CYP1B1*3 a new factor of influence? *Pharmacogenomics*, **9** (7), 969–974.
33 Kiemeney, L.A., Thorlacius, S., Sulem, P., Geller, F., Aben, K.K., Stacey, S.N., Gudmundsson, J., Jakobsdottir, M., Bergthorsson, J.T., Sigurdsson, A., Blondal, T., Witjes, J.A., Vermeulen, S.H., Hulsbergen-van de Kaa, C.A., Swinkels, D.W., Ploeg, M., Cornel, E.B., Vergunst, H., Thorgeirsson, T.E., Gudbjartsson, D., Gudjonsson, S.A., Thorleifsson, G., Kristinsson, K.T., Mouy, M., Snorradottir, S., Placidi, D., Campagna, M., Arici, C., Koppova, K., Gurzau, E., Rudnai, P., Kellen, E., Polidoro, S., Guarrera, S., Sacerdote, C., Sanchez, M., Saez, B., Valdivia, G., Ryk, C., de Verdier, P., Lindblom, A., Golka, K., Bishop, D.T., Knowles, M.A., Nikulasson, S., Petursdottir, V., Jonsson, E., Geirsson,

G., Kristjansson, B., Mayordomo, J.I., Steineck, G., Porru, S., Buntinx, F., Zeegers, M.P., Fletcher, T., Kumar, R., Matullo, G., Vineis, P., Kiltie, A.E., Gulcher, J.R., Thorsteinsdottir, U., Kong, A., Rafnar, T., and Stefansson, K. (2008) Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nature Genetics*, **40** (11), 1307–1312.

34 Saravana Devi, S., Vinayagamoorthy, N., Agrawal, M., Biswas, A., Biswas, R., Naoghare, P., Kumbhakar, S., Krishnamurthi, K., Hengstler, J.G., Hermes, M., and Chakrabarti, T. (2008) Distribution of detoxifying genes polymorphism in Maharastrian population of central India. *Chemosphere*, **70** (10), 1835–1839.

35 Cadenas, C., Franckenstein, D., Schmidt, M., Gehrmann, M., Hermes, M., Geppert, B., Schormann, W., Maccoux, L.J., Schug, M., Schumann, A., Wilhelm, C., Freis, E., Ickstadt, K., Rahnenfuhrer, J., Baumbach, J.I., Sickmann, A., and Hengstler, J.G. (2010) Role of thioredoxin reductase 1 and thioredoxin interacting protein in prognosis of breast cancer. *Breast Cancer Research*, **12** (3), R44.

36 Hellwig, B., Hengstler, J.G., Schmidt, M., Gehrmann, M.C., Schormann, W., and Rahnenfuhrer, J. (2010) Comparison of scores for bimodality of gene expression distributions and genome-wide evaluation of the prognostic relevance of high-scoring genes. *BMC Bioinformatics*, **11**, 276.

37 Kiemeney, L.A., Sulem, P., Besenbacher, S., Vermeulen, S.H., Sigurdsson, A., Thorleifsson, G., Gudbjartsson, D.F., Stacey, S.N., Gudmundsson, J., Zanon, C., Kostic, J., Masson, G., Bjarnason, H., Palsson, S.T., Skarphedinsson, O.B., Gudjonsson, S.A., Witjes, J.A., Grotenhuis, A.J., Verhaegh, G.W., Bishop, D.T., Sak, S.C., Choudhury, A., Elliott, F., Barrett, J.H., Hurst, C.D., de Verdier, P.J., Ryk, C., Rudnai, P., Gurzau, E., Koppova, K., Vineis, P., Polidoro, S., Guarrera, S., Sacerdote, C., Campagna, M., Placidi, D., Arici, C., Zeegers, M.P., Kellen, E., Gutierrez, B.S., Sanz-Velez, J.I., Sanchez-Zalabardo, M., Valdivia, G., Garcia-Prats, M.D., Hengstler, J.G., Blaszkewicz, M., Dietrich, H., Ophoff, R.A., van den Berg, L.H., Alexiusdottir, K., Kristjansson, K., Geirsson, G., Nikulasson, S., Petursdottir, V., Kong, A., Thorgeirsson, T., Mungan, N.A., Lindblom, A., van Es, M.A., Porru, S., Buntinx, F., Golka, K., Mayordomo, J.I., Kumar, R., Matullo, G., Steineck, G., Kiltie, A.E., Aben, K.K., Jonsson, E., Thorsteinsdottir, U., Knowles, M.A., Rafnar, T., and Stefansson, K. (2010) A sequence variant at 4p16.3 confers susceptibility to urinary bladder cancer. *Nature Genetics*, **42** (5), 415–419.

38 Golka, K., Hermes, M., Selinski, S., Blaszkewicz, M., Bolt, H.M., Roth, G., Dietrich, H., Prager, H.M., Ickstadt, K., and Hengstler, J.G. (2009) Susceptibility to urinary bladder cancer: relevance of rs9642880[T], GSTM1 0/0 and occupational exposure. *Pharmacogenetics and Genomics*, **19** (11), 903–906.

39 Rafnar, T., Sulem, P., Stacey, S.N., Geller, F., Gudmundsson, J., Sigurdsson, A., Jakobsdottir, M., Helgadottir, H., Thorlacius, S., Aben, K.K., Blondal, T., Thorgeirsson, T.E., Thorleifsson, G., Kristjansson, K., Thorisdottir, K., Ragnarsson, R., Sigurgeirsson, B., Skuladottir, H., Gudbjartsson, T., Isaksson, H.J., Einarsson, G.V., Benediktsdottir, K.R., Agnarsson, B.A., Olafsson, K., Salvarsdottir, A., Bjarnason, H., Asgeirsdottir, M., Kristinsson, K.T., Matthiasdottir, S., Sveinsdottir, S.G., Polidoro, S., Hoiom, V., Botella-Estrada, R., Hemminki, K., Rudnai, P., Bishop, D.T., Campagna, M., Kellen, E., Zeegers, M.P., de Verdier, P., Ferrer, A., Isla, D., Vidal, M.J., Andres, R., Saez, B., Juberias, P., Banzo, J., Navarrete, S., Tres, A., Kan, D., Lindblom, A., Gurzau, E., Koppova, K., de Vegt, F., Schalken, J.A., van der Heijden, H.F., Smit, H.J., Termeer, R.A., Oosterwijk, E., van Hooij, O., Nagore, E., Porru, S., Steineck, G., Hansson, J., Buntinx, F., Catalona, W.J., Matullo, G., Vineis, P., Kiltie, A.E., Mayordomo, J.I., Kumar, R., Kiemeney, L.A., Frigge, M.L., Jonsson, T., Saemundsson, H., Barkardottir, R.B., Jonsson, E., Jonsson, S., Olafsson, J.H., Gulcher, J.R., Masson, G., Gudbjartsson, D.F., Kong, A., Thorsteinsdottir, U., and Stefansson, K.

(2009) Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. *Nature Genetics*, **41** (2), 221–227.

40 Wu, X., Ye, Y., Kiemeney, L.A., Sulem, P., Rafnar, T., Matullo, G., Seminara, D., Yoshida, T., Saeki, N., Andrew, A.S., Dinney, C.P., Czerniak, B., Zhang, Z.F., Kiltie, A.E., Bishop, D.T., Vineis, P., Porru, S., Buntinx, F., Kellen, E., Zeegers, M.P., Kumar, R., Rudnai, P., Gurzau, E., Koppova, K., Mayordomo, J.I., Sanchez, M., Saez, B., Lindblom, A., de Verdier, P., Steineck, G., Mills, G.B., Schned, A., Guarrera, S., Polidoro, S., Chang, S.C., Lin, J., Chang, D.W., Hale, K.S., Majewski, T., Grossman, H.B., Thorlacius, S., Thorsteinsdottir, U., Aben, K.K., Witjes, J.A., Stefansson, K., Amos, C.I., Karagas, M.R., and Gu, J. (2009) Genetic variation in the prostate stem cell antigen gene PSCA confers susceptibility to urinary bladder cancer. *Nature Genetics*, **41** (9), 991–995.

41 Lehmann, M.L., Selinski, S., Blaszkewicz, M., Orlich, M., Ovsiannikov, D., Moormann, O., Guballa, C., Kress, A., Truss, M.C., Gerullis, H., Otto, T., Barski, D., Niegisch, G., Albers, P., Frees, S., Brenner, W., Thuroff, J.W., Angeli-Greaves, M., Seidel, T., Roth, G., Dietrich, H., Ebbinghaus, R., Prager, H.M., Bolt, H. M., Falkenstein, M., Zimmermann, A., Klein, T., Reckwitz, T., Roemer, H.C., Lohlein, D., Weistenhofer, W., Schops, W., Beg, A.E., Aslam, M., Banfi, G., Romics, I., Ickstadt, K., Schwender, H., Winterpacht, A., Hengstler, J.G., and Golka, K. (2010) Rs710521[A] on chromosome 3q28 close to TP63 is associated with increased urinary bladder cancer risk. *Archives of Toxicology*, **84** (12), 967–978.

42 Rothman, N., Garcia-Closas, M., Chatterjee, N., Malats, N., Wu, X., Figueroa, J.D., Real, F.X., Van Den Berg, D., Matullo, G., Baris, D., Thun, M., Kiemeney, L.A., Vineis, P., De Vivo, I., Albanes, D., Purdue, M.P., Rafnar, T., Hildebrandt, M.A., Kiltie, A.E., Cussenot, O., Golka, K., Kumar, R., Taylor, J.A., Mayordomo, J.I., Jacobs, K.B., Kogevinas, M., Hutchinson, A., Wang, Z., Fu, Y.P., Prokunina-Olsson, L., Burdett, L., Yeager, M., Wheeler, W., Tardon, A., Serra, C., Carrato, A., Garcia-Closas, R., Lloreta, J., Johnson, A., Schwenn, M., Karagas, M.R., Schned, A., Andriole, G. Jr., Grubb, R. 3rd, Black, A., Jacobs, E.J., Diver, W.R., Gapstur, S.M., Weinstein, S.J., Virtamo, J., Cortessis, V.K., Gago-Dominguez, M., Pike, M.C., Stern, M.C., Yuan, J.M., Hunter, D.J., McGrath, M., Dinney, C.P., Czerniak, B., Chen, M., Yang, H., Vermeulen, S.H., Aben, K.K., Witjes, J.A., Makkinje, R.R., Sulem, P., Besenbacher, S., Stefansson, K., Riboli, E., Brennan, P., Panico, S., Navarro, C., Allen, N.E., Bueno-de-Mesquita, H.B., Trichopoulos, D., Caporaso, N., Landi, M.T., Canzian, F., Ljungberg, B., Tjonneland, A., Clavel-Chapelon, F., Bishop, D.T., Teo, M.T., Knowles, M.A., Guarrera, S., Polidoro, S., Ricceri, F., Sacerdote, C., Allione, A., Cancel-Tassin, G., Selinski, S., Hengstler, J.G., Dietrich, H., Fletcher, T., Rudnai, P., Gurzau, E., Koppova, K., Bolick, S.C., Godfrey, A., Xu, Z., Sanz-Velez, J.I., García-Prats, M.D., Sanchez, M., Valdivia, G., Porru, S., Benhamou, S., Hoover, R.N., Fraumeni, J.F. Jr., Silverman, D.T., and Chanock, S.J. (2010) A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. *Nature Genetics*, **42** (11), 978–984.

43 Dudek, A.M., Grotenhuis, A.J., Vermeulen, S.H., Kiemeney, L.A., and Verhaegh, G.W. (2013) Urinary bladder cancer susceptibility markers. What do we know about functional mechanisms? *International Journal of Molecular Sciences*, **14** (6), 12346–12366.

44 Chung, C.C., Magalhaes, W.C., Gonzalez-Bosquet, J., and Chanock, S.J. (2010) Genome-wide association studies in cancer – current and future directions. *Carcinogenesis*, **31** (1), 111–120.

45 Sanchez-Carbayo, M. (2012) Urine epigenomics: a promising path for bladder cancer diagnostics. *Expert Review of Molecular Diagnostics*, **12** (5), 429–432.

46 Shimizu, T., Suzuki, H., Nojima, M., Kitamura, H., Yamamoto, E., Maruyama, R., Ashida, M., Hatahira, T., Kai, M.,

Masumori, N., Tokino, T., Imai, K., Tsukamoto, T., and Toyota, M. (2013) Methylation of a panel of microRNA genes is a novel biomarker for detection of bladder cancer. *European Urology*, **63** (6), 1091–1100.

47 Schaefer, A., Stephan, C., Busch, J., Yousef, G.M., and Jung, K. (2010) Diagnostic, prognostic and therapeutic implications of microRNAs in urologic tumors. *Nature Reviews Urology*, **7** (5), 286–297.

48 Fendler, A., Stephan, C., Yousef, G.M., and Jung, K. (2011) MicroRNAs as regulators of signal transduction in urological tumors. *Clinical Chemistry*, **57** (7), 954–968.

49 Catto, J.W., Miah, S., Owen, H.C., Bryant, H., Myers, K., Dudziec, E., Larre, S., Milo, M., Rehman, I., Rosario, D.J., Di Martino, E., Knowles, M.A., Meuth, M., Harris, A.L., and Hamdy, F.C. (2009) Distinct microRNA alterations characterize high- and low-grade bladder cancer. *Cancer Research*, **69** (21), 8472–8481.

50 Adam, L., Zhong, M., Choi, W., Qi, W., Nicoloso, M., Arora, A., Calin, G., Wang, H., Siefker-Radtke, A., McConkey, D., Bar-Eli, M., and Dinney, C. (2009) miR-200 expression regulates epithelial-to-mesenchymal transition in bladder cancer cells and reverses resistance to epidermal growth factor receptor therapy. *Clinical Cancer Research*, **15** (16), 5060–5072.

51 Lin, T., Dong, W., Huang, J., Pan, Q., Fan, X., Zhang, C., and Huang, L. (2009) MicroRNA-143 as a tumor suppressor for bladder cancer. *Journal of Urology*, **181** (3), 1372–1380.

52 Veerla, S., Lindgren, D., Kvist, A., Frigyesi, A., Staaf, J., Persson, H., Liedberg, F., Chebil, G., Gudjonsson, S., Borg, A., Mansson, W., Rovira, C., and Hoglund, M. (2009) MiRNA expression in urothelial carcinomas: important roles of miR-10a, miR-222, miR-125b, miR-7 and miR-452 for tumor stage and metastasis, and frequent homozygous losses of miR-31. *International Journal of Cancer*, **124** (9), 2236–2242.

53 Miah, S., Dudziec, E., Drayton, R.M., Zlotta, A.R., Morgan, S.L., Rosario, D.J., Hamdy, F.C., and Catto, J.W. (2012) An evaluation of urinary microRNA reveals a high sensitivity for bladder cancer. *British Journal of Cancer*, **107** (1), 123–128.

54 Hanke, M., Hoefig, K., Merz, H., Feller, A.C., Kausch, I., Jocham, D., Warnecke, J.M., and Sczakiel, G. (2010) A robust methodology to study urine microRNA as tumor marker: microRNA-126 and microRNA-182 are related to urinary bladder cancer. *Urologic Oncology*, **28** (6), 655–661.

55 Yamada, Y., Enokida, H., Kojima, S., Kawakami, K., Chiyomaru, T., Tatarano, S., Yoshino, H., Kawahara, K., Nishiyama, K., Seki, N., and Nakagawa, M. (2011) MiR-96 and miR-183 detection in urine serve as potential tumor markers of urothelial carcinoma: correlation with stage and grade, and comparison with urinary cytology. *Cancer Science*, **102** (3), 522–529.

56 Wang, G., Chan, E.S., Kwan, B.C., Li, P.K., Yip, S.K., Szeto, C.C., and Ng, C.F. (2012) Expression of microRNAs in the urine of patients with bladder cancer. *Clinical Genitourinary Cancer*, **10** (2), 106–113.

57 Yun, S.J., Jeong, P., Kim, W.T., Kim, T.H., Lee, Y.S., Song, P.H., Choi, Y.H., Kim, I.Y., Moon, S.K., and Kim, W.J. (2012) Cell-free microRNAs in urine as diagnostic and prognostic biomarkers of bladder cancer. *International Journal of Oncology*, **41** (5), 1871–1878.

58 Puerta-Gil, P., Garcia-Baquero, R., Jia, A.Y., Ocana, S., Alvarez-Mugica, M., Alvarez-Ossorio, J.L., Cordon-Cardo, C., Cava, F., and Sanchez-Carbayo, M. (2012) miR-143, miR-222, and miR-452 are useful as tumor stratification and noninvasive diagnostic biomarkers for bladder cancer. *American Journal of Pathology*, **180** (5), 1808–1815.

59 Tolle, A., Jung, M., Rabenhorst, S., Kilic, E., Jung, K., and Weikert, S. (2013) Identification of microRNAs in blood and urine as tumour markers for the detection of urinary bladder cancer. *Oncology Reports*, **30** (4), 1949–1956.

60 Mengual, L., Lozano, J.J., Ingelmo-Torres, M., Gazquez, C., Ribal, M.J., and Alcaraz, A. (2013) Using microRNA profiling in

urine samples to develop a non-invasive test for bladder cancer. *International Journal of Cancer*, **133** (11), 2631–2641.

61 Adam, L., Wszolek, M.F., Liu, C.G., Jing, W., Diao, L., Zien, A., Zhang, J.D., Jackson, D., and Dinney, C.P. (2012) Plasma microRNA profiles for bladder cancer detection. *Urologic Oncology*, **31** (8), 1701–1708.

62 Johnson, D.E. and Williams, L.T. (1993) Structural and functional diversity in the FGF receptor multigene family. *Advances in Cancer Research*, **60** (1), 1–41.

63 Cappellen, D., De Oliveira, C., Ricol, D., de Medina, S., Bourdin, J., Sastre-Garau, X., Chopin, D., Thiery, J.P., and Radvanyi, F. (1999) Frequent activating mutations of FGFR3 in human bladder and cervix carcinomas. *Nature Genetics*, **23** (1), 18–20.

64 Billerey, C., Chopin, D., Aubriot-Lorton, M.H., Ricol, D., Gil Diez de Medina, S., Van Rhijn, B., Bralet, M.P., Lefrere-Belda, M.A., Lahaye, J.B., Abbou, C.C., Bonaventure, J., Zafrani, E.S., van der Kwast, T., Thiery, J.P., and Radvanyi, F. (2001) Frequent FGFR3 mutations in papillary non-invasive bladder (pTa) tumors. *American Journal of Pathology*, **158** (6), 1955–1959.

65 Kimura, T., Suzuki, H., Ohashi, T., Asano, K., Kiyota, H., and Eto, Y. (2001) The incidence of thanatophoric dysplasia mutations in FGFR3 gene is higher in low-grade or superficial bladder carcinomas. *Cancer*, **92** (10), 2555–2561.

66 Sibley, K., Cuthbert-Heavens, D., and Knowles, M.A. (2001) Loss of heterozygosity at 4p16.3 and mutation of FGFR3 in transitional cell carcinoma. *Oncogene*, **20** (6), 686–691.

67 van Rhijn, B.W., Lurkin, I., Radvanyi, F., Kirkels, W.J., van der Kwast, T.H., and Zwarthoff, E.C. (2001) The fibroblast growth factor receptor 3 (FGFR3) mutation is a strong indicator of superficial bladder cancer with low recurrence rate. *Cancer Research*, **61** (4), 1265–1268.

68 van Rhijn, B.W., Montironi, R., Zwarthoff, E.C., Jobsis, A.C., and van der Kwast, T.H. (2002) Frequent FGFR3 mutations in urothelial papilloma. *Journal of Pathology*, **198** (2), 245–251.

69 van Rhijn, B.W., van Tilborg, A.A., Lurkin, I., Bonaventure, J., de Vries, A., Thiery, J.P., van der Kwast, T.H., Zwarthoff, E.C., and Radvanyi, F. (2002) Novel fibroblast growth factor receptor 3 (FGFR3) mutations in bladder cancer previously identified in non-lethal skeletal disorders. *European Journal of Human Genetics*, **10** (12), 819–824.

70 Bakkar, A.A., Wallerand, H., Radvanyi, F., Lahaye, J.B., Pissard, S., Lecerf, L., Kouyoumdjian, J.C., Abbou, C.C., Pairon, J.C., Jaurand, M.C., Thiery, J.P., Chopin, D.K., and de Medina, S.G. (2003) FGFR3 and TP53 gene mutations define two distinct pathways in urothelial cell carcinoma of the bladder. *Cancer Research*, **63** (23), 8108–8112.

71 Rieger-Christ, K.M., Mourtzinos, A., Lee, P.J., Zagha, R.M., Cain, J., Silverman, M., Libertino, J.A., and Summerhayes, I.C. (2003) Identification of fibroblast growth factor receptor 3 mutations in urine sediment DNA samples complements cytology in bladder tumor detection. *Cancer*, **98** (4), 737–744.

72 van Rhijn, B.W., Lurkin, I., Chopin, D.K., Kirkels, W.J., Thiery, J.P., van der Kwast, T.H., Radvanyi, F., and Zwarthoff, E.C. (2003) Combined microsatellite and FGFR3 mutation analysis enables a highly sensitive detection of urothelial cell carcinoma in voided urine. *Clinical Cancer Research*, **9** (1), 257–263.

73 van Rhijn, B.W., Vis, A.N., van der Kwast, T.H., Kirkels, W.J., Radvanyi, F., Ooms, E.C., Chopin, D.K., Boeve, E.R., Jobsis, A.C., and Zwarthoff, E.C. (2003) Molecular grading of urothelial cell carcinoma with fibroblast growth factor receptor 3 and MIB-1 is superior to pathologic grade for the prediction of clinical outcome. *Journal of Clinical Oncology*, **21** (10), 1912–1921.

74 Hernandez, S., Lopez-Knowles, E., Lloreta, J., Kogevinas, M., Jaramillo, R., Amoros, A., Tardon, A., Garcia-Closas, R., Serra, C., Carrato, A., Malats, N., and Real, F.X. (2005) FGFR3 and Tp53 mutations in T1G3 transitional bladder carcinomas: independent distribution and

lack of association with prognosis. *Clinical Cancer Research*, **11** (15), 5444–5450.

75 Wallerand, H., Bakkar, A.A., de Medina, S.G., Pairon, J.C., Yang, Y.C., Vordos, D., Bittard, H., Fauconnet, S., Kouyoumdjian, J.C., Jaurand, M.C., Zhang, Z.F., Radvanyi, F., Thiery, J.P., and Chopin, D.K. (2005) Mutations in TP53, but not FGFR3, in urothelial cell carcinoma of the bladder are influenced by smoking: contribution of exogenous versus endogenous carcinogens. *Carcinogenesis*, **26** (1), 177–184.

76 Jebar, A.H., Hurst, C.D., Tomlinson, D.C., Johnston, C., Taylor, C.F., and Knowles, M.A. (2005) FGFR3 and Ras gene mutations are mutually exclusive genetic events in urothelial cell carcinoma. *Oncogene*, **24** (33), 5218–5225.

77 Hernandez, S., Lopez-Knowles, E., Lloreta, J., Kogevinas, M., Amoros, A., Tardon, A., Carrato, A., Serra, C., Malats, N., and Real, F.X. (2006) Prospective study of FGFR3 mutations as a prognostic factor in nonmuscle invasive urothelial bladder carcinomas. *Journal of Clinical Oncology*, **24** (22), 3664–3671.

78 Lamy, A., Gobet, F., Laurent, M., Blanchard, F., Varin, C., Moulin, C., Andreou, A., Frebourg, T., and Pfister, C. (2006) Molecular profiling of bladder tumors based on the detection of FGFR3 and TP53 mutations. *Journal of Urology*, **176** (6 Pt 1), 2686–2689.

79 Lindgren, D., Liedberg, F., Andersson, A., Chebil, G., Gudjonsson, S., Borg, A., Mansson, W., Fioretos, T., and Hoglund, M. (2006) Molecular characterization of early-stage bladder carcinomas by expression profiles, FGFR3 mutation status, and loss of 9q. *Oncogene*, **25** (18), 2685–2696.

80 Eble, J.N., Sauter, G., Epstein, J.I., and Sesterhenn, I.A. (2004) *Pathology and Genetics of Tumours of the Urinary System and Male Genital Organs*, 3rd edn, IARC Press, Lyon.

81 Tomlinson, D.C., Baldo, O., Harnden, P., and Knowles, M.A. (2007) FGFR3 protein expression and its relationship to mutation status and prognostic variables in bladder cancer. *Journal of Pathology*, **213** (1), 91–98.

82 Hart, K.C., Robertson, S.C., Kanemitsu, M.Y., Meyer, A.N., Tynan, J.A., and Donoghue, D.J. (2000) Transformation and Stat activation by derivatives of FGFR1, FGFR3, and FGFR4. *Oncogene*, **19** (29), 3309–3320.

83 Lopez-Knowles, E., Hernandez, S., Malats, N., Kogevinas, M., Lloreta, J., Carrato, A., Tardon, A., Serra, C., and Real, F.X. (2006) *PIK3CA* mutations are an early genetic alteration associated with FGFR3 mutations in superficial papillary bladder tumors. *Cancer Research*, **66** (15), 7401–7404.

84 Platt, F.M., Hurst, C.D., Taylor, C.F., Gregory, W.M., Harnden, P., and Knowles, M.A. (2009) Spectrum of phosphatidylinositol 3-kinase pathway gene alterations in bladder cancer. *Clinical Cancer Research*, **15** (19), 6008–6017.

85 Hurst, C.D., Zuiverloon, T.C., Hafner, C., Zwarthoff, E.C., and Knowles, M.A. (2009) A SNaPshot assay for the rapid and simple detection of four common hotspot codon mutations in the *PIK3CA* gene. *BMC Research Notes*, **2**, 66.

86 Mozaffari, M., Hoogeveen-Westerveld, M., Kwiatkowski, D., Sampson, J., Ekong, R., Povey, S., den Dunnen, J.T., van den Ouweland, A., Halley, D., and Nellist, M. (2009) Identification of a region required for TSC1 stability by functional analysis of TSC1 missense mutations found in individuals with tuberous sclerosis complex. *BMC Medical Genetics*, **10**, 88.

87 Nellist, M., van den Heuvel, D., Schluep, D., Exalto, C., Goedbloed, M., Maat-Kievit, A., van Essen, T., van Spaendonck-Zwarts, K., Jansen, F., Helderman, P., Bartalini, G., Vierimaa, O., Penttinen, M., van den Ende, J., van den Ouweland, A., and Halley, D. (2009) Missense mutations to the TSC1 gene cause tuberous sclerosis complex. *European Journal of Human Genetics*, **17** (3), 319–328.

88 Henske, E.P., Scheithauer, B.W., Short, M.P., Wollmann, R., Nahmias, J., Hornigold, N., van Slegtenhorst, M., Welsh, C.T., and Kwiatkowski, D.J.

(1996) Allelic loss is frequent in tuberous sclerosis kidney lesions but rare in brain lesions. *American Journal of Human Genetics*, **59** (2), 400–406.

89 Hornigold, N., Devlin, J., Davies, A.M., Aveyard, J.S., Habuchi, T., and Knowles, M.A. (1999) Mutation of the 9q34 gene TSC1 in sporadic bladder cancer. *Oncogene*, **18** (16), 2657–2661.

90 Adachi, H., Igawa, M., Shiina, H., Urakami, S., Shigeno, K., and Hino, O. (2003) Human bladder tumors with 2-hit mutations of tumor suppressor gene TSC1 and decreased expression of p27. *Journal of Urology*, **170** (2), 601–604.

91 Cairns, P., Shaw, M.E., and Knowles, M.A. (1993) Initiation of bladder cancer may involve deletion of a tumour-suppressor gene on chromosome 9. *Oncogene*, **8** (4), 1083–1085.

92 Habuchi, T., Devlin, J., Elder, P.A., and Knowles, M.A. (1995) Detailed deletion mapping of chromosome 9q in bladder cancer: evidence for two tumour suppressor loci. *Oncogene*, **11** (8), 1671–1674.

93 Simoneau, M., Aboulkassim, T.O., LaRue, H., Rousseau, F., and Fradet, Y. (1999) Four tumor suppressor loci on chromosome 9q in bladder cancer: evidence for two novel candidate regions at 9q22.3 and 9q31. *Oncogene*, **18** (1), 157–163.

94 van Tilborg, A.A., Groenfeld, L.E., van der Kwast, T.H., and Zwarthoff, E.C. (1999) Evidence for two candidate tumour suppressor loci on chromosome 9q in transitional cell carcinoma (TCC) of the bladder but no homozygous deletions in bladder tumour cell lines. *British Journal of Cancer*, **80** (3–4), 489–494.

95 Knowles, M.A., Habuchi, T., Kennedy, W., and Cuthbert-Heavens, D. (2003) Mutation spectrum of the 9q34 tuberous sclerosis gene TSC1 in transitional cell carcinoma of the bladder. *Cancer Research*, **63** (22), 7652–7656.

96 Pymar, L.S., Platt, F.M., Askham, J.M., Morrison, E.E., and Knowles, M.A. (2008) Bladder tumour-derived somatic TSC1 missense mutations cause loss of function via distinct mechanisms. *Human Molecular Genetics*, **17** (13), 2006–2017.

97 Carpten, J.D., Faber, A.L., Horn, C., Donoho, G.P., Briggs, S.L., Robbins, C.M., Hostetter, G., Boguslawski, S., Moses, T. Y., Savage, S., Uhlik, M., Lin, A., Du, J., Qian, Y.W., Zeckner, D.J., Tucker-Kellogg, G., Touchman, J., Patel, K., Mousses, S., Bittner, M., Schevitz, R., Lai, M.H., Blanchard, K.L., and Thomas, J.E. (2007) A transforming mutation in the pleckstrin homology domain of AKT1 in cancer. *Nature*, **448** (7152), 439–444.

98 Askham, J.M., Platt, F., Chambers, P.A., Snowden, H., Taylor, C.F., and Knowles, M.A. (2010) AKT1 mutations in bladder cancer: identification of a novel oncogenic mutation that can co-operate with E17K. *Oncogene*, **29** (1), 150–155.

99 Fujimoto, K., Yamada, Y., Okajima, E., Kakizoe, T., Sasaki, H., Sugimura, T., and Terada, M. (1992) Frequent association of p53 gene mutation in invasive bladder cancer. *Cancer Research*, **52** (6), 1393–1398.

100 Spruck, C.H. 3rd, Ohneseit, P.F., Gonzalez-Zulueta, M., Esrig, D., Miyao, N., Tsai, Y.C., Lerner, S.P., Schmutte, C., Yang, A.S., Cote, R. *et al.* (1994) Two molecular pathways to transitional cell carcinoma of the bladder. *Cancer Research*, **54** (3), 784–788.

101 Uchida, T., Wada, C., Ishida, H., Wang, C., Egawa, S., Yokoyama, E., Kameya, T., and Koshiba, K. (1995) p53 mutations and prognosis in bladder tumors. *Journal of Urology*, **153** (4), 1097–1104.

102 Soussi, T. and Beroud, C. (2001) Assessing TP53 status in human tumours to evaluate clinical outcome. *Nature Reviews Cancer*, **1** (3), 233–240.

103 Petitjean, A., Achatz, M.I., Borresen-Dale, A.L., Hainaut, P., and Olivier, M. (2007) TP53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes. *Oncogene*, **26** (15), 2157–2165.

104 Olivier, M., Eeles, R., Hollstein, M., Khan, M.A., Harris, C.C., and Hainaut, P. (2002) The IARC TP53 database: new online mutation analysis and recommendations to users. *Human Mutation*, **19** (6), 607–614.

105 Blandino, G., Levine, A.J., and Oren, M. (1999) Mutant p53 gain of function: differential effects of different p53 mutants on resistance of cultured cells to chemotherapy. *Oncogene*, **18** (2), 477–485.

106 Strano, S., Dell'Orso, S., Di Agostino, S., Fontemaggi, G., Sacchi, A., and Blandino, G. (2007) Mutant p53: an oncogenic transcription factor. *Oncogene*, **26** (15), 2212–2219.

107 Weisz, L., Oren, M., and Rotter, V. (2007) Transcription regulation by mutant p53. *Oncogene*, **26** (15), 2202–2211.

108 Soussi, T. (2007) p53 alterations in human cancer: more questions than answers. *Oncogene*, **26** (15), 2145–2156.

109 George, B., Datar, R.H., Wu, L., Cai, J., Patten, N., Beil, S.J., Groshen, S., Stein, J., Skinner, D., Jones, P.A., and Cote, R.J. (2007) p53 gene and protein status: the role of p53 alterations in predicting outcome in patients with bladder cancer. *Journal of Clinical Oncology*, **25** (34), 5352–5358.

110 Kelsey, K.T., Hirao, T., Schned, A., Hirao, S., Devi-Ashok, T., Nelson, H.H., Andrew, A., and Karagas, M.R. (2004) A population-based study of immunohistochemical detection of p53 alteration in bladder cancer. *British Journal of Cancer*, **90** (8), 1572–1576.

111 Lopez-Knowles, E., Hernandez, S., Kogevinas, M., Lloreta, J., Amoros, A., Tardon, A., Carrato, A., Kishore, S., Serra, C., Malats, N., Real, F.X., and Investigators, E.S. (2006) The p53 pathway and outcome among patients with T1G3 bladder tumors. *Clinical Cancer Research*, **12** (20), 6029–6036.

112 Simon, R., Struckmann, K., Schraml, P., Wagner, U., Forster, T., Moch, H., Fijan, A., Bruderer, J., Wilber, K., Mihatsch, M. J., Gasser, T., and Sauter, G. (2002) Amplification pattern of 12q13–q15 genes (MDM2, CDK4,GLI) in urinary bladder cancer. *Oncogene*, **21** (16), 2476–2483.

113 Liang, H., Atkins, H., Abdel-Fattah, R., Jones, S.N., and Lunec, J. (2004) Genomic organisation of the human MDM2 oncogene and relationship to its alternatively spliced mRNAs. *Gene*, **338** (2), 217–223.

114 Schlott, T., Quentin, T., Korabiowska, M., Budd, B., and Kunze, E. (2004) Alteration of the MDM2–p73–$P14^{ARF}$ pathway related to tumour progression during urinary bladder carcinogenesis. *International Journal of Molecular Medicine*, **14** (5), 825–836.

115 Sanchez-Carbayo, M., Socci, N.D., Kirchoff, T., Erill, N., Offit, K., Bochner, B.H., and Cordon-Cardo, C. (2007) A polymorphism in HDM2 (SNP309) associates with early onset in superficial tumors, TP53 mutations, and poor outcome in invasive bladder cancer. *Clinical Cancer Research*, **13** (11), 3215–3220.

116 Markl, I.D. and Jones, P.A. (1998) Presence and location of TP53 mutation determines pattern of CDKN2A/ARF pathway inactivation in bladder cancer. *Cancer Research*, **58** (23), 5348–5353.

117 Berggren, P., Kumar, R., Sakano, S., Hemminki, L., Wada, T., Steineck, G., Adolfsson, J., Larsson, P., Norming, U., Wijkstrom, H., and Hemminki, K. (2003) Detecting homozygous deletions in the CDKN2A($p16^{INK4a}$)/ARF($p14^{ARF}$) gene in urinary bladder cancer using real-time quantitative PCR. *Clinical Cancer Research*, **9** (1), 235–242.

118 Chang, L.L., Yeh, W.T., Yang, S.Y., Wu, W.J., and Huang, C.H. (2003) Genetic alterations of $p16^{INK4a}$ and $p14^{ARF}$ genes in human bladder cancer. *Journal of Urology*, **170** (2), 595–600.

119 Le Frere-Belda, M.A., Gil Diez de Medina, S., Daher, A., Martin, N., Albaud, B., Heudes, D., Abbou, C.C., Thiery, J.P., Zafrani, E.S., Radvanyi, F., and Chopin, D. (2004) Profiles of the 2 INK4a gene products, p16 and $p14^{ARF}$, in human reference urothelium and bladder carcinomas, according to pRb and p53 protein status. *Human Pathology*, **35** (7), 817–824.

120 Kawamoto, K., Enokida, H., Gotanda, T., Kubo, H., Nishiyama, K., Kawahara, M., and Nakagawa, M. (2006) $p16^{INK4a}$ and $p14^{ARF}$ methylation as a potential biomarker for human bladder cancer.

Biochemical and Biophysical Research Communications, **339** (3), 790–796.

121 Stein, J.P., Ginsberg, D.A., Grossfeld, G.D., Chatterjee, S.J., Esrig, D., Dickinson, M.G., Groshen, S., Taylor, C.R., Jones, P.A., Skinner, D.G., and Cote, R.J. (1998) Effect of $p21^{WAF1/CIP1}$ expression on tumor progression in bladder cancer. *Journal of the National Cancer Institute*, **90** (14), 1072–1079.

122 Chatterjee, S.J., Datar, R., Youssefzadeh, D., George, B., Goebell, P.J., Stein, J.P., Young, L., Shi, S.R., Gee, C., Groshen, S., Skinner, D.G., and Cote, R.J. (2004) Combined effects of p53, p21, and pRb expression in the progression of bladder transitional cell carcinoma. *Journal of Clinical Oncology*, **22** (6), 1007–1013.

123 Shariat, S.F., Chade, D.C., Karakiewicz, P.I., Ashfaq, R., Isbarn, H., Fradet, Y., Bastian, P.J., Nielsen, M.E., Capitanio, U., Jeldres, C., Montorsi, F., Lerner, S.P., Sagalowsky, A.I., Cote, R.J., and Lotan, Y. (2010) Combination of multiple molecular markers can improve prognostication in patients with locally advanced and lymph node positive bladder cancer. *Journal of Urology*, **183** (1), 68–75.

124 Cairns, P., Proctor, A.J., and Knowles, M. A. (1991) Loss of heterozygosity at the RB locus is frequent and correlates with muscle invasion in bladder carcinoma. *Oncogene*, **6** (12), 2305–2309.

125 Benedict, W.F., Lerner, S.P., Zhou, J., Shen, X., Tokunaga, H., and Czerniak, B. (1999) Level of retinoblastoma protein expression correlates with p16 (MTS-1/INK4A/CDKN2) status in bladder cancer. *Oncogene*, **18** (5), 1197–1203.

126 Shariat, S.F., Tokunaga, H., Zhou, J., Kim, J., Ayala, G.E., Benedict, W.F., and Lerner, S.P. (2004) p53, p21, pRB, and p16 expression predict clinical outcome in cystectomy with bladder cancer. *Journal of Clinical Oncology*, **22** (6), 1014–1024.

127 Cordon-Cardo, C., Wartinger, D., Petrylak, D., Dalbagni, G., Fair, W.R., Fuks, Z., and Reuter, V.E. (1992) Altered expression of the retinoblastoma gene product: prognostic indicator in bladder cancer. *Journal of the National Cancer Institute*, **84** (16), 1251–1256.

128 Logothetis, C.J., Xu, H.J., Ro, J.Y., Hu, S.X., Sahin, A., Ordonez, N., and Benedict, W.F. (1992) Altered expression of retinoblastoma protein and known prognostic variables in locally advanced bladder cancer. *Journal of the National Cancer Institute*, **84** (16), 1256–1261.

129 Hurst, C.D., Tomlinson, D.C., Williams, S.V., Platt, F.M., and Knowles, M.A. (2008) Inactivation of the Rb pathway and overexpression of both isoforms of E2F3 are obligate events in bladder tumours with 6p22 amplification. *Oncogene*, **27** (19), 2716–2727.

130 Feber, A., Clark, J., Goodwin, G., Dodson, A.R., Smith, P.H., Fletcher, A., Edwards, S., Flohr, P., Falconer, A., Roe, T., Kovacs, G., Dennis, N., Fisher, C., Wooster, R., Huddart, R., Foster, C.S., and Cooper, C.S. (2004) Amplification and overexpression of E2F3 in human bladder cancer. *Oncogene*, **23** (8), 1627–1630.

131 Oeggerli, M., Tomovska, S., Schraml, P., Calvano-Forte, D., Schafroth, S., Simon, R., Gasser, T., Mihatsch, M.J., and Sauter, G. (2004) E2F3 amplification and overexpression is associated with invasive tumor growth and rapid tumor cell proliferation in urinary bladder cancer. *Oncogene*, **23** (33), 5616–5623.

132 Steinthorsdottir, V., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Jonsdottir, T., Walters, G.B., Styrkarsdottir, U., Gretarsdottir, S., Emilsson, V., Ghosh, S., Baker, A., Snorradottir, S., Bjarnason, H., Ng, M.C., Hansen, T., Bagger, Y., Wilensky, R.L., Reilly, M.P., Adeyemo, A., Chen, Y., Zhou, J., Gudnason, V., Chen, G., Huang, H., Lashley, K., Doumatey, A., So, W.Y., Ma, R.C., Andersen, G., Borch-Johnsen, K., Jorgensen, T., van Vliet-Ostaptchouk, J.V., Hofker, M.H., Wijmenga, C., Christiansen, C., Rader, D.J., Rotimi, C., Gurney, M., Chan, J.C., Pedersen, O., Sigurdsson, G., Gulcher, J.R., Thorsteinsdottir, U., Kong, A., and Stefansson, K. (2007) A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nature Genetics*, **39** (6), 770–775.

133 Oeggerli, M., Schraml, P., Ruiz, C., Bloch, M., Novotny, H., Mirlacher, M., Sauter, G., and Simon, R. (2006) E2F3 is the main target gene of the 6p22 amplicon with high specificity for human bladder cancer. *Oncogene*, **25** (49), 6538–6543.

134 Olsson, A.Y., Feber, A., Edwards, S., Te Poele, R., Giddings, I., Merson, S., and Cooper, C.S. (2007) Role of E2F3 expression in modulating cellular proliferation rate in human bladder and prostate cancer cells. *Oncogene*, **26** (7), 1028–1037.

135 Xu, H.J., Cairns, P., Hu, S.X., Knowles, M.A., and Benedict, W.F. (1993) Loss of RB protein expression in primary bladder cancer correlates with loss of heterozygosity at the RB locus and tumor progression. *International Journal of Cancer*, **53** (5), 781–784.

136 Adams, J., Williams, S.V., Aveyard, J.S., and Knowles, M.A. (2005) Loss of heterozygosity analysis and DNA copy number measurement on 8p in bladder cancer reveals two mechanisms of allelic loss. *Cancer Research*, **65** (1), 66–75.

137 Choi, C., Kim, M.H., Juhng, S.W., and Oh, B.R. (2000) Loss of heterozygosity at chromosome segments 8p22 and 8p11.2–21.1 in transitional-cell carcinoma of the urinary bladder. *International Journal of Cancer*, **86**, 501–505.

138 Stoehr, R., Wissmann, C., Suzuki, H., Knuechel, R., Krieg, R.C., Klopocki, E., Dahl, E., Wild, P., Blaszyk, H., Sauter, G., Simon, R., Schmitt, R., Zaak, D., Hofstaedter, F., Rosenthal, A., Baylin, S. B., Pilarsky, C., and Hartmann, A. (2004). Deletions of chromosome 8p and loss of sFRP1 expression are progression markers of papillary bladder cancer. *Laboratory Investigation*, **84**, 465–478.

139 Takle, L.A. and Knowles, M.A. (1996) Deletion mapping implicates two tumor suppressor genes on chromosome 8p in the development of bladder cancer. *Oncogene*, **12** (5), 1083–1087.

140 Eydmann, M.E. and Knowles, M.A. (1997) Mutation analysis of 8p genes POLB and PPP2CB in bladder cancer. *Cancer Genetics and Cytogenetics*, **93** (2), 167–171.

141 Thompson, T.E., Rogan, P.K., Risinger, J.I., and Taylor, J.A. (2002) Splice variants but not mutations of DNA polymerase beta are common in bladder cancer. *Cancer Research*, **62** (11), 3251–3256.

142 Adams, J., Cuthbert-Heavens, D., Bass, S., and Knowles, M.A. (2005) Infrequent mutation of TRAIL receptor 2 (TRAIL-R2/DR5) in transitional cell carcinoma of the bladder with 8p21 loss of heterozygosity. *Cancer Letters*, **220** (2), 137–144.

143 Knowles, M.A., Aveyard, J.S., Taylor, C.F., Harnden, P., and Bass, S. (2005) Mutation analysis of the 8p candidate tumour suppressor genes DBC2 (RHOBTB2) and LZTS1 in bladder cancer. *Cancer Letters*, **225** (1), 121–130.

144 Richter, J., Beffa, L., Wagner, U., Schraml, P., Gasser, T.C., Moch, H., Mihatsch, M.J., and Sauter, G. (1998) Patterns of chromosomal imbalances in advanced urinary bladder cancer detected by comparative genomic hybridization. *American Journal of Pathology*, **153** (5), 1615–1621.

145 Blaveri, E., Brewer, J.L., Roydasgupta, R., Fridlyand, J., DeVries, S., Koppie, T., Pejavar, S., Mehta, K., Carroll, P., Simko, J.P., and Waldman, F.M. (2005) Bladder cancer stage and outcome by array-based comparative genomic hybridization. *Clinical Cancer Research*, **11** (19), 7012–7022.

146 Cappellen, D., Gil Diez de Medina, S., Chopin, D., Thiery, J.P., and Radvanyi, F. (1997) Frequent loss of heterozygosity on chromosome 10q in muscle-invasive transitional cell carcinomas of the bladder. *Oncogene*, **14** (25), 3059–3066.

147 Kagan, J., Liu, J., Stein, J.D., Wagner, S.S., Babkowski, R., Grossman, B.H., and Katz, R.L. (1998) Cluster of allele losses within a 2.5 cM region of chromosome 10 in high-grade invasive bladder cancer. *Oncogene*, **16** (7), 909–913.

148 Aveyard, J.S., Skilleter, A., Habuchi, T., and Knowles, M.A. (1999) Somatic mutation of PTEN in bladder carcinoma. *British Journal of Cancer*, **80** (5–6), 904–908.

149 Liu, J., Babaian, D.C., Liebert, M., Steck, P.A., and Kagan, J. (2000) Inactivation of

MMAC1 in bladder transitional-cell carcinoma cell lines and specimens. *Molecular Carcinogenesis*, **29** (3), 143–150.

150 Wang, D.S., Rieger-Christ, K., Latini, J.M., Moinzadeh, A., Stoffel, J., Pezza, J.A., Saini, K., Libertino, J.A., and Summerhayes, I.C. (2000) Molecular analysis of PTEN and MXI1 in primary bladder carcinoma. *International Journal of Cancer*, **88** (4), 620–625.

151 Kwabi-Addo, B., Giri, D., Schmidt, K., Podsypanina, K., Parsons, R., Greenberg, N., and Ittmann, M. (2001) Haploinsufficiency of the Pten tumor suppressor gene promotes prostate cancer progression. *Proceedings of the National Academy of Sciences of the United States of America*, **98** (20), 11563–11568.

152 Kwon, C.H., Zhao, D., Chen, J., Alcantara, S., Li, Y., Burns, D.K., Mason, R.P., Lee, E.Y., Wu, H., and Parada, L.F. (2008) Pten haploinsufficiency accelerates formation of high-grade astrocytomas. *Cancer Research*, **68** (9), 3286–3294.

153 Gildea, J.J., Herlevsen, M., Harding, M.A., Gulding, K.M., Moskaluk, C.A., Frierson, H.F., and Theodorescu, D. (2004) PTEN can inhibit *in vitro* organotypic and *in vivo* orthotopic invasion of human bladder cancer cells even in the absence of its lipid phosphatase activity. *Oncogene*, **23** (40), 6788–6797.

154 Tsuruta, H., Kishimoto, H., Sasaki, T., Horie, Y., Natsui, M., Shibata, Y., Hamada, K., Yajima, N., Kawahara, K., Sasaki, M., Tsuchiya, N., Enomoto, K., Mak, T.W., Nakano, T., Habuchi, T., and Suzuki, A. (2006) Hyperplasia and carcinomas in Pten-deficient mice and reduced PTEN protein in human bladder cancer patients. *Cancer Research*, **66** (17), 8389–8396.

155 Coombs, L.M., Pigott, D.A., Sweeney, E., Proctor, A.J., Eydmann, M.E., Parkinson, C., and Knowles, M.A. (1991) Amplification and over-expression of c-erbB-2 in transitional cell carcinoma of the urinary bladder. *British Journal of Cancer*, **63** (4), 601–608.

156 Sauter, G., Moch, H., Moore, D., Carroll, P., Kerschmann, R., Chew, K., Mihatsch, M.J., Gudat, F., and Waldman, F. (1993) Heterogeneity of erbB-2 gene amplification in bladder cancer. *Cancer Research*, **53** (10 Suppl.), 2199–2203.

157 Lonn, U., Lonn, S., Friberg, S., Nilsson, B., Silfversward, C., and Stenkvist, B. (1995) Prognostic value of amplification of c-erb-B2 in bladder carcinoma. *Clinical Cancer Research*, **1** (10), 1189–1194.

158 Miyamoto, H., Kubota, Y., Noguchi, S., Takase, K., Matsuzaki, J., Moriyama, M., Takebayashi, S., Kitamura, H., and Hosaka, M. (2000) C-ERBB-2 gene amplification as a prognostic marker in human bladder cancer. *Urology*, **55** (5), 679–683.

159 Simon, R., Atefy, R., Wagner, U., Forster, T., Fijan, A., Bruderer, J., Wilber, K., Mihatsch, M.J., Gasser, T., and Sauter, G. (2003) HER-2 and TOP2A coamplification in urinary bladder cancer. *International Journal of Cancer*, **107** (5), 764–772.

160 Hovey, R.M., Chu, L., Balazs, M., DeVries, S., Moore, D., Sauter, G., Carroll, P.R., and Waldman, F.M. (1998) Genetic alterations in primary bladder cancers and their metastases. *Cancer Research*, **58** (16), 3555–3560.

161 Simon, R., Burger, H., Semjonow, A., Hertle, L., Terpe, H.J., and Bocker, W. (2000) Patterns of chromosomal imbalances in muscle invasive bladder cancer. *International Journal of Oncology*, **17** (5), 1025–1029.

162 Veltman, J.A., Fridlyand, J., Pejavar, S., Olshen, A.B., Korkola, J.E., DeVries, S., Carroll, P., Kuo, W.L., Pinkel, D., Albertson, D., Cordon-Cardo, C., Jain, A.N., and Waldman, F.M. (2003) Array-based comparative genomic hybridization for genome-wide screening of DNA copy number in bladder tumors. *Cancer Research*, **63** (11), 2872–2880.

163 Simon, R., Eltze, E., Schafer, K.L., Burger, H., Semjonow, A., Hertle, L., Dockhorn-Dworniczak, B., Terpe, H.J., and Bocker, W. (2001) Cytogenetic analysis of multifocal bladder cancer supports a monoclonal origin and intraepithelial spread of tumor cells. *Cancer Research*, **61** (1), 355–362.

164 Schaffer, A.A., Simon, R., Desper, R., Richter, J., and Sauter, G. (2001) Tree models for dependent copy number changes in bladder cancer. *International Journal of Oncology*, **18** (2), 349–354.

165 Blaveri, E., Simko, J.P., Korkola, J.E., Brewer, J.L., Baehner, F., Mehta, K., Devries, S., Koppie, T., Pejavar, S., Carroll, P., and Waldman, F.M. (2005) Bladder cancer outcome and subtype classification by gene expression. *Clinical Cancer Research*, **11** (11), 4044–4055.

166 Tsai, Y.C., Nichols, P.W., Hiti, A.L., Williams, Z., Skinner, D.G., and Jones, P.A. (1990) Allelic losses of chromosomes 9, 11, and 17 in human bladder cancer. *Cancer Research*, **50** (1), 44–47.

167 Linnenbach, A.J., Pressler, L.B., Seng, B.A., Kimmel, B.S., Tomaszewski, J.E., and Malkowicz, S.B. (1993) Characterization of chromosome 9 deletions in transitional cell carcinoma by microsatellite assay. *Human Molecular Genetics*, **2** (9), 1407–1411.

168 Gibas, Z., Prout, G.R., Jr., Connolly, J.G., Pontes, J.E., and Sandberg, A.A. (1984) Nonrandom chromosomal changes in transitional cell carcinoma of the bladder. *Cancer Research*, **44** (3), 1257–1264.

169 Fadl-Elmula, I., Gorunova, L., Mandahl, N., Elfving, P., Lundgren, R., Mitelman, F., and Heim, S. (2000) Karyotypic characterization of urinary bladder transitional cell carcinomas. *Genes, Chromosomes & Cancer*, **29** (3), 256–265.

170 Cairns, P., Mao, L., Merlo, A., Lee, D.J., Schwab, D., Eby, Y., Tokino, K., van der Riet, P., Blaugrund, J.E., and Sidransky, D. (1994) Rates of p16 (MTS1) mutations in primary tumors with 9p loss. *Science*, **265** (5170), 415–417.

171 Devlin, J., Keen, A.J., and Knowles, M.A. (1994) Homozygous deletion mapping at 9p21 in bladder carcinoma defines a critical region within 2cM of IFNA. *Oncogene*, **9** (9), 2757–2760.

172 Orlow, I., Lacombe, L., Hannon, G.J., Serrano, M., Pellicer, I., Dalbagni, G., Reuter, V.E., Zhang, Z.F., Beach, D., and Cordon-Cardo, C. (1995) Deletion of the p16 and p15 genes in human bladder tumors. *Journal of the National Cancer Institute*, **87** (20), 1524–1529.

173 Williamson, M.P., Elder, P.A., Shaw, M.E., Devlin, J., and Knowles, M.A. (1995) p16 (CDKN2) is a major deletion target at 9p21 in bladder cancer. *Human Molecular Genetics*, **4** (9), 1569–1577.

174 Chapman, E.J., Harnden, P., Chambers, P., Johnston, C., and Knowles, M.A. (2005) Comprehensive analysis of CDKN2A status in microdissected urothelial cell carcinoma reveals potential haploinsufficiency, a high frequency of homozygous co-deletion and associations with clinical phenotype. *Clinical Cancer Research*, **11** (16), 5740–5747.

175 Bartoletti, R., Cai, T., Nesi, G., Roberta Girardi, L., Baroni, G., and Dal Canto, M. (2007) Loss of P16 expression and chromosome 9p21 LOH in predicting outcome of patients affected by superficial bladder cancer. *Journal of Surgical Research*, **143** (2), 422–427.

176 Berggren de Verdier, P.J., Kumar, R., Adolfsson, J., Larsson, P., Norming, U., Onelov, E., Wijkstrom, H., Steineck, G., and Hemminki, K. (2006) Prognostic significance of homozygous deletions and multiple duplications at the CDKN2A ($p16^{INK4a}$)/ARF ($p14^{ARF}$) locus in urinary bladder cancer. *Scandinavian Journal of Urology and Nephrology*, **40** (5), 363–369.

177 McGarvey, T.W., Maruta, Y., Tomaszewski, J.E., Linnenbach, A.J., and Malkowicz, S.B. (1998) PTCH gene mutations in invasive transitional cell carcinoma of the bladder. *Oncogene*, **17** (9), 1167–1172.

178 Aboulkassim, T.O., LaRue, H., Lemieux, P., Rousseau, F., and Fradet, Y. (2003) Alteration of the PATCHED locus in superficial bladder cancer. *Oncogene*, **22** (19), 2967–2971.

179 Habuchi, T., Luscombe, M., Elder, P.A., and Knowles, M.A. (1998) Structure and methylation-based silencing of a gene (DBCCR1) within a candidate bladder cancer tumor suppressor region at 9q32–q33. *Genomics*, **48** (3), 277–288.

180 Nishiyama, H., Hornigold, N., Davies, A.M., and Knowles, M.A. (1999) A sequence-ready 840-kb PAC contig spanning the candidate tumor suppressor locus DBC1 on human chromosome 9q32–q33. *Genomics*, **59** (3), 335–338.

181 Stadler, W.M., Steinberg, G., Yang, X., Hagos, F., Turner, C., and Olopade, O.I. (2001) Alterations of the 9p21 and 9q33 chromosomal bands in clinical bladder cancer specimens by fluorescence *in situ* hybridization. *Clinical Cancer Research*, **7** (6), 1676–1682.

182 Habuchi, T., Yoshida, O., and Knowles, M.A. (1997) A novel candidate tumour suppressor locus at 9q32–33 in bladder cancer: localization of the candidate region within a single 840kb YAC. *Human Molecular Genetics*, **6** (6), 913–919.

183 Nishiyama, H., Takahashi, T., Kakehi, Y., Habuchi, T., and Knowles, M.A. (1999) Homozygous deletion at the 9q32–33 candidate tumor suppressor locus in primary human bladder cancer. *Genes, Chromosomes & Cancer*, **26** (2), 171–175.

184 Fujiwara, H., Emi, M., Nagai, H., Ohgaki, K., Imoto, I., Akimoto, M., Ogawa, O., and Habuchi, T. (2001) Definition of a 1-Mb homozygous deletion at 9q32–q33 in a human bladder-cancer cell line. *Journal of Human Genetics*, **46** (7), 372–377.

185 Salem, C., Liang, G., Tsai, Y.C., Coulter, J., Knowles, M.A., Feng, A.C., Groshen, S., Nichols, P.W., and Jones, P.A. (2000) Progressive increases in de novo methylation of CpG islands in bladder cancer. *Cancer Research*, **60** (9), 2473–2476.

186 Habuchi, T., Takahashi, T., Kakinuma, H., Wang, L., Tsuchiya, N., Satoh, S., Akao, T., Sato, K., Ogawa, O., Knowles, M.A., and Kato, T. (2001) Hypermethylation at 9q32–33 tumour suppressor region is age-related in normal urothelium and an early and frequent alteration in bladder cancer. *Oncogene*, **20** (4), 531–537.

187 Nishiyama, H., Gill, J.H., Pitt, E., Kennedy, W., and Knowles, M.A. (2001) Negative regulation of G_1/S transition by the candidate bladder tumour suppressor gene DBCCR1. *Oncogene*, **20** (23), 2956–2964.

188 Hirao, S., Hirao, T., Marsit, C.J., Hirao, Y., Schned, A., Devi-Ashok, T., Nelson, H.H., Andrew, A., Karagas, M.R., and Kelsey, K.T. (2005) Loss of heterozygosity on chromosome 9q and p53 alterations in human bladder cancer. *Cancer*, **104** (9), 1918–1923.

189 Ferlay, J., Shin, H.R., Bray, F., Forman, D., Mathers, C., and Parkin, D.M. (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International Journal of Cancer*, **127** (12), 2893–2917.

190 Alvarez-Cubero, M.J., Saiz, M., Martinez-Gonzalez, L.J., Alvarez, J.C., Lorente, J.A., and Cozar, J.M. (2013) Genetic analysis of the principal genes related to prostate cancer: a review. *Urologic Oncology*, **31** (8), 1419–1429.

191 Lee, J., Demissie, K., Lu, S.E., and Rhoads, G.G. (2007) Cancer incidence among Korean-American immigrants in the United States and native Koreans in South Korea. *Cancer Control: Journal of the Moffitt Cancer Center*, **14** (1), 78–85.

192 Schulz, W.A. (2005) *Molecular Biology of Human Cancers: An Advanced Student's Textbook*, 1st edn, Springer, Dordrecht.

193 Choudhury, A.D., Eeles, R., Freedland, S. J., Isaacs, W.B., Pomerantz, M.M., Schalken, J.A., Tammela, T.L., and Visakorpi, T. (2012) The role of genetic markers in the management of prostate cancer. *European Urology*, **62** (4), 577–587.

194 Simard, J., Dumont, M., Labuda, D., Sinnett, D., Meloche, C., El-Alfy, M., Berger, L., Lees, E., Labrie, F., and Tavtigian, S.V. (2003) Prostate cancer susceptibility genes: lessons learned and challenges posed. *Endocrine-Related Cancer*, **10** (2), 225–259.

195 Maier, C., Herkommer, K., Hoegel, J., Vogel, W., and Paiss, T. (2005) A genomewide linkage analysis for prostate cancer susceptibility genes in families from Germany. *European Journal of Human Genetics*, **13** (3), 352–360.

196 Castro, E. and Eeles, R. (2012) The role of BRCA1 and BRCA2 in prostate cancer. *Asian Journal of Andrology*, **14** (3), 409–414.

197 Eeles, R.A., Al Olama, A.A., Benlloch, S., Saunders, E.J., Leongamornlert, D.A., Tymrakiewicz, M., Ghoussaini, M., Luccarini, C., Dennis, J., Jugurnauth-Little, S., Dadaev, T., Neal, D.E., Hamdy, F.C., Donovan, J.L., Muir, K., Giles, G.G.,

Severi, G., Wiklund, F., Gronberg, H., Haiman, C.A., Schumacher, F., Henderson, B.E., Le Marchand, L., Lindstrom, S., Kraft, P., Hunter, D.J., Gapstur, S., Chanock, S.J., Berndt, S.I., Albanes, D., Andriole, G., Schleutker, J., Weischer, M., Canzian, F., Riboli, E., Key, T.J., Travis, R.C., Campa, D., Ingles, S.A., John, E.M., Hayes, R.B., Pharoah, P.D., Pashayan, N., Khaw, K.T., Stanford, J.L., Ostrander, E.A., Signorello, L.B., Thibodeau, S.N., Schaid, D., Maier, C., Vogel, W., Kibel, A.S., Cybulski, C., Lubinski, J., Cannon-Albright, L., Brenner, H., Park, J.Y., Kaneva, R., Batra, J., Spurdle, A.B., Clements, J.A., Teixeira, M.R., Dicks, E., Lee, A., Dunning, A.M., Baynes, C., Conroy, D., Maranian, M.J., Ahmed, S., Govindasami, K., Guy, M., Wilkinson, R.A., Sawyer, E.J., Morgan, A., Dearnaley, D.P., Horwich, A., Huddart, R.A., Khoo, V.S., Parker, C.C., Van As, N.J., Woodhouse, C.J., Thompson, A., Dudderidge, T., Ogden, C., Cooper, C.S., Lophatananon, A., Cox, A., Southey, M.C., Hopper, J.L., English, D.R., Aly, M., Adolfsson, J., Xu, J., Zheng, S.L., Yeager, M., Kaaks, R., Diver, W.R., Gaudet, M.M., Stern, M.C., Corral, R., Joshi, A.D., Shahabi, A., Wahlfors, T., Tammela, T.L., Auvinen, A., Virtamo, J., Klarskov, P., Nordestgaard, B.G., Roder, M.A., Nielsen, S.F., Bojesen, S.E., Siddiq, A., Fitzgerald, L.M., Kolb, S., Kwon, E.M., Karyadi, D.M., Blot, W.J., Zheng, W., Cai, Q., McDonnell, S.K., Rinckleb, A.E., Drake, B., Colditz, G., Wokolorczyk, D., Stephenson, R.A., Teerlink, C., Muller, H., Rothenbacher, D., Sellers, T.A., Lin, H.Y., Slavov, C., Mitev, V., Lose, F., Srinivasan, S., Maia, S., Paulo, P., Lange, E., Cooney, K.A., Antoniou, A.C., Vincent, D., Bacot, F., Tessier, D.C., COGS – Cancer Research UK GWAS – ELLIPSE (part of GAME-ON) Initiative, Australian Prostate Cancer Bioresource, UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology, UK ProtecT (Prostate testing for cancer and Treatment) Study Collaborators, and PRACTICAL (Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome) Consortium, Kote-Jarai, Z., and Easton F D.F. (2013) Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nature Genetics*, **45** (4), 385–391, 391e381 382.

198 Olama, A.A., Kote-Jarai, Z., Schumacher, F.R., Wiklund, F., Berndt, S.I., Benlloch, S., Giles, G.G., Severi, G., Neal, D.E., Hamdy, F.C., Donovan, J.L., Hunter, D.J., Henderson, B.E., Thun, M.J., Gaziano, M., Giovannucci, E.L., Siddiq, A., Travis, R.C., Cox, D.G., Canzian, F., Riboli, E., Key, T.J., Andriole, G., Albanes, D., Hayes, R.B., Schleutker, J., Auvinen, A., Tammela, T.L., Weischer, M., Stanford, J.L., Ostrander, E.A., Cybulski, C., Lubinski, J., Thibodeau, S.N., Schaid, D.J., Sorensen, K.D., Batra, J., Clements, J.A., Chambers, S., Aitken, J., Gardiner, R.A., Maier, C., Vogel, W., Dork, T., Brenner, H., Habuchi, T., Ingles, S., John, E.M., Dickinson, J.L., Cannon-Albright, L., Teixeira, M.R., Kaneva, R., Zhang, H.W., Lu, Y.J., Park, J.Y., Cooney, K.A., Muir, K.R., Leongamornlert, D.A., Saunders, E., Tymrakiewicz, M., Mahmud, N., Guy, M., Govindasami, K., O'Brien, L.T., Wilkinson, R.A., Hall, A.L., Sawyer, E.J., Dadaev, T., Morrison, J., Dearnaley, D.P., Horwich, A., Huddart, R.A., Khoo, V.S., Parker, C.C., Van As, N., Woodhouse, C.J., Thompson, A., Dudderidge, T., Ogden, C., Cooper, C.S., Lophatonanon, A., Southey, M.C., Hopper, J.L., English, D., Virtamo, J., Le Marchand, L., Campa, D., Kaaks, R., Lindstrom, S., Diver, W.R., Gapstur, S., Yeager, M., Cox, A., Stern, M.C., Corral, R., Aly, M., Isaacs, W., Adolfsson, J., Xu, J., Zheng, S.L., Wahlfors, T., Taari, K., Kujala, P., Klarskov, P., Nordestgaard, B.G., Roder, M.A., Frikke-Schmidt, R., Bojesen, S.E., FitzGerald, L.M., Kolb, S., Kwon, E.M., Karyadi, D.M., Orntoft, T.F., Borre, M., Rinckleb, A., Luedeke, M., Herkommer, K., Meyer, A., Serth, J., Marthick, J.R., Patterson, B., Wokolorczyk, D., Spurdle, A., Lose, F., McDonnell, S.K., Joshi, A.D., Shahabi, A., Pinto, P., Santos, J., Ray, A., Sellers, T.A., Lin, H.Y., Stephenson, R.A., Teerlink, C., Muller, H., Rothenbacher,

D., Tsuchiya, N., Narita, S., Cao, G.W., Slavov, C., Mitev, V., UK Genetic Prostate Cancer Study Collaborators/ British Association of Urological Surgeons' Section of Oncology, UK ProtecT Study Collaborators, Australian Prostate Cancer Bioresource, PRACTICAL Consortium, Chanock, S., Gronberg, H., Haiman, C.A., Kraft F P. Easton F D.F., and Eeles F R.A. (2013) A meta-analysis of genome-wide association studies to identify prostate cancer susceptibility loci associated with aggressive and non-aggressive disease. *Human Molecular Genetics*, **22** (2), 408–415.

199 Castro, E., Goh, C., Olmos, D., Saunders, E., Leongamornlert, D., Tymrakiewicz, M., Mahmud, N., Dadaev, T., Govindasami, K., Guy, M., Sawyer, E., Wilkinson, R., Ardern-Jones, A., Ellis, S., Frost, D., Peock, S., Evans, D.G., Tischkowitz, M., Cole, T., Davidson, R., Eccles, D., Brewer, C., Douglas, F., Porteous, M.E., Donaldson, A., Dorkins, H., Izatt, L., Cook, J., Hodgson, S., Kennedy, M.J., Side, L.E., Eason, J., Murray, A., Antoniou, A.C., Easton, D.F., Kote-Jarai, Z., and Eeles, R. (2013) Germline BRCA mutations are associated with higher risk of nodal involvement, distant metastasis, and poor survival outcomes in prostate cancer. *Journal of Clinical Oncology*, **31** (14), 1748–1757.

200 Ewing, C.M., Ray, A.M., Lange, E.M., Zuhlke, K.A., Robbins, C.M., Tembe, W. D., Wiley, K.E., Isaacs, S.D., Johng, D., Wang, Y., Bizon, C., Yan, G., Gielzak, M., Partin, A.W., Shanmugam, V., Izatt, T., Sinari, S., Craig, D.W., Zheng, S.L., Walsh, P.C., Montie, J.E., Xu, J., Carpten, J.D., Isaacs, W.B., and Cooney, K.A. (2012) Germline mutations in HOXB13 and prostate-cancer risk. *New England Journal of Medicine*, **366** (2), 141–149.

201 Economides, K.D. and Capecchi, M.R. (2003) Hoxb13 is required for normal differentiation and secretory function of the ventral prostate. *Development*, **130** (10), 2061–2069.

202 Jung, C., Kim, R.S., Zhang, H.J., Lee, S.J., and Jeng, M.H. (2004) HOXB13 induces growth suppression of prostate cancer cells as a repressor of hormone-activated androgen receptor signaling. *Cancer Research*, **64** (24), 9185–9192.

203 Norris, J.D., Chang, C.Y., Wittmann, B. M., Kunder, R.S., Cui, H., Fan, D., Joseph, J.D., and McDonnell, D.P. (2009) The homeodomain protein HOXB13 regulates the cellular response to androgens. *Molecular Cell*, **36** (3), 405–416.

204 Vesprini, D., Liu, S., and Nam, R. (2013) Predicting high risk disease using serum and DNA biomarkers. *Current Opinion in Urology*, **23** (3), 252–260.

205 Carpten, J., Nupponen, N., Isaacs, S., Sood, R., Robbins, C., Xu, J., Faruque, M., Moses, T., Ewing, C., Gillanders, E., Hu, P., Bujnovszky, P., Makalowska, I., Baffoe-Bonnie, A., Faith, D., Smith, J., Stephan, D., Wiley, K., Brownstein, M., Gildea, D., Kelly, B., Jenkins, R., Hostetter, G., Matikainen, M., Schleutker, J., Klinger, K., Connors, T., Xiang, Y., Wang, Z., De Marzo, A., Papadopoulos, N., Kallioniemi, O.P., Burk, R., Meyers, D., Gronberg, H., Meltzer, P., Silverman, R., Bailey-Wilson, J., Walsh, P., Isaacs, W., and Trent, J. (2002) Germline mutations in the ribonuclease L gene in families showing linkage with HPC1. *Nature Genetics*, **30** (2), 181–184.

206 Emi, M., Asaoka, H., Matsumoto, A., Itakura, H., Kurihara, Y., Wada, Y., Kanamori, H., Yazaki, Y., Takahashi, E., Lepert, M. *et al.* (1993) Structure, organization, and chromosomal mapping of the human macrophage scavenger receptor gene. *Journal of Biological Chemistry*, **268** (3), 2120–2125.

207 Niegemann, E. and Brendel, M. (1994) A single amino acid change in SNM1-encoded protein leads to thermoconditional deficiency for DNA cross-link repair in *Saccharomyces cerevisiae*. *Mutation Research*, **315** (3), 275–279.

208 Seimon, T.A., Obstfeld, A., Moore, K.J., Golenbock, D.T., and Tabas, I. (2006) Combinatorial pattern recognition receptor signaling alters the balance of life and death in macrophages. *Proceedings of the National Academy of*

Sciences of the United States of America, **103** (52), 19794–19799.

209 Korver, W., Guevara, C., Chen, Y., Neuteboom, S., Bookstein, R., Tavtigian, S., and Lees, E. (2003) The product of the candidate prostate cancer susceptibility gene ELAC2 interacts with the gamma-tubulin complex. *International Journal of Cancer*, **104** (3), 283–288.

210 Beuten, J., Gelfond, J.A., Franke, J.L., Shook, S., Johnson-Pais, T.L., Thompson, I.M., and Leach, R.J. (2010) Single and multivariate associations of MSR1, ELAC2, and RNASEL with prostate cancer in an ethnic diverse cohort of men. *Cancer Epidemiology, Biomarkers & Prevention*, **19** (2), 588–599.

211 Bensen, J.T., Xu, Z., Smith, G.J., Mohler, J.L., Fontham, E.T., and Taylor, J.A. (2013) Genetic polymorphism and prostate cancer aggressiveness: a case-only study of 1,536 GWAS and candidate SNPs in African-Americans and European-Americans. *Prostate*, **73** (1), 11–22.

212 Landi, D., Gemignani, F., and Landi, S. (2012) Role of variations within microRNA-binding sites in cancer. *Mutagenesis*, **27** (2), 205–210.

213 He, B., Pan, Y., Cho, W.C., Xu, Y., Gu, L., Nie, Z., Chen, L., Song, G., Gao, T., Li, R., and Wang, S. (2012) The association between four genetic variants in microRNAs (rs11614913, rs2910164, rs3746444, rs2292832) and cancer risk: evidence from published studies. *PLoS One*, **7** (11), e49032.

214 Kachroo, N. and Gnanapragasam, V.J. (2013) The role of treatment modality on the utility of predictive tissue biomarkers in clinical prostate cancer: a systematic review. *Journal of Cancer Research and Clinical Oncology*, **139** (1), 1–24.

215 Stephan, C., Cammann, H., Meyer, H.A., Lein, M., and Jung, K. (2007) PSA and new biomarkers within multivariate models to improve early detection of prostate cancer. *Cancer Letters*, **249** (1), 18–29.

216 Hoffman, R.M. (2011) Clinical practice. Screening for prostate cancer. *New England Journal of Medicine*, **365** (21), 2013–2019.

217 Djavan, B., Remzi, M., Zlotta, A.R., Ravery, V., Hammerer, P., Reissigl, A., Dobronski, P., Kaisary, A., and Marberger, M. (2002) Complexed prostate-specific antigen, complexed prostate-specific antigen density of total and transition zone, complexed/total prostate-specific antigen ratio, free-to-total prostate-specific antigen ratio, density of total and transition zone prostate-specific antigen: results of the prospective multicenter European trial. *Urology*, **60** (4 Suppl. 1), 4–9.

218 Schroder, F.H. (2012) Landmarks in prostate cancer screening. *BJU International*, (110 Suppl.), 13–17.

219 Schroder, F.H., Hugosson, J., Roobol, M.J., Tammela, T.L., Ciatto, S., Nelen, V., Kwiatkowski, M., Lujan, M., Lilja, H., Zappa, M., Denis, L.J., Recker, F., Paez, A., Maattanen, L., Bangma, C.H., Aus, G., Carlsson, S., Villers, A., Rebillard, X., van der Kwast, T., Kujala, P.M., Blijenberg, B.G., Stenman, U.H., Huber, A., Taari, K., Hakama, M., Moss, S.M., de Koning, H.J., Auvinen, A., and Investigators, E. (2012) Prostate-cancer mortality at 11 years of follow-up. *New England Journal of Medicine*, **366** (11), 981–990.

220 Wever, E.M., Heijnsdijk, E.A., Draisma, G., Bangma, C.H., Roobol, M.J., Schroder, F.H., and de Koning, H.J. (2013) Treatment of local-regional prostate cancer detected by PSA screening: benefits and harms according to prognostic factors. *British Journal of Cancer*, **108** (10), 1971–1977.

221 Ivanovic, V., Melman, A., Davis-Joseph, B., Valcic, M., and Geliebter, J. (1995) Elevated plasma levels of TGF-beta 1 in patients with invasive prostate cancer. *Nature Medicine*, **1** (4), 282–284.

222 Miyake, H., Hara, I., Yamanaka, K., Gohji, K., Arakawa, S., and Kamidono, S. (1999) Elevation of serum levels of urokinase-type plasminogen activator and its receptor is associated with disease progression and prognosis in patients with prostate cancer. *Prostate*, **39** (2), 123–129.

223 Nakashima, J., Tachibana, M., Horiguchi, Y., Oya, M., Ohigashi, T., Asakura, H., and Murai, M. (2000) Serum interleukin

6 as a prognostic factor in patients with prostate cancer. *Clinical Cancer Research*, **6** (7), 2702–2706.

224 Haese, A., Graefen, M., Becker, C., Noldus, J., Katz, J., Cagiannos, I., Kattan, M., Scardino, P.T., Huland, E., Huland, H., and Lilja, H. (2003) The role of human glandular kallikrein 2 for prediction of pathologically organ confined prostate cancer. *Prostate*, **54** (3), 181–186.

225 Michalaki, V., Syrigos, K., Charles, P., and Waxman, J. (2004) Serum levels of IL-6 and TNF-alpha correlate with clinicopathological features and patient survival in patients with prostate cancer. *British Journal of Cancer*, **90** (12), 2312–2316.

226 Shariat, S.F., Kattan, M.W., Traxel, E., Andrews, B., Zhu, K., Wheeler, T.M., and Slawin, K.M. (2004) Association of pre- and postoperative plasma levels of transforming growth factor $beta_1$ and interleukin 6 and its soluble receptor with prostate cancer progression. *Clinical Cancer Research*, **10** (6), 1992–1999.

227 Shariat, S.F., Roehrborn, C.G., McConnell, J.D., Park, S., Alam, N., Wheeler, T.M., and Slawin, K.M. (2007) Association of the circulating levels of the urokinase system of plasminogen activation with the presence of prostate cancer and invasion, progression, and metastasis. *Journal of Clinical Oncology*, **25** (4), 349–355.

228 Steuber, T., Vickers, A.J., Serio, A.M., Vaisanen, V., Haese, A., Pettersson, K., Eastham, J.A., Scardino, P.T., Huland, H., and Lilja, H. (2007) Comparison of free and total forms of serum human kallikrein 2 and prostate-specific antigen for prediction of locally advanced and recurrent prostate cancer. *Clinical Chemistry*, **53** (2), 233–240.

229 Shariat, S.F., Karam, J.A., Walz, J., Roehrborn, C.G., Montorsi, F., Margulis, V., Saad, F., Slawin, K.M., and Karakiewicz, P.I. (2008) Improved prediction of disease relapse after radical prostatectomy through a panel of preoperative blood-based biomarkers. *Clinical Cancer Research*, **14** (12), 3785–3791.

230 Milanese, G., Dellabella, M., Fazioli, F., Pierpaoli, E., Polito, M., Siednius, N., Montironi, R., Blasi, F., and Muzzonigro, G. (2009) Increased urokinase-type plasminogen activator receptor and epidermal growth factor receptor in serum of patients with prostate cancer. *Journal of Urology*, **181** (3), 1393–1400.

231 Nagaraj, N.S. and Datta, P.K. (2010) Targeting the transforming growth factor-beta signaling pathway in human cancer. *Expert Opinion on Investigational Drugs*, **19** (1), 77–91.

232 Almasi, C.E., Brasso, K., Iversen, P., Pappot, H., Hoyer-Hansen, G., Dano, K., and Christensen, I.J. (2011) Prognostic and predictive value of intact and cleaved forms of the urokinase plasminogen activator receptor in metastatic prostate cancer. *Prostate*, **71** (8), 899–907.

233 Terracciano, D., Bruzzese, D., Ferro, M., Autorino, R., di Lorenzo, G., Buonerba, C., Mariano, A., Macchia, V., Altieri, V., and di Carlo, A. (2011) Soluble interleukin-6 receptor to interleukin-6 (sIL6R/IL-6) ratio in serum as a predictor of high Gleason sum at radical prostatectomy. *Oncology Letters*, **2** (5), 861–864.

234 Grubisha, M.J. and DeFranco, D.B. (2013) Local endocrine, paracrine and redox signaling networks impact estrogen and androgen crosstalk in the prostate cancer microenvironment. *Steroids*, **78** (6), 538–541.

235 Duffy, M.J. and Duggan, C. (2004) The urokinase plasminogen activator system: a rich source of tumour markers for the individualised management of patients with cancer. *Clinical Biochemistry*, **37** (7), 541–548.

236 Guo, Y., Xu, F., Lu, T., Duan, Z., and Zhang, Z. (2012) Interleukin-6 signaling pathway in targeted therapy for cancer. *Cancer Treatment Reviews*, **38** (7), 904–910.

237 Vickers, A.J., Cronin, A.M., Roobol, M.J., Savage, C.J., Peltola, M., Pettersson, K., Scardino, P.T., Schroder, F.H., and Lilja, H. (2010) A four-kallikrein panel predicts prostate cancer in men with recent screening: data from the European Randomized Study of Screening for

Prostate Cancer, Rotterdam. *Clinical Cancer Research*, **16** (12), 3232–3239.

238 Liu, N., Liang, W., Ma, X., Li, X., Ning, B., Cheng, C., Ou, G., Wang, B., Zhang, J., and Gao, Z. (2013) Simultaneous and combined detection of multiple tumor biomarkers for prostate cancer in human serum by suspension array technology. *Biosensors & Bioelectronics*, **47**, 92–98.

239 Hermans, K.G., Boormans, J.L., Gasi, D., van Leenders, G.J., Jenster, G., Verhagen, P.C., and Trapman, J. (2009) Overexpression of prostate-specific TMPRSS2(exon 0)–ERG fusion transcripts corresponds with favorable prognosis of prostate cancer. *Clinical Cancer Research*, **15** (20), 6398–6403.

240 Tomlins, S.A., Aubin, S.M., Siddiqui, J., Lonigro, R.J., Sefton-Miller, L., Miick, S., Williamsen, S., Hodge, P., Meinke, J., Blase, A., Penabella, Y., Day, J.R., Varambally, R., Han, B., Wood, D., Wang, L., Sanda, M.G., Rubin, M.A., Rhodes, D. R., Hollenbeck, B., Sakamoto, K., Silberstein, J.L., Fradet, Y., Amberson, J.B., Meyers, S., Palanisamy, N., Rittenhouse, H., Wei, J.T., Groskopf, J., and Chinnaiyan, A.M. (2011) Urine TMPRSS2–ERG fusion transcript stratifies prostate cancer risk in men with elevated serum PSA. *Science Translational Medicine*, **3** (94), 94ra72.

241 Hessels, D., Smit, F.P., Verhaegh, G.W., Witjes, J.A., Cornel, E.B., and Schalken, J. A. (2007) Detection of TMPRSS2–ERG fusion transcripts and prostate cancer antigen 3 in urinary sediments may improve diagnosis of prostate cancer. *Clinical Cancer Research*, **13** (17), 5103–5108.

242 Rice, K.R., Chen, Y., Ali, A., Whitman, E.J., Blase, A., Ibrahim, M., Elsamanoudi, S., Brassell, S., Furusato, B., Stingle, N., Sesterhenn, I.A., Petrovics, G., Miick, S., Rittenhouse, H., Groskopf, J., McLeod, D.G., and Srivastava, S. (2010) Evaluation of the ETS-related gene mRNA in urine for the detection of prostate cancer. *Clinical Cancer Research*, **16** (5), 1572–1576.

243 Zhou, A.G., Owens, C.L., Cosar, E.F., and Jiang, Z. (2013) Clinical implications of current developments in genitourinary pathology. *Archives of Pathology & Laboratory Medicine*, **137** (7), 887–893.

244 Cooper, C.S., Campbell, C., and Jhavar, S. (2007) Mechanisms of Disease: biomarkers and molecular targets from microarray gene expression studies in prostate cancer. *Nature Clinical Practice Urology*, **4** (12), 677–687.

245 Irshad, S., Bansal, M., Castillo-Martin, M., Zheng, T., Aytes, A., Wenske, S., Le Magnen, C., Guarnieri, P., Sumazin, P., Benson, M.C., Shen, M.M., Califano, A., and Abate-Shen, C. (2013) A molecular signature predictive of indolent prostate cancer. *Science Translational Medicine*, **5** (202), 202ra122.

246 Mattick, J.S. (2001) Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Reports*, **2** (11), 986–991.

247 Huttenhofer, A., Schattner, P., and Polacek, N. (2005) Non-coding RNAs: hope or hype? *Trends in Genetics*, **21** (5), 289–297.

248 Petrovics, G., Zhang, W., Makarem, M., Street, J.P., Connelly, R., Sun, L., Sesterhenn, I.A., Srikantan, V., Moul, J.W., and Srivastava, S. (2004) Elevated expression of PCGEM1, a prostate-specific gene with cell growth-promoting function, is associated with high-risk prostate cancer patients. *Oncogene*, **23** (2), 605–611.

249 Chung, S., Nakagawa, H., Uemura, M., Piao, L., Ashikawa, K., Hosono, N., Takata, R., Akamatsu, S., Kawaguchi, T., Morizono, T., Tsunoda, T., Daigo, Y., Matsuda, K., Kamatani, N., Nakamura, Y., and Kubo, M. (2011) Association of a novel long non-coding RNA in 8q24 with prostate cancer susceptibility. *Cancer Science*, **102** (1), 245–252.

250 Prensner, J.R., Iyer, M.K., Balbin, O.A., Dhanasekaran, S.M., Cao, Q., Brenner, J.C., Laxman, B., Asangani, I.A., Grasso, C.S., Kominsky, H.D., Cao, X., Jing, X., Wang, X., Siddiqui, J., Wei, J.T., Robinson, D., Iyer, H.K., Palanisamy, N., Maher, C.A., and Chinnaiyan, A.M. (2011) Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nature Biotechnology*, **29** (8), 742–749.

251 Ferreira, L.B., Palumbo, A., de Mello, K.D., Sternberg, C., Caetano, M.S., de Oliveira, F.L., Neves, A.F., Nasciutti, L.E., Goulart, L.R., and Gimba, E.R. (2012) PCA3 noncoding RNA is involved in the control of prostate-cancer cell survival and modulates androgen receptor signaling. *BMC Cancer*, **12**, 507.

252 Bussemakers, M.J., van Bokhoven, A., Verhaegh, G.W., Smit, F.P., Karthaus, H.F., Schalken, J.A., Debruyne, F.M., Ru, N., and Isaacs, W.B. (1999) DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. *Cancer Research*, **59** (23), 5975–5979.

253 US Food and Drug Administration (2012) Medcial Devices: PROGENSA PCA3 assay, http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/DeviceApprovalsandClearances/Recently-ApprovedDevices/ucm294907.htm (2 July 2013).

254 Bradley, L.A., Palomaki, G.E., Gutman, S., Samson, D., and Aronson, N. (2013) Comparative effectiveness review: prostate cancer antigen 3 testing for the diagnosis and management of prostate cancer. *Journal of Urology*, **190** (2), 389–398.

255 Sung, Y.Y. and Cheung, E. (2013) Antisense now makes sense: dual modulation of androgen-dependent transcription by CTBP1-AS. *EMBO Journal*, **32** (12), 1653–1654.

256 Takayama, K., Horie-Inoue, K., Katayama, S., Suzuki, T., Tsutsumi, S., Ikeda, K., Urano, T., Fujimura, T., Takagi, K., Takahashi, S., Homma, Y., Ouchi, Y., Aburatani, H., Hayashizaki, Y., and Inoue, S. (2013) Androgen-responsive long noncoding RNA CTBP1-AS promotes prostate cancer. *EMBO Journal*, **32** (12), 1665–1680.

257 Kong, Y.W., Ferland-McCollough, D., Jackson, T.J., and Bushell, M. (2012) microRNAs in cancer management. *Lancet Oncology*, **13** (6), e249–e258.

258 Esquela-Kerscher, A. and Slack, F.J. (2006) Oncomirs – microRNAs with a role in cancer. *Nature Reviews Cancer*, **6** (4), 259–269.

259 Gandellini, P., Folini, M., and Zaffaroni, N. (2009) Towards the definition of prostate cancer-related microRNAs: where are we now? *Trends in Molecular Medicine*, **15** (9), 381–390.

260 Catto, J.W., Alcaraz, A., Bjartell, A.S., De Vere White, R., Evans, C.P., Fussel, S., Hamdy, F.C., Kallioniemi, O., Mengual, L., Schlomm, T., and Visakorpi, T. (2011) MicroRNA in prostate, bladder, and kidney cancer: a systematic review. *European Urology*, **59** (5), 671–681.

261 Maugeri-Sacca, M., Coppola, V., Bonci, D., and De Maria, R. (2012) MicroRNAs and prostate cancer: from preclinical research to translational oncology. *Cancer Journal*, **18** (3), 253–261.

262 O'Kelly, F., Marignol, L., Meunier, A., Lynch, T.H., Perry, A.S., and Hollywood, D. (2012) MicroRNAs as putative mediators of treatment response in prostate cancer. *Nature Reviews Urology*, **9** (7), 397–407.

263 Fang, Y.X. and Gao, W.Q. (2013) Roles of microRNAs during prostatic tumorigenesis and tumor progression. *Oncogene*, **33** (2), 135–147.

264 Cortez, M.A., Bueso-Ramos, C., Ferdin, J., Lopez-Berestein, G., Sood, A.K., and Calin, G.A. (2011) MicroRNAs in body fluids–the mix of hormones and biomarkers. *Nature Reviews Clinical Oncology*, **8** (8), 467–477.

265 Sita-Lumsden, A., Dart, D.A., Waxman, J., and Bevan, C.L. (2013) Circulating microRNAs as potential new biomarkers for prostate cancer. *British Journal of Cancer*, **108** (10), 1925–1930.

266 Mitchell, P.S., Parkin, R.K., Kroh, E.M., Fritz, B.R., Wyman, S.K., Pogosova-Agadjanyan, E.L., Peterson, A., Noteboom, J., O'Briant, K.C., Allen, A., Lin, D.W., Urban, N., Drescher, C.W., Knudsen, B.S., Stirewalt, D.L., Gentleman, R., Vessella, R.L., Nelson, P.S., Martin, D.B., and Tewari, M. (2008) Circulating microRNAs as stable blood-based markers for cancer detection. *Proceedings of the National Academy of Sciences of the United States of America*, **105** (30), 10513–10518.

267 Brase, J.C., Johannes, M., Schlomm, T., Falth, M., Haese, A., Steuber, T.,

Beissbarth, T., Kuner, R., and Sultmann, H. (2011) Circulating miRNAs are correlated with tumor progression in prostate cancer. *International Journal of Cancer*, **128** (3), 608–616.

268 Bryant, R.J., Pawlowski, T., Catto, J.W., Marsden, G., Vessella, R.L., Rhees, B., Kuslich, C., Visakorpi, T., and Hamdy, F.C. (2012) Changes in circulating microRNA levels associated with prostate cancer. *British Journal of Cancer*, **106** (4), 768–774.

269 Selth, L.A., Townley, S., Gillis, J.L., Ochnik, A.M., Murti, K., Macfarlane, R.J., Chi, K.N., Marshall, V.R., Tilley, W.D., and Butler, L.M. (2012) Discovery of circulating microRNAs associated with human prostate cancer using a mouse model of disease. *International Journal of Cancer*, **131** (3), 652–661.

270 Madhavan, D., Cuk, K., Burwinkel, B., and Yang, R. (2013) Cancer diagnosis and prognosis decoded by blood-based circulating microRNA signatures. *Frontiers in Genetics*, **4**, 116.

271 Nguyen, H.C., Xie, W., Yang, M., Hsieh, C.L., Drouin, S., Lee, G.S., and Kantoff, P.W. (2013) Expression differences of circulating microRNAs in metastatic castration resistant prostate cancer and low-risk, localized prostate cancer. *Prostate*, **73** (4), 346–354.

272 Yaman Agaoglu, F., Kovancilar, M., Dizdar, Y., Darendeliler, E., Holdenrieder, S., Dalay, N., and Gezer, U. (2011) Investigation of miR-21, miR-141, and miR-221 in blood circulation of patients with prostate cancer. *Tumour Biology*, **32** (3), 583–588.

273 Moltzahn, F., Olshen, A.B., Baehner, L., Peek, A., Fong, L., Stoppler, H., Simko, J., Hilton, J.F., Carroll, P., and Blelloch, R. (2011) Microfluidic-based multiplex qRT-PCR identifies diagnostic and prognostic microRNA signatures in the sera of prostate cancer patients. *Cancer Research*, **71** (2), 550–560.

274 Shen, J., Hruby, G.W., McKiernan, J.M., Gurvich, I., Lipsky, M.J., Benson, M.C., and Santella, R.M. (2012) Dysregulation of circulating microRNAs and prediction of aggressive prostate cancer. *Prostate*, **72** (13), 1469–1477.

275 Zhang, H.L., Yang, L.F., Zhu, Y., Yao, X.D., Zhang, S.L., Dai, B., Zhu, Y.P., Shen, Y.J., Shi, G.H., and Ye, D.W. (2011) Serum miRNA-21: elevated levels in patients with metastatic hormone-refractory prostate cancer and potential predictive factor for the efficacy of docetaxel-based chemotherapy. *Prostate*, **71** (3), 326–331.

276 Gonzales, J.C., Fink, L.M., Goodman, O.B. Jr., Symanowski, J.T., Vogelzang, N.J., and Ward, D.C. (2011) Comparison of circulating MicroRNA 141 to circulating tumor cells, lactate dehydrogenase, and prostate-specific antigen for determining treatment response in patients with metastatic prostate cancer. *Clinical Genitourinary Cancer*, **9** (1), 39–45.

277 Keller, A., Leidinger, P., Bauer, A., Elsharawy, A., Haas, J., Backes, C., Wendschlag, A., Giese, N., Tjaden, C., Ott, K., Werner, J., Hackert, T., Ruprecht, K., Huwer, H., Huebers, J., Jacobs, G., Rosenstiel, P., Dommisch, H., Schaefer, A., Muller-Quernheim, J., Wullich, B., Keck, B., Graf, N., Reichrath, J., Vogel, B., Nebel, A., Jager, S.U., Staehler, P., Amarantos, I., Boisguerin, V., Staehler, C., Beier, M., Scheffler, M., Buchler, M.W., Wischhusen, J., Haeusler, S.F., Dietl, J., Hofmann, S., Lenhof, H.P., Schreiber, S., Katus, H.A., Rottbauer, W., Meder, B., Hoheisel, J.D., Franke, A., and Meese, E. (2011) Toward the blood-borne miRNome of human diseases. *Nature Methods*, **8** (10), 841–843.

278 Tong, A.W., Fulgham, P., Jay, C., Chen, P., Khalil, I., Liu, S., Senzer, N., Eklund, A.C., Han, J., and Nemunaitis, J. (2009) MicroRNA profile analysis of human prostate cancers. *Cancer Gene Therapy*, **16** (3), 206–216.

279 Schaefer, A., Jung, M., Mollenkopf, H.J., Wagner, I., Stephan, C., Jentzmik, F., Miller, K., Lein, M., Kristiansen, G., and Jung, K. (2010) Diagnostic and prognostic implications of microRNA profiling in prostate carcinoma. *International Journal of Cancer*, **126** (5), 1166–1176.

280 Wach, S., Nolte, E., Szczyrba, J., Stohr, R., Hartmann, A., Orntoft, T., Dyrskjot, L., Eltze, E., Wieland, W., Keck, B., Ekici, A.B., Grasser, F., and Wullich, B. (2012)

MicroRNA profiles of prostate carcinoma detected by multiplatform microRNA screening. *International Journal of Cancer*, **130** (3), 611–621.

281 Spahn, M., Kneitz, S., Scholz, C.J., Stenger, N., Rudiger, T., Strobel, P., Riedmiller, H., and Kneitz, B. (2010) Expression of microRNA-221 is progressively reduced in aggressive prostate cancer and metastasis and predicts clinical recurrence. *International Journal of Cancer*, **127** (2), 394–403.

282 Salmena, L., Carracedo, A., and Pandolfi, P.P. (2008) Tenets of PTEN tumor suppression. *Cell*, **133** (3), 403–414.

283 Carracedo, A., Alimonti, A., and Pandolfi, P.P. (2011) PTEN level in tumor suppression: how much is too little? *Cancer Research*, **71** (3), 629–633.

284 Li, Y., Su, J., DingZhang, X., Zhang, J., Yoshimoto, M., Liu, S., Bijian, K., Gupta, A., Squire, J.A., Alaoui Jamali, M.A., and Bismar, T.A. (2011) PTEN deletion and heme oxygenase-1 overexpression cooperate in prostate cancer progression and are associated with adverse clinical outcome. *Journal of Pathology*, **224** (1), 90–100.

285 Attard, G., Swennenhuis, J.F., Olmos, D., Reid, A.H., Vickers, E., A'Hern, R., Levink, R., Coumans, F., Moreira, J., Riisnaes, R., Oommen, N.B., Hawche, G., Jameson, C., Thompson, E., Sipkema, R., Carden, C.P., Parker, C., Dearnaley, D., Kaye, S.B., Cooper, C.S., Molina, A., Cox, M.E., Terstappen, L.W., and de Bono, J.S. (2009) Characterization of ERG, AR and PTEN gene status in circulating tumor cells from patients with castration-resistant prostate cancer. *Cancer Research*, **69** (7), 2912–2918.

286 Reid, A.H., Attard, G., Ambroisine, L., Fisher, G., Kovacs, G., Brewer, D., Clark, J., Flohr, P., Edwards, S., Berney, D.M., Foster, C.S., Fletcher, A., Gerald, W.L., Moller, H., Reuter, V.E., Scardino, P.T., Cuzick, J., de Bono, J.S., Cooper, C.S., and Transatlantic Prostate, G. (2010) Molecular characterisation of ERG, ETV1 and PTEN gene loci identifies patients at low and high risk of death from prostate cancer. *British Journal of Cancer*, **102** (4), 678–684.

287 Yoshimoto, M., Joshua, A.M., Cunha, I.W., Coudry, R.A., Fonseca, F.P., Ludkovski, O., Zielenska, M., Soares, F.A., and Squire, J.A. (2008) Absence of TMPRSS2–ERG fusions and PTEN losses in prostate cancer is associated with a favorable outcome. *Modern Pathology*, **21** (12), 1451–1460.

288 Chaux, A., Peskoe, S.B., Gonzalez-Roibon, N., Schultz, L., Albadine, R., Hicks, J., De Marzo, A.M., Platz, E.A., and Netto, G.J. (2012) Loss of PTEN expression is associated with increased risk of recurrence after prostatectomy for clinically localized prostate cancer. *Modern Pathology*, **25** (11), 1543–1549.

289 Krohn, A., Diedler, T., Burkhardt, L., Mayer, P.S., De Silva, C., Meyer-Kornblum, M., Kotschau, D., Tennstedt, P., Huang, J., Gerhauser, C., Mader, M., Kurtz, S., Sirma, H., Saad, F., Steuber, T., Graefen, M., Plass, C., Sauter, G., Simon, R., Minner, S., and Schlomm, T. (2012) Genomic deletion of PTEN is associated with tumor progression and early PSA recurrence in ERG fusion-positive and fusion-negative prostate cancer. *American Journal of Pathology*, **181** (2), 401–412.

290 Yoshimoto, M., Ding, K., Sweet, J.M., Ludkovski, O., Trottier, G., Song, K.S., Joshua, A.M., Fleshner, N.E., Squire, J.A., and Evans, A.J. (2013) PTEN losses exhibit heterogeneity in multifocal prostatic adenocarcinoma and are associated with higher Gleason grade. *Modern Pathology*, **26** (3), 435–447.

291 Carver, B.S., Chapinski, C., Wongvipat, J., Hieronymus, H., Chen, Y., Chandarlapaty, S., Arora, V.K., Le, C., Koutcher, J., Scher, H., Scardino, P.T., Rosen, N., and Sawyers, C.L. (2011) Reciprocal feedback regulation of PI3K and androgen receptor signaling in PTEN-deficient prostate cancer. *Cancer Cell*, **19** (5), 575–586.

292 Mulholland, D.J., Tran, L.M., Li, Y., Cai, H., Morim, A., Wang, S., Plaisier, S., Garraway, I.P., Huang, J., Graeber, T.G., and Wu, H. (2011) Cell autonomous role of PTEN in regulating castration-resistant prostate cancer growth. *Cancer Cell*, **19** (6), 792–804.

293 Thompson, T.C. (2011) Turning reciprocal feedback regulation into combination therapy. *Cancer Cell*, **19** (6), 697–699.

294 Nie, Z., Hu, G., Wei, G., Cui, K., Yamane, A., Resch, W., Wang, R., Green, D.R., Tessarollo, L., Casellas, R., Zhao, K., and Levens, D. (2012) c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells. *Cell*, **151** (1), 68–79.

295 Van Dang, C. and McMahon, S.B. (2010) Emerging concepts in the analysis of transcriptional targets of the MYC oncoprotein: are the targets targetable? *Genes & Cancer*, **1** (6), 560–567.

296 Gurel, B., Iwata, T., Koh, C.M., Jenkins, R.B., Lan, F., Van Dang, C., Hicks, J.L., Morgan, J., Cornish, T.C., Sutcliffe, S., Isaacs, W.B., Luo, J., and De Marzo, A.M. (2008) Nuclear MYC protein overexpression is an early alteration in human prostate carcinogenesis. *Modern Pathology*, **21** (9), 1156–1167.

297 Fromont, G., Godet, J., Peyret, A., Irani, J., Celhay, O., Rozet, F., Cathelineau, X., and Cussenot, O. (2013) 8q24 amplification is associated with Myc expression and prostate cancer progression and is an independent predictor of recurrence after radical prostatectomy. *Human Pathology*, **44** (8), 1617–1623.

298 Zafarana, G., Ishkanian, A.S., Malloff, C.A., Locke, J.A., Sykes, J., Thoms, J., Lam, W.L., Squire, J.A., Yoshimoto, M., Ramnarine, V.R., Meng, A., Ahmed, O., Jurisca, I., Milosevic, M., Pintilie, M., van der Kwast, T., and Bristow, R.G. (2012) Copy number alterations of c-MYC and PTEN are prognostic factors for relapse after prostate cancer radiotherapy. *Cancer*, **118** (16), 4053–4062.

299 Korkmaz, K.S., Korkmaz, C.G., Ragnhildstveit, E., Kizildag, S., Pretlow, T.G., and Saatcioglu, F. (2000) Full-length cDNA sequence and genomic organization of human NKX3A – alternative forms and regulation by both androgens and estrogens. *Gene*, **260** (1–2), 25–36.

300 Wang, X., Kruithof-de Julio, M., Economides, K.D., Walker, D., Yu, H., Halili, M.V., Hu, Y.P., Price, S.M., Abate-Shen, C., and Shen, M.M. (2009) A luminal epithelial stem cell that is a cell of origin for prostate cancer. *Nature*, **461** (7263), 495–500.

301 Locke, J.A., Zafarana, G., Ishkanian, A.S., Milosevic, M., Thoms, J., Have, C.L., Malloff, C.A., Lam, W.L., Squire, J.A., Pintilie, M., Sykes, J., Ramnarine, V.R., Meng, A., Ahmed, O., Jurisica, I., van der Kwast, T., and Bristow, R.G. (2012) NKX3.1 haploinsufficiency is prognostic for prostate cancer relapse following surgery or image-guided radiotherapy. *Clinical Cancer Research*, **18** (1), 308–316.

302 Zaman, M.S., Chen, Y., Deng, G., Shahryari, V., Suh, S.O., Saini, S., Majid, S., Liu, J., Khatri, G., Tanaka, Y., and Dahiya, R. (2010) The functional significance of microRNA-145 in prostate cancer. *British Journal of Cancer*, **103** (2), 256–264.

303 Suh, S.O., Chen, Y., Zaman, M.S., Hirata, H., Yamamura, S., Shahryari, V., Liu, J., Tabatabai, Z.L., Kakar, S., Deng, G., Tanaka, Y., and Dahiya, R. (2011) MicroRNA-145 is regulated by DNA methylation and p53 gene mutation in prostate cancer. *Carcinogenesis*, **32** (5), 772–778.

304 Tomlins, S.A., Rhodes, D.R., Perner, S., Dhanasekaran, S.M., Mehra, R., Sun, X.W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R., Lee, C., Montie, J.E., Shah, R.B., Pienta, K.J., Rubin, M.A., and Chinnaiyan, A.M. (2005) Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science*, **310** (5748), 644–648.

305 Seth, A. and Watson, D.K. (2005) ETS transcription factors and their emerging roles in human cancer. *European Journal of Cancer*, **41** (16), 2462–2478.

306 Park, K., Tomlins, S.A., Mudaliar, K.M., Chiu, Y.L., Esgueva, R., Mehra, R., Suleman, K., Varambally, S., Brenner, J.C., MacDonald, T., Srivastava, A., Tewari, A.K., Sathyanarayana, U., Nagy, D., Pestano, G., Kunju, L.P., Demichelis, F., Chinnaiyan, A.M., and Rubin, M.A. (2010) Antibody-based detection of ERG

rearrangement-positive prostate cancer. *Neoplasia*, **12** (7), 590–598.

307 Hart, A.H., Corrick, C.M., Tymms, M.J., Hertzog, P.J., and Kola, I. (1995) Human ERG is a proto-oncogene with mitogenic and transforming activity. *Oncogene*, **10** (7), 1423–1430.

308 Mani, R.S., Iyer, M.K., Cao, Q., Brenner, J.C., Wang, L., Ghosh, A., Cao, X., Lonigro, R.J., Tomlins, S.A., Varambally, S., and Chinnaiyan, A.M. (2011) TMPRSS2–ERG-mediated feed-forward regulation of wild-type ERG in human prostate cancers. *Cancer Research*, **71** (16), 5387–5392.

309 Carver, B.S., Tran, J., Gopalan, A., Chen, Z., Shaikh, S., Carracedo, A., Alimonti, A., Nardella, C., Varmeh, S., Scardino, P.T., Cordon-Cardo, C., Gerald, W., and Pandolfi, P.P. (2009) Aberrant ERG expression cooperates with loss of PTEN to promote cancer progression in the prostate. *Nature Genetics*, **41** (5), 619–624.

310 Toubaji, A., Albadine, R., Meeker, A.K., Isaacs, W.B., Lotan, T., Haffner, M.C., Chaux, A., Epstein, J.I., Han, M., Walsh, P.C., Partin, A.W., De Marzo, A.M., Platz, E.A., and Netto, G.J. (2011) Increased gene copy number of ERG on chromosome 21 but not TMPRSS2–ERG fusion predicts outcome in prostatic adenocarcinomas. *Modern Pathology*, **24** (11), 1511–1520.

311 Pettersson, A., Graff, R.E., Bauer, S.R., Pitt, M.J., Lis, R.T., Stack, E.C., Martin, N.E., Kunz, L., Penney, K.L., Ligon, A.H., Suppan, C., Flavin, R., Sesso, H.D., Rider, J.R., Sweeney, C., Stampfer, M.J., Fiorentino, M., Kantoff, P.W., Sanda, M.G., Giovannucci, E.L., Ding, E.L., Loda, M., and Mucci, L.A. (2012) The TMPRSS2–ERG rearrangement, ERG expression, and prostate cancer outcomes: a cohort study and meta-analysis. *Cancer Epidemiology, Biomarkers & Prevention*, **21** (9), 1497–1509.

312 Ferlay, J., Parkin, D.M., and Steliarova-Foucher, E. (2010) Estimates of cancer incidence and mortality in Europe in 2008. *European Journal of Cancer*, **46** (4), 765–781.

313 Hammerschmied, C.G., Walter, B., and Hartmann, A. (2008) [Renal cell carcinoma 2008. Histopathology, molecular genetics and new therapeutic options]. *Pathologe*, **29** (5), 354–363.

314 Chow, W.H. and Devesa, S.S. (2008) Contemporary epidemiology of renal cell cancer. *Cancer Journal*, **14** (5), 288–301.

315 Hunt, J.D., van der Hel, O.L., McMillan, G.P., Boffetta, P., and Brennan, P. (2005) Renal cell carcinoma in relation to cigarette smoking: meta-analysis of 24 studies. *International Journal of Cancer*, **114** (1), 101–108.

316 Chow, W.H., Gridley, G., Fraumeni, J.F. Jr., and Jarvholm, B. (2000) Obesity, hypertension, and the risk of kidney cancer in men. *New England Journal of Medicine*, **343** (18), 1305–1311.

317 Weikert, S., Boeing, H., Pischon, T., Weikert, C., Olsen, A., Tjonneland, A., Overvad, K., Becker, N., Linseisen, J., Trichopoulou, A., Mountokalakis, T., Trichopoulos, D., Sieri, S., Palli, D., Vineis, P., Panico, S., Peeters, P.H., Bueno-de-Mesquita, H.B., Verschuren, W.M., Ljungberg, B., Hallmans, G., Berglund, G., Gonzalez, C.A., Dorronsoro, M., Barricarte, A., Tormo, M.J., Allen, N., Roddam, A., Bingham, S., Khaw, K.T., Rinaldi, S., Ferrari, P., Norat, T., and Riboli, E. (2008) Blood pressure and risk of renal cell carcinoma in the European prospective investigation into cancer and nutrition. *American Journal of Epidemiology*, **167** (4), 438–446.

318 Pavlovich, C.P. and Schmidt, L.S. (2004) Searching for the hereditary causes of renal-cell carcinoma. *Nature Reviews Cancer*, **4** (5), 381–393.

319 Cairns, P. (2010) Renal cell carcinoma. *Cancer Biomarkers: Section A of Disease Markers*, **9** (1–6), 461–473.

320 Verine, J., Pluvinage, A., Bousquet, G., Lehmann-Che, J., de Bazelaire, C., Soufir, N., and Mongiat-Artus, P. (2010) Hereditary renal cancer syndromes: an update of a systematic review. *European Urology*, **58** (5), 701–710.

321 Hagenkord, J.M., Gatalica, Z., Jonasch, E., and Monzon, F.A. (2011) Clinical genomics of renal epithelial tumors. *Cancer Genetics*, **204** (6), 285–297.

322 Maher, E.R. (2004) Von Hippel–Lindau disease. *Current Molecular Medicine*, **4** (8), 833–842.

323 Blankenship, C., Naglich, J.G., Whaley, J.M., Seizinger, B., and Kley, N. (1999) Alternate choice of initiation codon produces a biologically active product of the von Hippel Lindau gene with tumor suppressor activity. *Oncogene*, **18** (8), 1529–1535.

324 Chen, F., Kishida, T., Yao, M., Hustad, T., Glavac, D., Dean, M., Gnarra, J.R., Orcutt, M.L., Duh, F.M., Glenn, G. *et al.* (1995) Germline mutations in the von Hippel–Lindau disease tumor suppressor gene: correlations with phenotype. *Human Mutation*, **5** (1), 66–75.

325 Zbar, B., Kishida, T., Chen, F., Schmidt, L., Maher, E.R., Richards, F.M., Crossey, P.A., Webster, A.R., Affara, N.A., Ferguson-Smith, M.A., Brauch, H., Glavac, D., Neumann, H.P., Tisherman, S., Mulvihill, J.J., Gross, D.J., Shuin, T., Whaley, J., Seizinger, B., Kley, N., Olschwang, S., Boisson, C., Richard, S., Lips, C.H., Lerman, M. *et al.* (1996) Germline mutations in the Von Hippel–Lindau disease (VHL) gene in families from North America, Europe, and Japan. *Human Mutation*, **8** (4), 348–357.

326 Stolle, C., Glenn, G., Zbar, B., Humphrey, J.S., Choyke, P., Walther, M., Pack, S., Hurley, K., Andrey, C., Klausner, R., and Linehan, W.M. (1998) Improved detection of germline mutations in the von Hippel–Lindau disease tumor suppressor gene. *Human Mutation*, **12** (6), 417–423.

327 Vortmeyer, A.O., Huang, S.C., Pack, S.D., Koch, C.A., Lubensky, I.A., Oldfield, E.H., and Zhuang, Z. (2002) Somatic point mutation of the wild-type allele detected in tumors of patients with VHL germline deletion. *Oncogene*, **21** (8), 1167–1170.

328 Banks, R.E., Tirukonda, P., Taylor, C., Hornigold, N., Astuti, D., Cohen, D., Maher, E.R., Stanley, A.J., Harnden, P., Joyce, A., Knowles, M., and Selby, P.J. (2006) Genetic and epigenetic analysis of von Hippel–Lindau (VHL) gene alterations and relationship with clinical variables in sporadic renal cancer. *Cancer Research*, **66** (4), 2000–2011.

329 Herman, J.G., Latif, F., Weng, Y., Lerman, M.I., Zbar, B., Liu, S., Samid, D., Duan, D.S., Gnarra, J.R., Linehan, W.M. *et al.* (1994) Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proceedings of the National Academy of Sciences of the United States of America*, **91** (21), 9700–9704.

330 Maher, E.R. and Kaelin, W.G. Jr. (1997) von Hippel–Lindau disease. *Medicine*, **76** (6), 381–391.

331 Walther, M.M., Choyke, P.L., Weiss, G., Manolatos, C., Long, J., Reiter, R., Alexander, R.B., and Linehan, W.M. (1995) Parenchymal sparing surgery in patients with hereditary renal cell carcinoma. *Journal of Urology*, **153** (3), 913–916.

332 Friedrich, C.A. (2001) Genotype-phenotype correlation in von Hippel–Lindau syndrome. *Human Molecular Genetics*, **10** (7), 763–767.

333 Kaelin, W.G. (2007) Von Hippel–Lindau disease. *Annual Review of Pathology*, 2145–173.

334 Clifford, S.C., Cockman, M.E., Smallwood, A.C., Mole, D.R., Woodward, E.R., Maxwell, P.H., Ratcliffe, P.J., and Maher, E.R. (2001) Contrasting effects on HIF-1alpha regulation by disease-causing pVHL mutations correlate with patterns of tumourigenesis in von Hippel–Lindau disease. *Human Molecular Genetics*, **10** (10), 1029–1038.

335 Nordstrom-O'Brien, M., van der Luijt, R.B., van Rooijen, E., van den Ouweland, A.M., Majoor-Krakauer, D.F., Lolkema, M.P., van Brussel, A., Voest, E.E., and Giles, R.H. (2010) Genetic analysis of von Hippel–Lindau disease. *Human Mutation*, **31** (5), 521–537.

336 Zbar, B., Glenn, G., Lubensky, I., Choyke, P., Walther, M.M., Magnusson, G., Bergerheim, U.S., Pettersson, S., Amin, M., and Hurley, K. (1995) Hereditary papillary renal cell carcinoma: clinical studies in 10 families. *Journal of Urology*, **153** (3), 907–912.

337 Linehan, W.M., Bratslavsky, G., Pinto, P.A., Schmidt, L.S., Neckers, L., Bottaro, D.P., and Srinivasan, R. (2010) Molecular

diagnosis and therapy of kidney cancer. *Annual Review of Medicine*, **61**, 329–343.

338 Ornstein, D.K., Lubensky, I.A., Venzon, D., Zbar, B., Linehan, W.M., and Walther, M.M. (2000) Prevalence of microscopic tumors in normal appearing renal parenchyma of patients with hereditary papillary renal cancer. *Journal of Urology*, **163** (2), 431–433.

339 Schmidt, L., Duh, F.M., Chen, F., Kishida, T., Glenn, G., Choyke, P., Scherer, S.W., Zhuang, Z., Lubensky, I., Dean, M., Allikmets, R., Chidambaram, A., Bergerheim, U.R., Feltis, J.T., Casadevall, C., Zamarron, A., Bernues, M., Richard, S., Lips, C.J., Walther, M.M., Tsui, L.C., Geil, L., Orcutt, M.L., Stackhouse, T., Lipan, J., Slife, L., Brauch, H., Decker, J., Niehans, G., Hughson, M.D., Moch, H., Storkel, S., Lerman, M.I., Linehan, W.M., and Zbar, B. (1997) Germline and somatic mutations in the tyrosine kinase domain of the MET proto-oncogene in papillary renal carcinomas. *Nature Genetics*, **16** (1), 68–73.

340 Schmidt, L., Junker, K., Nakaigawa, N., Kinjerski, T., Weirich, G., Miller, M., Lubensky, I., Neumann, H.P., Brauch, H., Decker, J., Vocke, C., Brown, J.A., Jenkins, R., Richard, S., Bergerheim, U., Gerrard, B., Dean, M., Linehan, W.M., and Zbar, B. (1999) Novel mutations of the MET proto-oncogene in papillary renal carcinomas. *Oncogene*, **18** (14), 2343–2350.

341 Schmidt, L.S., Nickerson, M.L., Angeloni, D., Glenn, G.M., Walther, M.M., Albert, P.S., Warren, M.B., Choyke, P.L., Torres-Cabala, C.A., Merino, M.J., Brunet, J., Berez, V., Borras, J., Sesia, G., Middelton, L., Phillips, J.L., Stolle, C., Zbar, B., Pautler, S.E., and Linehan, W.M. (2004) Early onset hereditary papillary renal carcinoma: germline missense mutations in the tyrosine kinase domain of the met proto-oncogene. *Journal of Urology*, **172** (4 Pt 1), 1256–1261.

342 Zbar, B. and Lerman, M. (1998) Inherited carcinomas of the kidney. *Advances in Cancer Research*, **75**, 163–201.

343 Jeffers, M., Schmidt, L., Nakaigawa, N., Webb, C.P., Weirich, G., Kishida, T., Zbar, B., and Vande Woude, G.F. (1997) Activating mutations for the met tyrosine kinase receptor in human cancer. *Proceedings of the National Academy of Sciences of the United States of America*, **94** (21), 11445–11450.

344 Jeffers, M.F. and Vande Woude, G.F. (1999) Activating mutations in the Met receptor overcome the requirement for autophosphorylation of tyrosines crucial for wild type signaling. *Oncogene*, **18** (36), 5120–5125.

345 Lubensky, I.A., Schmidt, L., Zhuang, Z., Weirich, G., Pack, S., Zambrano, N., Walther, M.M., Choyke, P., Linehan, W. M., and Zbar, B. (1999) Hereditary and sporadic papillary renal carcinomas with c-met mutations share a distinct morphological phenotype. *American Journal of Pathology*, **155** (2), 517–526.

346 Zhuang, Z., Park, W.S., Pack, S., Schmidt, L., Vortmeyer, A.O., Pak, E., Pham, T., Weil, R.J., Candidus, S., Lubensky, I.A., Linehan, W.M., Zbar, B., and Weirich, G. (1998) Trisomy 7-harbouring non-random duplication of the mutant MET allele in hereditary papillary renal carcinomas. *Nature Genetics*, **20** (1), 66–69.

347 Alam, N.A., Bevan, S., Churchman, M., Barclay, E., Barker, K., Jaeger, E.E., Nelson, H.M., Healy, E., Pembroke, A.C., Friedmann, P.S., Dalziel, K., Calonje, E., Anderson, J., August, P.J., Davies, M.G., Felix, R., Munro, C.S., Murdoch, M., Rendall, J., Kennedy, S., Leigh, I.M., Kelsell, D.P., Tomlinson, I.P., and Houlston, R.S. (2001) Localization of a gene (MCUL1) for multiple cutaneous leiomyomata and uterine fibroids to chromosome 1q42.3–q43. *American Journal of Human Genetics*, **68** (5), 1264–1269.

348 Tomlinson, I.P., Alam, N.A., Rowan, A.J., Barclay, E., Jaeger, E.E., Kelsell, D., Leigh, I., Gorman, P., Lamlum, H., Rahman, S., Roylance, R.R., Olpin, S., Bevan, S., Barker, K., Hearle, N., Houlston, R.S., Kiuru, M., Lehtonen, R., Karhu, A., Vilkki, S., Laiho, P., Eklund, C., Vierimaa, O., Aittomaki, K., Hietala, M., Sistonen, P., Paetau, A., Salovaara, R., Herva, R., Launonen, V., Aaltonen, L.A., and Multiple Leiomyoma, C. (2002) Germline

mutations in FH predispose to dominantly inherited uterine fibroids, skin leiomyomata and papillary renal cell cancer. *Nature Genetics*, **30** (4), 406–410.

349 Pithukpakorn, M., Wei, M.H., Toure, O., Steinbach, P.J., Glenn, G.M., Zbar, B., Linehan, W.M., and Toro, J.R. (2006) Fumarate hydratase enzyme activity in lymphoblastoid cells and fibroblasts of individuals in families with hereditary leiomyomatosis and renal cell cancer. *Journal of Medical Genetics*, **43** (9), 755–762.

350 Toro, J.R., Nickerson, M.L., Wei, M.H., Warren, M.B., Glenn, G.M., Turner, M. L., Stewart, L., Duray, P., Tourre, O., Sharma, N., Choyke, P., Stratton, P., Merino, M., Walther, M.M., Linehan, W. M., Schmidt, L.S., and Zbar, B. (2003) Mutations in the fumarate hydratase gene cause hereditary leiomyomatosis and renal cell cancer in families in North America. *American Journal of Human Genetics*, **73** (1), 95–106.

351 Merino, M.J., Torres-Cabala, C., Pinto, P., and Linehan, W.M. (2007) The morphologic spectrum of kidney tumors in hereditary leiomyomatosis and renal cell carcinoma (HLRCC) syndrome. *American Journal of Surgical Pathology*, **31** (10), 1578–1585.

352 Rosner, I., Bratslavsky, G., Pinto, P.A., and Linehan, W.M. (2009) The clinical implications of the genetics of renal cell carcinoma. *Urologic Oncology*, **27** (2), 131–136.

353 Gardie, B., Remenieras, A., Kattygnarath, D., Bombled, J., Lefevre, S., Perrier-Trudova, V., Rustin, P., Barrois, M., Slama, A., Avril, M.F., Bessis, D., Caron, O., Caux, F., Collignon, P., Coupier, I., Cremin, C., Dollfus, H., Dugast, C., Escudier, B., Faivre, L., Field, M., Gilbert-Dussardier, B., Janin, N., Leport, Y., Leroux, D., Lipsker, D., Malthieu, F., McGilliwray, B., Maugard, C., Mejean, A., Mortemousque, I., Plessis, G., Poppe, B., Pruvost-Balland, C., Rooker, S., Roume, J., Soufir, N., Steinraths, M., Tan, M.H., Theodore, C., Thomas, L., Vabres, P., Van Glabeke, E., Meric, J.B., Verkarre, V., Lenoir, G., Joulin, V., Deveaux, S., Cusin, V., Feunteun, J., Teh, B.T., Bressac-de Paillerets, B., and Richard, S. (2011) Novel FH mutations in families with hereditary leiomyomatosis and renal cell cancer (HLRCC) and patients with isolated type 2 papillary renal cell carcinoma. *Journal of Medical Genetics*, **48** (4), 226–234.

354 Birt, A.R., Hogg, G.R., and Dube, W.J. (1977) Hereditary multiple fibrofolliculomas with trichodiscomas and acrochordons. *Archives of Dermatology*, **113** (12), 1674–1677.

355 Pavlovich, C.P., Grubb, R.L. 3rd, Hurley, K., Glenn, G.M., Toro, J., Schmidt, L.S., Torres-Cabala, C., Merino, M.J., Zbar, B., Choyke, P., Walther, M.M., and Linehan, W.M. (2005) Evaluation and management of renal tumors in the Birt–Hogg–Dube syndrome. *Journal of Urology*, **173** (5), 1482–1486.

356 Pavlovich, C.P., Walther, M.M., Eyler, R. A., Hewitt, S.M., Zbar, B., Linehan, W.M., and Merino, M.J. (2002) Renal tumors in the Birt–Hogg–Dube syndrome. *American Journal of Surgical Pathology*, **26** (12), 1542–1552.

357 Tickoo, S.K., Reuter, V.E., Amin, M.B., Srigley, J.R., Epstein, J.I., Min, K.W., Rubin, M.A., and Ro, J.Y. (1999) Renal oncocytosis: a morphologic study of fourteen cases. *American Journal of Surgical Pathology*, **23** (9), 1094–1101.

358 Schmidt, L.S., Warren, M.B., Nickerson, M.L., Weirich, G., Matrosova, V., Toro, J. R., Turner, M.L., Duray, P., Merino, M., Hewitt, S., Pavlovich, C.P., Glenn, G., Greenberg, C.R., Linehan, W.M., and Zbar, B. (2001) Birt–Hogg–Dube syndrome, a genodermatosis associated with spontaneous pneumothorax and kidney neoplasia, maps to chromosome 17p11.2. *American journal of human genetics*, **69** (4), 876–882.

359 Khoo, S.K., Giraud, S., Kahnoski, K., Chen, J., Motorna, O., Nickolov, R., Binet, O., Lambert, D., Friedel, J., Levy, R., Ferlicot, S., Wolkenstein, P., Hammel, P., Bergerheim, U., Hedblad, M.A., Bradley, M., Teh, B.T., Nordenskjold, M., and Richard, S. (2002) Clinical and genetic studies of Birt–Hogg–Dube syndrome. *Journal of Medical Genetics*, **39** (12), 906–912.

360 Nickerson, M.L., Warren, M.B., Toro, J.R., Matrosova, V., Glenn, G., Turner, M.L., Duray, P., Merino, M., Choyke, P., Pavlovich, C.P., Sharma, N., Walther, M., Munroe, D., Hill, R., Maher, E., Greenberg, C., Lerman, M.I., Linehan, W.M., Zbar, B., and Schmidt, L.S. (2002) Mutations in a novel gene lead to kidney tumors, lung wall defects, and benign tumors of the hair follicle in patients with the Birt–Hogg–Dube syndrome. *Cancer Cell*, **2** (2), 157–164.

361 Toro, J.R., Wei, M.H., Glenn, G.M., Weinreich, M., Toure, O., Vocke, C., Turner, M., Choyke, P., Merino, M.J., Pinto, P.A., Steinberg, S.M., Schmidt, L.S., and Linehan, W.M. (2008) BHD mutations, clinical and molecular genetic investigations of Birt–Hogg–Dube syndrome: a new series of 50 families and a review of published reports. *Journal of Medical Genetics*, **45** (6), 321–331.

362 da Silva, N.F., Gentle, D., Hesson, L.B., Morton, D.G., Latif, F., and Maher, E.R. (2003) Analysis of the Birt–Hogg–Dube (BHD) tumour suppressor gene in sporadic renal cell carcinoma and colorectal cancer. *Journal of Medical Genetics*, **40** (11), 820–824.

363 Khoo, S.K., Kahnoski, K., Sugimura, J., Petillo, D., Chen, J., Shockley, K., Ludlow, J., Knapp, R., Giraud, S., Richard, S., Nordenskjold, M., and Teh, B.T. (2003) Inactivation of BHD in sporadic renal tumors. *Cancer Research*, **63** (15), 4583–4587.

364 Crino, P.B., Nathanson, K.L., and Henske, E.P. (2006) The tuberous sclerosis complex. *New England Journal of Medicine*, **355** (13), 1345–1356.

365 Lin, L., Zhang, J.H., Panicker, L.M., and Simonds, W.F. (2008) The parafibromin tumor suppressor protein inhibits cell proliferation by repression of the c-myc proto-oncogene. *Proceedings of the National Academy of Sciences of the United States of America*, **105** (45), 17420–17425.

366 Ricketts, C., Woodward, E.R., Killick, P., Morris, M.R., Astuti, D., Latif, F., and Maher, E.R. (2008) Germline SDHB mutations and familial renal cell carcinoma. *Journal of the National Cancer Institute*, **100** (17), 1260–1262.

367 Linehan, W.M., Srinivasan, R., and Schmidt, L.S. (2010) The genetic basis of kidney cancer: a metabolic disease. *Nature Reviews Urology*, **7** (5), 277–285.

368 Gnarra, J.R., Tory, K., Weng, Y., Schmidt, L., Wei, M.H., Li, H., Latif, F., Liu, S., Chen, F., Duh, F.M. *et al.* (1994) Mutations of the VHL tumour suppressor gene in renal carcinoma. *Nature Genetics*, **7** (1), 85–90.

369 Dalgliesh, G.L., Furge, K., Greenman, C., Chen, L., Bignell, G., Butler, A., Davies, H., Edkins, S., Hardy, C., Latimer, C., Teague, J., Andrews, J., Barthorpe, S., Beare, D., Buck, G., Campbell, P.J., Forbes, S., Jia, M., Jones, D., Knott, H., Kok, C.Y., Lau, K.W., Leroy, C., Lin, M.L., McBride, D.J., Maddison, M., Maguire, S., McLay, K., Menzies, A., Mironenko, T., Mulderrig, L., Mudie, L., O'Meara, S., Pleasance, E., Rajasingham, A., Shepherd, R., Smith, R., Stebbings, L., Stephens, P., Tang, G., Tarpey, P.S., Turrell, K., Dykema, K.J., Khoo, S.K., Petillo, D., Wondergem, B., Anema, J., Kahnoski, R. J., Teh, B.T., Stratton, M.R., and Futreal, P.A. (2010) Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature*, **463** (7279), 360–363.

370 Dulaimi, E., Ibanez de Caceres, I., Uzzo, R.G., Al-Saleem, T., Greenberg, R.E., Polascik, T.J., Babb, J.S., Grizzle, W.E., and Cairns, P. (2004) Promoter hypermethylation profile of kidney cancer. *Clinical Cancer Research*, **10** (12 Pt 1), 3972–3979.

371 Monzon, F.A., Alvarez, K., Peterson, L., Truong, L., Amato, R.J., Hernandez-McClain, J., Tannir, N., Parwani, A.V., and Jonasch, E. (2011) Chromosome 14q loss defines a molecular subtype of clear-cell renal cell carcinoma associated with poor prognosis. *Modern Pathology*, **24** (11), 1470–1479.

372 Beroukhim, R., Brunet, J.P., Di Napoli, A., Mertz, K.D., Seeley, A., Pires, M.M., Linhart, D., Worrell, R.A., Moch, H., Rubin, M.A., Sellers, W.R., Meyerson, M., Linehan, W.M., Kaelin, W.G. Jr., and Signoretti, S. (2009) Patterns of gene

expression and copy-number alterations in von-hippel lindau disease-associated and sporadic clear cell carcinoma of the kidney. *Cancer Research*, **69** (11), 4674–4681.

373 Fischer, J., Palmedo, G., von Knobloch, R., Bugert, P., Prayer-Galetti, T., Pagano, F., and Kovacs, G. (1998) Duplication and overexpression of the mutant allele of the MET proto-oncogene in multiple hereditary papillary renal cell tumours. *Oncogene*, **17** (6), 733–739.

374 Kiuru, M., Lehtonen, R., Arola, J., Salovaara, R., Jarvinen, H., Aittomaki, K., Sjoberg, J., Visakorpi, T., Knuutila, S., Isola, J., Delahunt, B., Herva, R., Launonen, V., Karhu, A., and Aaltonen, L.A. (2002) Few FH mutations in sporadic counterparts of tumor types observed in hereditary leiomyomatosis and renal cell cancer families. *Cancer Research*, **62** (16), 4554–4557.

375 Kovacs, G., Fuzesi, L., Emanual, A., and Kung, H.F. (1991) Cytogenetics of papillary renal cell tumors. *Genes, Chromosomes & Cancer*, **3** (4), 249–255.

376 Kovacs, G. (1993) Molecular cytogenetics of renal cell tumors. *Advances in Cancer Research*, **62**, 89–124.

377 Speicher, M.R., Schoell, B., du Manoir, S., Schrock, E., Ried, T., Cremer, T., Storkel, S., Kovacs, A., and Kovacs, G. (1994) Specific loss of chromosomes 1, 2, 6, 10, 13, 17, and 21 in chromophobe renal cell carcinomas revealed by comparative genomic hybridization. *American Journal of Pathology*, **145** (2), 356–364.

378 Contractor, H., Zariwala, M., Bugert, P., Zeisler, J., and Kovacs, G. (1997) Mutation of the p53 tumour suppressor gene occurs preferentially in the chromophobe type of renal cell tumour. *Journal of Pathology*, **181** (2), 136–139.

379 Argani, P. and Ladanyi, M. (2005) Translocation carcinomas of the kidney. *Clinics in Laboratory Medicine*, **25** (2), 363–378.

380 Komai, Y., Fujiwara, M., Fujii, Y., Mukai, H., Yonese, J., Kawakami, S., Yamamoto, S., Migita, T., Ishikawa, Y., Kurata, M., Nakamura, T., and Fukui, I. (2009) Adult Xp11 translocation renal cell carcinoma diagnosed by cytogenetics and immunohistochemistry. *Clinical Cancer Research*, **15** (4), 1170–1176.

381 Meyer, P.N., Clark, J.I., Flanigan, R.C., and Picken, M.M. (2007) Xp11.2 translocation renal cell carcinoma with very aggressive course in five adults. *American Journal of Clinical Pathology*, **128** (1), 70–79.

382 Koie, T., Yoneyama, T., Hashimoto, Y., Kamimura, N., Kusumi, T., Kijima, H., and Ohyama, C. (2009) An aggressive course of Xp11 translocation renal cell carcinoma in a 28-year-old man. *International Journal of Urology*, **16** (3), 333–335.

383 Purdue, M.P., Johansson, M., Zelenika, D., Toro, J.R., Scelo, G., Moore, L.E., Prokhortchouk, E., Wu, X., Kiemeney, L.A., Gaborieau, V., Jacobs, K.B., Chow, W.H., Zaridze, D., Matveev, V., Lubinski, J., Trubicka, J., Szeszenia-Dabrowska, N., Lissowska, J., Rudnai, P., Fabianova, E., Bucur, A., Bencko, V., Foretova, L., Janout, V., Boffetta, P., Colt, J.S., Davis, F.G., Schwartz, K.L., Banks, R.E., Selby, P.J., Harnden, P., Berg, C.D., Hsing, A.W., Grubb, R.L. 3rd, Boeing, H., Vineis, P., Clavel-Chapelon, F., Palli, D., Tumino, R., Krogh, V., Panico, S., Duell, E.J., Quiros, J.R., Sanchez, M.J., Navarro, C., Ardanaz, E., Dorronsoro, M., Khaw, K.T., Allen, N.E., Bueno-de-Mesquita, H.B., Peeters, P.H., Trichopoulos, D., Linseisen, J., Ljungberg, B., Overvad, K., Tjonneland, A., Romieu, I., Riboli, E., Mukeria, A., Shangina, O., Stevens, V.L., Thun, M.J., Diver, W.R., Gapstur, S.M., Pharoah, P. D., Easton, D.F., Albanes, D., Weinstein, S.J., Virtamo, J., Vatten, L., Hveem, K., Njolstad, I., Tell, G.S., Stoltenberg, C., Kumar, R., Koppova, K., Cussenot, O., Benhamou, S., Oosterwijk, E., Vermeulen, S.H., Aben, K.K., van der Marel, S.L., Ye, Y., Wood, C.G., Pu, X., Mazur, A.M., Boulygina, E.S., Chekanov, N.N., Foglio, M., Lechner, D., Gut, I., Heath, S., Blanche, H., Hutchinson, A., Thomas, G., Wang, Z., Yeager, M., Fraumeni, J.F. Jr., Skryabin, K.G., McKay, J.D., Rothman, N., Chanock, S.J., Lathrop, M., and Brennan, P. (2011) Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21

and 11q13.3. *Nature Genetics*, **43** (1), 60–65.

384 Wu, X., Scelo, G., Purdue, M.P., Rothman, N., Johansson, M., Ye, Y., Wang, Z., Zelenika, D., Moore, L.E., Wood, C.G., Prokhortchouk, E., Gaborieau, V., Jacobs, K.B., Chow, W.H., Toro, J.R., Zaridze, D., Lin, J., Lubinski, J., Trubicka, J., Szeszenia-Dabrowska, N., Lissowska, J., Rudnai, P., Fabianova, E., Mates, D., Jinga, V., Bencko, V., Slamova, A., Holcatova, I., Navratilova, M., Janout, V., Boffetta, P., Colt, J.S., Davis, F.G., Schwartz, K.L., Banks, R.E., Selby, P.J., Harnden, P., Berg, C.D., Hsing, A.W., Grubb, R.L. 3rd, Boeing, H., Vineis, P., Clavel-Chapelon, F., Palli, D., Tumino, R., Krogh, V., Panico, S., Duell, E.J., Quiros, J.R., Sanchez, M.J., Navarro, C., Ardanaz, E., Dorronsoro, M., Khaw, K.T., Allen, N.E., Bueno-de-Mesquita, H.B., Peeters, P.H., Trichopoulos, D., Linseisen, J., Ljungberg, B., Overvad, K., Tjonneland, A., Romieu, I., Riboli, E., Stevens, V.L., Thun, M.J., Diver, W.R., Gapstur, S.M., Pharoah, P.D., Easton, D.F., Albanes, D., Virtamo, J., Vatten, L., Hveem, K., Fletcher, T., Koppova, K., Cussenot, O., Cancel-Tassin, G., Benhamou, S., Hildebrandt, M.A., Pu, X., Foglio, M., Lechner, D., Hutchinson, A., Yeager, M., Fraumeni, J.F. Jr., Lathrop, M., Skryabin, K.G., McKay, J.D., Gu, J., Brennan, P., and Chanock, S.J. (2012) A genome-wide association study identifies a novel susceptibility locus for renal cell carcinoma on 12p11.23. *Human Molecular Genetics*, **21** (2), 456–462.

385 Han, S.S., Yeager, M., Moore, L.E., Wei, M.H., Pfeiffer, R., Toure, O., Purdue, M.P., Johansson, M., Scelo, G., Chung, C.C., Gaborieau, V., Zaridze, D., Schwartz, K., Szeszenia-Dabrowska, N., Davis, F., Bencko, V., Colt, J.S., Janout, V., Matveev, V., Foretova, L., Mates, D., Navratilova, M., Boffetta, P., Berg, C.D., Grubb, R.L. 3rd, Stevens, V.L., Thun, M.J., Diver, W.R., Gapstur, S.M., Albanes, D., Weinstein, S.J., Virtamo, J., Burdett, L., Brisuda, A., McKay, J.D., Fraumeni, J.F. Jr., Chatterjee, N., Rosenberg, P.S., Rothman, N., Brennan, P., Chow, W.H., Tucker, M.A., Chanock, S.J., and Toro, J.R. (2012) The chromosome 2p21 region harbors a complex genetic architecture for association with risk for renal cell carcinoma. *Human Molecular Genetics*, **21** (5), 1190–1200.

386 Henrion, M., Frampton, M., Scelo, G., Purdue, M., Ye, Y., Broderick, P., Ritchie, A., Kaplan, R., Meade, A., McKay, J., Johansson, M., Lathrop, M., Larkin, J., Rothman, N., Wang, Z., Chow, W.H., Stevens, V.L., Ryan Diver, W., Gapstur, S.M., Albanes, D., Virtamo, J., Wu, X., Brennan, P., Chanock, S., Eisen, T., and Houlston, R.S. (2013) Common variation at 2q22.3 (ZEB2) influences the risk of renal cancer. *Human Molecular Genetics*, **22** (4), 825–831.

387 Horikawa, Y., Wood, C.G., Yang, H., Zhao, H., Ye, Y., Gu, J., Lin, J., Habuchi, T., and Wu, X. (2008) Single nucleotide polymorphisms of microRNA machinery genes modify the risk of renal cell carcinoma. *Clinical Cancer Research*, **14** (23), 7956–7962.

388 Zhao, A., Li, G., Peoc'h, M., Genin, C., and Gigante, M. (2013) Serum miR-210 as a novel biomarker for molecular diagnosis of clear cell renal cell carcinoma. *Experimental and Molecular Pathology*, **94** (1), 115–120.

389 Youssef, Y.M., White, N.M., Grigull, J., Krizova, A., Samy, C., Mejia-Guerrero, S., Evans, A., and Yousef, G.M. (2011) Accurate molecular classification of kidney cancer subtypes using microRNA signature. *European Urology*, **59** (5), 721–730.

390 Wach, S., Nolte, E., Theil, A., Stohr, C., Rau, T.T., Hartmann, A., Ekici, A., Keck, B., Taubert, H., and Wullich, B. (2013) MicroRNA profiles classify papillary renal cell carcinoma subtypes. *British Journal of Cancer*, **109** (3), 714–722.

391 de Martino, M., Klatte, T., Haitel, A., and Marberger, M. (2012) Serum cell-free DNA in renal cell carcinoma: a diagnostic and prognostic marker. *Cancer*, **118** (1), 82–90.

392 Ellinger, J., Muller, D.C., Muller, S.C., Hauser, S., Heukamp, L.C., von Ruecker, A., Bastian, P.J., and Walgenbach-Brunagel, G. (2012) Circulating mitochondrial DNA in serum: a universal

diagnostic biomarker for patients with urological malignancies. *Urologic Oncology*, **30** (4), 509–515.

393 Urakami, S., Shiina, H., Enokida, H., Hirata, H., Kawamoto, K., Kawakami, T., Kikuno, N., Tanaka, Y., Majid, S., Nakagawa, M., Igawa, M., and Dahiya, R. (2006) Wnt antagonist family genes as biomarkers for diagnosis, staging, and prognosis of renal cell carcinoma using tumor and serum DNA. *Clinical Cancer Research*, **12** (23), 6989–6997.

394 Verine, J., Lehmann-Che, J., Soliman, H., Feugeas, J.P., Vidal, J.S., Mongiat-Artus, P., Belhadj, S., Philippe, J., Lesage, M., Wittmer, E., Chanel, S., Couvelard, A., Ferlicot, S., Rioux-Leclercq, N., Vignaud, J.M., Janin, A., and Germain, S. (2010) Determination of angptl4 mRNA as a diagnostic marker of primary and metastatic clear cell renal-cell carcinoma. *PLoS One*, **5** (4), e10421.

395 Paradis, V., Bieche, I., Dargere, D., Bonvoust, F., Ferlicot, S., Olivi, M., Lagha, N.B., Blanchet, P., Benoit, G., Vidaud, M., and Bedossa, P. (2001) hTERT expression in sporadic renal cell carcinomas. *Journal of Pathology*, **195** (2), 209–217.

396 Diegmann, J., Junker, K., Gerstmayer, B., Bosio, A., Hindermann, W., Rosenhahn, J., and von Eggeling, F. (2005) Identification of CD70 as a diagnostic biomarker for clear cell renal cell carcinoma by gene expression profiling, real-time RT-PCR and immunohistochemistry. *European Journal of Cancer*, **41** (12), 1794–1801.

397 Nakada, C., Matsuura, K., Tsukamoto, Y., Tanigawa, M., Yoshimoto, T., Narimatsu, T., Nguyen, L.T., Hijiya, N., Uchida, T., Sato, F., Mimata, H., Seto, M., and Moriyama, M. (2008) Genome-wide microRNA expression profiling in renal cell carcinoma: significant down-regulation of miR-141 and miR-200c. *Journal of Pathology*, **216** (4), 418–427.

398 Jung, M., Mollenkopf, H.J., Grimm, C., Wagner, I., Albrecht, M., Waller, T., Pilarsky, C., Johannsen, M., Stephan, C., Lehrach, H., Nietfeld, W., Rudel, T., Jung, K., and Kristiansen, G. (2009) MicroRNA profiling of clear cell renal cell cancer identifies a robust signature to define renal malignancy. *Journal of Cellular and Molecular Medicine*, **13** (9B), 3918–3928.

399 Juan, D., Alexe, G., Antes, T., Liu, H., Madabhushi, A., Delisi, C., Ganesan, S., Bhanot, G., and Liou, L.S. (2010) Identification of a microRNA panel for clear-cell kidney cancer. *Urology*, **75** (4), 835–841.

400 Xue, Y.J., Xiao, R.H., Long, D.Z., Zou, X.F., Wang, X.N., Zhang, G.X., Yuan, Y. H., Wu, G.Q., Yang, J., Wu, Y.T., Xu, H., Liu, F.L., and Liu, M. (2012) Overexpression of FoxM1 is associated with tumor progression in patients with clear cell renal cell carcinoma. *Journal of Translational Medicine*, **10**, 200.

401 Shioi, K., Komiya, A., Hattori, K., Huang, Y., Sano, F., Murakami, T., Nakaigawa, N., Kishida, T., Kubota, Y., Nagashima, Y., and Yao, M. (2006) Vascular cell adhesion molecule 1 predicts cancer-free survival in clear cell renal carcinoma patients. *Clinical Cancer Research*, **12** (24), 7339–7346.

402 Zhang, X., Lin, J., Wu, X., Lin, Z., Ning, B., Kadlubar, S., and Kadlubar, F.F. (2012) Association between GSTM1 copy number, promoter variants and susceptibility to urinary bladder cancer. *International Journal of Molecular Epidemiology and Genetics*, **3** (3), 228–236.

403 Pesch, B., Gawrych, K., Rabstein, S., Weiss, T., Casjens, S., Rihs, H.P., Ding, H., Angerer, J., Illig, T., Klopp, N., Bueno-de-Mesquita, H.B., Ros, M.M., Kaaks, R., Chang-Claude, J., Roswall, N., Tjonneland, A., Overvad, K., Clavel-Chapelon, F., Boutron-Ruault, M.C., Dossus, L., Boeing, H., Weikert, S., Trichopoulos, D., Palli, D., Sieri, S., Tumino, R., Panico, S., Quiros, J.R., Gonzalez, C.A., Sanchez, M.J., Dorronsoro, M., Navarro, C., Barricarte, A., Ljungberg, B., Johansson, M., Ulmert, D., Ehrnstrom, R., Khaw, K.T., Wareham, N., Key, T.J., Ferrari, P., Romieu, I., Riboli, E., Bruning, T., and Vineis, P. (2013) *N*-actyltransferase 2 acetylator phenotype, occupation and bladder cancer risk: results from the EPIC cohort. *Cancer Epidemiology, Biomarkers & Prevention*, **22** (11), 2055–2065.

404 Schulz, W.A. (2006) Understanding urothelial carcinoma through cancer pathways. *International Journal of Cancer*, **119** (7), 1513–1518.

405 Ahmad, I., Sansom, O.J., and Leung, H.Y. (2012) Exploring molecular genetics of bladder cancer: lessons learned from mouse models. *Disease Models & Mechanisms*, **5** (3), 323–332.

406 Alam, N.A., Rowan, A.J., Wortham, N.C., Pollard, P.J., Mitchell, M., Tyrer, J.P., Barclay, E., Calonje, E., Manek, S., Adams, S.J., Bowers, P.W., Burrows, N.P., Charles-Holmes, R., Cook, L.J., Daly, B.M., Ford, G.P., Fuller, L.C., Hadfield-Jones, S.E., Hardwick, N., Highet, A.S., Keefe, M., MacDonald-Hull, S.P., Potts, E.D., Crone, M., Wilkinson, S., Camacho-Martinez, F., Jablonska, S., Ratnavel, R., MacDonald, A., Mann, R.J., Grice, K., Guillet, G., Lewis-Jones, M.S., McGrath, H., Seukeran, D.C., Morrison, P.J., Fleming, S., Rahman, S., Kelsell, D., Leigh, I., Olpin, S., and Tomlinson, I.P. (2003) Genetic and functional analyses of FH mutations in multiple cutaneous and uterine leiomyomatosis, hereditary leiomyomatosis and renal cancer, and fumarate hydratase deficiency. *Human Molecular Genetics*, **12** (11), 1241–1252.

6
From the Genetic Make-Up to the Molecular Signature of Non-Coding RNA in Breast Cancer

Michael Schrauder and Reiner Strick

6.1
Introduction

Breast cancer is the most common cancer in women in industrial countries. Worldwide, about 1.4 million women are diagnosed with breast cancer each year and approximately 460 000 deaths are recorded per year [1]. Breast cancer incidence shows strong regional differences with significantly higher incidence rates in more developed countries compared with less developed countries (round 72 versus 29 per 100 000). In developed countries, mortality rates depend on different tumor and patient characteristics. A favorable prognosis is mainly dependent on early cancer detection. A diagnosis of breast cancer with a tumor size smaller than 1 cm results in an average 10-year survival rate of higher than 95%; a primary diagnosis with a tumor size larger than 2 cm results in an average 10 year survival rate of about 85%, but with the tumor already larger than 5 cm, the survival rate reduces to 60% [2–6]. The potency of the planned therapy is also strongly dependent on the stage of the tumor at the time of the diagnosis. Therefore, many breast cancer chemotherapies, with their serious side-effects, could be prevented by an early diagnosis (e.g., by mammography screening). Different studies have shown that mammography screening can reduce breast cancer mortality by about 10–30% [7–9]. Sensitivity ranges from 47% for small cancers (pT1a) to 95% for larger cancers (pT2 and greater). False-positive rates of 8–10% are a major problem of mammography screening, leading to many unnecessary interventions [10]. Furthermore, precursors like ductal carcinoma *in situ* can be detected with mammography with a sensitivity of about 70% [11]. Several factors have been described to impact on the sensitivity and specificity of mammogram examinations, such as mammographic density [12,13], which is an intrinsic risk factor for breast cancer [14]. Detection of small tumors can be a challenge or even impossible in women with very dense breast tissue. In addition to breast tissue density, genetic and environmental factors are associated with breast cancer. Familial clustering, which was first described by Paul Broca in 1866 [15], is one of the most evident risk factors, and led to the discovery of

Nucleic Acids as Molecular Diagnostics, First Edition. Edited by Andreas Keller and Eckart Meese.

BRCA1 and *BRCA2* in the early 1990s [16,17]. Mutations in *BRCA1* and *BRCA2* result in a high penetrance of breast cancer in affected individuals and in early appearance of the disease. In addition to the high-penetrance breast cancer genes, a series of low-penetrance genes has been discovered in the last 15 years.

Other diagnostic methods (e.g., ultrasound, magnetic resonance imaging, and tomosynthesis) have proven useful for breast cancer detection in addition to mammography. However, none of these methods are part of screening modalities worldwide and relatively high false-positive rates are a problem for all diagnostic methods in current use [13,18–22]. As well as imaging, several molecular methods have been described which could improve the early detection of breast cancer. Biomarkers such as circulating tumor cells (CTCs), proteins, or nucleic acids are the subject of intense worldwide research. It is assumed that CTCs could be found in the blood of about 15% of all primary breast cancer patients [23,24]. In addition to the prognostic relevance of these markers, the detection of CTCs or other biomarkers in a patient with a negative imaging result could direct the process of diagnostics to additional tests (e.g., magnetic resonance imaging) and intense surveillance.

This chapter will focus on nucleic acids as molecular diagnostic, prognostic, and predictive tools in patients with breast cancer.

6.2 Molecular Breast Cancer Detection

6.2.1 Circulating Free DNA

It has been known for decades that patients with cancer have elevated blood levels of free DNA (circulating free DNA (cfDNA)) [25,26]. The origin of cfDNA and the responses to higher levels in the blood of cancer patients are still elusive. Different mechanisms, such as cell apoptosis, necrosis, and/or active release, have been suggested as possible explanations for this phenomenon. The plasma of cancer patients contains approximately fourfold greater levels of cfDNA compared with plasma of healthy individuals and persisting high levels of circulating nucleic acids after, for example, breast cancer treatment correlate with poorer overall survival rates [25,27]. Nevertheless, the information from simple cfDNA quantitation is not sufficient for clinical applications; therefore specific characteristics of cfDNAs, such as mutations, copy number variations (CNVs), loss of heterozygosity (LOH), and others, have been used in the attempt to establish cfDNAs as tumor-specific biomarkers. CNVs are amplified or deleted regions of the genome, varying in size. These genomic regions are one of the major causes of normal human genome variability and affect the individual phenotypic variation [28–30]. If one parental copy of a heterozygous region is lost, the difference at these otherwise polymorphic loci is also lost, resulting in the so-called LOH.

Circulating DNA fragments carrying tumor-specific sequence alterations seem to be the most promising subtype of cfDNA with regard to cancer detection. First results included the detection of *PIK3CA* and *TP53* mutations as well as microsatellite instability and LOH analyses of cfDNA, but larger studies are still missing [31–34]. A recent study attracted great attention by showing in a proof-of-concept analysis that circulating tumor DNA levels (identified by somatic genomic alterations) are better correlated with changes in tumor burden of metastatic breast cancer patients than the cancer antigen CA 15-3 or CTCs. According to this study, circulating tumor DNA provided the earliest measure of treatment response in more than 50% of the analyzed population. The authors linked cfDNA found in plasma to the primary tumor by sequencing *PIK3CA* and *TP53* – two of the most commonly mutated genes in breast cancer tissue [35,36]. Other somatic structural variants were identified in five additional patients. All patients with breast cancer accumulate mutations in the DNA of their tumor, but only around 60% of the analyzed patients (30 out of 52) presented these DNA changes which could be used to identify tumor-specific cfDNA [35,36]. Nevertheless, tumor-specific cfDNA identified by somatic alterations in individual patients seems to be a sensitive biomarker of tumor burden and will be of great interest in the near future.

Frequent cfDNA modifications with diagnostic potential are DNA methylation changes [37]. Genes encoding tumor suppressors and other cancer-relevant regulators are often deactivated by DNA methylation in the promoter region (so-called "epigenetic silencing") [38]. These DNA methylations occur at CpG dinucleotides, and correlations with cancer subtypes and other prognostic factors have been shown [39]. The detection of tumor-specific methylation changes in cfDNA offers new opportunities for cancer diagnostics.

The first studies in breast cancer patients showed that *RASSF1A* promoter methylation of cfDNA can predict the presence of breast cancer with a sensitivity of 15–75%. *APC* (adenomatous polyposis coli) promoter methylation, which is a well-established marker in colorectal cancer, showed a sensitivity of 2–47% in breast cancer patients [40–42]. Methylation changes in carcinogenesis are often heterogeneous and therefore gene panels rather than single genes are promising for diagnostic purposes. Our group examined promoter methylations of seven putative tumor-suppressor genes (*SFRP1*, *SFRP2*, *SFRP5*, *ITIH5*, *WIF1*, *DKK3*, and *RASSF1A*) in cfDNA extracted from serum of a test set ($n = 261$) and an independent validation set ($n = 343$). These genes have previously shown cancer-specific methylation in breast tissue. We identified promoter methylations of *ITIH5* and *DKK3* as candidate biomarkers for breast cancer detection. *ITIH5* and *DKK3* methylations achieved 41% sensitivity with a specificity of 93%. Combination of these genes with *RASSF1A* methylation further increased the sensitivity, suggesting that tumor-specific methylation of the three-gene panel (*ITIH5*, *DKK3*, and *RASSF1A*) might be a valuable biomarker for the early detection of breast cancer [43].

6.2.2 Long Intergenic Non-Coding RNA

Long intergenic non-coding RNAs (lincRNAs or lncRNAs) represent highly abundant RNAs, which are not translated into proteins. However, the transcription, RNA length, and splicing of lncRNAs are similar to mRNA, but lncRNAs do not contain open reading frames. The interaction of lncRNAs and mRNA with chromatin regulatory proteins is also comparable [44,45]. One key role of lncRNAs is probably the control of site (gene)-specific chromatin-modifying complexes in order to implement epigenetic changes and reprogram the chromatin state [46,47]. It is presumed that more than one-third of human lncRNAs bind to the polycomb repressive complex 2 (PRC2) or the chromatin-modifying proteins CoREST and SMCX [48]. In contrast, other lncRNAs bind to trithorax chromatin-activating complexes and/or activated chromatin [49].

The latest human lncRNA annotation by the GENCODE consortium within the framework of the ENCODE project detected 14 880 lncRNAs from 9277 genes [50]. They also found that lncRNAs had similar histone modification profiles, splicing signals, and equal exon/intron lengths, but mostly only two exons. Interestingly, recent ribosome profiling analyses found lncRNA, but also other ncRNAs, such as small nuclear RNA (snRNA) and microRNA (miRNA) precursors, bound to ribosomes with to date unknown biological implications [51,52]. However, the time of ribosome binding differed between non-coding and protein-coding RNAs [52]. The lncRNAs are responsible for fundamental biological processes and have a high impact on evolution and disease. Latest analysis showed that lncRNA and ncRNA could even promote chromosomal rearrangements [53]. Early lncRNAs, like *XIST* (X chromosome inactivation) and *H19* (*IGF2* imprinting) were identified and functionally demonstrated [54,55]. New approaches on the basis of chromatin structure, like methylated histones, revealed over 1600 long multi-exonic RNAs in mice [56].

Recently, a database (LncRNADisease; http://www.cuilab.cn/lncrnadisease) was established listing lncRNAs involved in 166 diseases, including breast cancer [57]. Although more and more lncRNAs are detected, and play an essential role in disease and development, especially for mammary development and breast cancer, the functional regulation of lncRNAs is only limited [58–60] (Table 6.1).

6.2.2.1 HOTAIR

HOTAIR (HOX transcript antisense RNA), also called *HOXAS* or *HOXC-AS*, is a lncRNA (NCRNA00072) on chromosome 12q13.13 within the homeobox C (*HOXC*) gene cluster, between *HOXC10* and *HOXC12* overlapping with *HOXC11*. This lncRNA is coexpressed with *HOXC* genes and represses the transcription of *HOXD* genes *in trans*. *HOTAIR* interacts and assembles with *PRC2* and *LSD1* (lysine-specific demethylase 1) complexes at the *HOXD* gene cluster [62,71]. Like many lncRNAs, *HOTAIR* is involved in epigenetic silencing of, for example, *HOXD* genes through H3K27 methylation and H3K4 demethylation. Interestingly, the methyltransferase EZH2, which methylates H3K27 and

Table 6.1 Examples of lncRNAs correlated with breast cancer.

Gene	Class	Function	Reference
HOTAIR	promotes metastasis	binds PRC2 and LSD1	[61,62]
H19	oncogenic; tumor suppressive	promotes cell growth and proliferation; activated by c-MYC; downregulated by prolonged cell proliferation	[63,64]
GAS5	tumor suppressive	induces apoptosis and growth arrest; binds glucocorticoid receptor and prevents glucocorticoid receptor-induced gene expression	[65,66]
LSINCT5	oncogenic	regulates cell proliferation	[67]
LOC554202	tumor suppressive	promotor hypermethylation; correlates with miR-31; down in triple-negative breast cancer	[68]
XIST	tumor suppressive	loss of XIST associated with basal-like breast cancer	[69]
SRA1 (nc)	oncogenic	activator of nuclear receptors	[70]

transcriptionally silent genes, interacts with *HOTAIR* and *XIST* [72]. Furthermore, BRCA1 was found to interact with EZH2 in breast cancer cells with the effect that BRCA1 inhibited the binding of EZH2 to HOTAIR, and loss of BRCA1 resulted in higher cancer cell migration and invasion [73]. *HOTAIR* is upregulated in breast cancer (and hepatocellular cancer), and is an independent predictor of overall survival and progression-free survival [61,74]. *HOTAIR* is thought to mediate epigenetic repression of *PRC2* target genes, by recruitment of PR2C and changing the chromatin structure through H3K27 and H3K4 methylation/demethylation [61]. *In situ* hybridizations showed a correlation of induced *HOTAIR* with positive estrogen receptor (ER) and progesterone receptor (PR) status [75].

6.2.2.2 H19

H19 is a paternally imprinted gene on chromosome 11p15.5 near the *IGF2* gene (also paternally imprinted). *H19* is a lncRNA and highly expressed in fetal tissues, but reduced in expression in most tissues after birth. Over 72% of breast adenocarcinomas showed increased *H19* gene expression compared with healthy controls, especially in stromal cells of tumors [76]. In addition, c-*MYC* expression associated with breast carcinomas and correlated directly with the *H19* expression. c-*MYC* downregulated the expression of *H19* and also *IGF2* [64]. Another regulation of *H19* was found dependent on hypoxia and p53 status. *H19* expression was induced under hypoxia, in the presence of HIF-1a and mutated p53 [63]. However, the function of *H19* as a tumor suppressor or oncogene, its regulation of possible target genes or overall chromatin, and especially the role of produced miRNAs haves to be further clarified [77].

6.2.2.3 GAS5

GAS5 (growth arrest-specific 5) is a lncRNA (NCRNA00030) on chromosome 1q25.1 overlapping with *GAS5* antisense and various *SNORD* genes. Some breast cancer cells showed growth arrest and apoptosis after *GAS5* induction, and human breast cancer samples showed reduced *GAS5* expression [66]. There are hints that *GAS5* binds to the DNA-binding domain of the glucocorticoid receptor, thus competing for DNA binding [65].

6.2.2.4 LSINCT5

LSINCT5 (long stress-induced non-coding transcript 5) on chromosome 5p15.33 is a 2.6-kb polyadenylated transcript localized in the nucleus. *LSINCT5* was shown to have increased expression in various breast (and ovarian) cancer cell lines and in 14 primary breast tumors. Inhibition of *LSINCT5* in cancer cells resulted in a decrease of proliferation [67].

6.2.2.5 LOC554202

The lncRNA LOC554202 is located on chromosome 9p21.3. Interestingly, the well-known miR-31, a major contributor to breast cancer metastasis, is transcribed from the first intron of LOC554202. miR-31 has a key role in the progression and metastasis of breast cancer [78,79]. LOC554202 and miR-31 were shown to be downregulated in triple-negative breast cancer [68].

6.2.2.6 SRA1

SRA1 (steroid receptor RNA activator 1) on chromosome 5q31.3 is present as protein coding and non-coding transcripts. The lncRNA *SRA1* functioned as a coactivator of several nuclear receptors, but also regulated non-nuclear receptor pathways. The coding *SRA1* interacted with the lncRNA *SRA1* transcripts and acted as a transcriptional repressor, but the *SRA1* protein was also found as a transcriptional repressor with other transcription factors [80]. SRA was found upregulated in breast tumorigenesis [81]. The lncRNA *SRA1* harbors a differentially spliced intron 1 [70]. Protein-coding and lncRNA *SRA1* expression were detected in breast cancer; however, significantly higher lncRNA *SRA1* was found in breast tumors with high PRs [70].

6.2.2.7 XIST

XIST (X-inactive specific transcript) is a lncRNA on chromosome Xq13.2. *XIST* is transcribed solely from the X inactivation center of the silenced X chromosome and is essential for the inactivation of transcription. A natural antisense transcript (NAT) of *XIST* has been detected and has been called *XIST-AS* or *TSIX*. *XIST* expression is epigenetically regulated [82,83]. *XIST* expression was found downregulated in female breast, ovarian, and cervical cancer cell lines [84]. Over 50% of basic-like breast cancer showed duplication of the active X chromosome, and loss of the inactive X chromosome and *XIST* expression. Interestingly, only some X-chromosomal genes were overexpressed in these cancers [69]. Predictions of *XIST* expression were made, like low *XIST* expression of

BRCA1$^-$ and p53$^-$ breast tumors correlated with high cisplatin sensitivity and benefit of platinum-based chemotherapy [85]. BRCA1 was found associated with *XIST* RNA, and played an important role in inactivating the X chromosome and stabilizing it [86]. However, a later analysis disputed this correlation and found no functional role of BRCA1 for *XIST* regulation or inactivation of the X chromosome [87]. Further analyses showed that BRCA1 was not involved in *XIST* localization, but BRCA1 presumably regulated *XIST* expression [83].

6.2.3
Natural Antisense Transcripts

A large number of antisense transcripts have been detected and it has been proposed that regulation of gene expression by double-stranded RNA could be a general mechanism in mammalian cells [88]. Sense and antisense transcripts can encode for proteins or be non-coding, although most natural antisense transcripts (NATs) are non-coding. NATs play an essential role in the regulation of transcription, expression, and development. Two different NATs can be defined: *cis*-NATs, transcribed from opposite DNA strands, but the same chromosome loci, and *trans*-NATs, transcribed from separate loci. According to their sequence complementarity, *cis*-NATs have a perfect match, but different orientation and overlap: fully overlapping (enclosed), head-to-head (5′ to 5′ = divergent), or tail-to-tail (3′ to 3′ = convergent); *trans*-NATs display only short or incomplete complementarities, but could therefore target many sense transcripts [89] (Figure 6.1). Transcriptional clusters by aligning expressed sequence tags (ESTs) and cDNAs yielded up to 22% of sense–antisense transcripts (5880 of 26 741 transcript clusters) [90]. Another analysis found a lower amount of sense–antisense transcripts, but different clusters than Chen *et al.* [90], so that the true percentage of NATs is probably even higher [91]. However, the detection of NATs is not uniform or random, but is localized to specific genes or regions of genes. It has been proposed that antisense transcripts initiate or terminate near

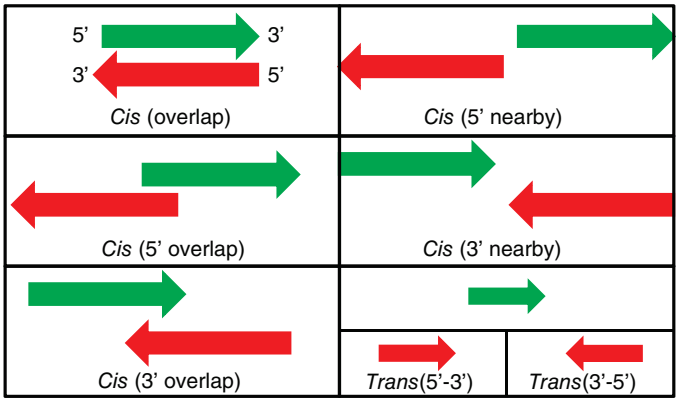


Figure 6.1 Possible structures of *cis*- and *trans*-NATs.

the corresponding promotors and terminators of the sense transcript [92]. Variable expression levels have been detected for sense and antisense transcripts in different tissues. An analysis of five human cell lines also showed that the expression of NATs did always depend on the expression of the sense transcript [92]. The regulatory impact of NATs for the sense transcript can be contrary – a knockdown of NATs can result in a decrease of the sense transcript (concordant) or even in an increase (discordant). The regulation of gene expression can be through transcriptional interference or post-transcriptional mechanisms [93]:

1) Inhibition of sense transcription by antisense transcription (transcription collision model).
2) Antisense transcription can induce DNA methylation or modify the chromatin so that the sense transcription is inhibited.
3) Formation of nuclear RNA duplexes (sense and antisense) inhibiting sense RNA processing, RNA stability, and/or splicing of sense RNA.
4) RNA duplexes could induce RNA editing leading to degradation of double-stranded RNA and/or change RNA secondary structures.
5) RNA duplexes could inhibit miRNA-binding sites and regulation.
6) DICER could process duplex RNA to siRNAs resulting in transcript regulation.
7) Cytoplasmic RNA duplex inhibit directly the translation.

An expression study of sense and antisense transcripts using microarrays showed that more than 2500 NATs were expressed in normal breast duct luminal cells and in primary breast tumors. The authors established 257 corresponding sense–antisense transcripts and found that antisense transcription was more prevalent in breast cancer [94]. However, a study using SAGE-Seq libraries found that antisense tags were significantly lower in breast tumors than in normal samples, but the overall difference in sense–antisense ratio could have been due to increased sense transcripts in tumors [95].

The inhibition of tumor suppressor genes through NATs, like p15 in leukemia, which act in *cis* and *trans* over heterochromatin formation, is crucial for carcinomas [96]. Using four breast cancer cell lines (MDA-MB-231, MDA-MB-435, triple-negative Hs578 T, and the HER2-overexpressing SKBR3) showed the expression of NATs for the *BRMS1* metastasis suppressor gene and the *RAR-β2* and *CST6* human tumor suppressor gene promoters [97].

Important NATs in breast cancer are listed in Table 6.2.

6.2.3.1 HIF-1a-AS

HIF-1a-AS (aHIF) was detected overexpressed in renal carcinoma and complementary to the 3′ untranslated region of *HIF-1a* [103]. Using 110 invasive breast cancers, Cayre *et al.* [98] found a significant positive association between *HIF-1a* and *HIF-1a-AS*, and an inverse correlation between *HIF-1a* and *HIF-1a-AS* expression and poor disease-free survival. *HIF-1a-AS* was therefore a marker for poor prognosis of breast cancer.

Table 6.2 Some NATs that play an important role in breast cancer.

NATs	Tissues	Function	Reference
HIF-1a-AS	normal breast and breast tumors	negative regulation of HIF-1a	[98]
H19-AS	breast tumors	regulates IGF2 expression *in trans*	[99]
SLC22A18-AS	normal breast and breast tumors		[100]
RPS6KA2-AS	breast cancer cell lines and ER-positive breast cancer patients	negative regulation of RPS6 KA2	[101]
ZFAS1	normal breast and downregulated in breast cancer	development, putative tumor suppressor	[102]

6.2.3.2 H19 and H19-AS (91H)

The *H19*/*IGF2* locus on chromosome 11p15.5 is imprinted with an expression of *H19* from the maternal allele and *IGF2* from the paternal alleles. *H19* is an lncRNA (see above) and over 72% of breast carcinomas showed increased *H19* expression compared with controls, especially in stromal cells, and correlated with the presence of hormone receptors [76]. *In vitro* studies with breast cancer cell lines showed that *H19* had a positive influence on anchorage-independent growth assays and promoted tumor progression in SCID mice [104]. A 120-kb transcript (91 H) was detected antisense to *H19* and maternally expressed. Interestingly, in breast cancer cells and breast tumors the *H19-AS* (91 H) was stabilized and overexpressed. A regulation of *IGF2* expression through *H19* and *H19-AS in trans* was discussed [99] and shown in mice [105].

6.2.3.3 SLC22A18-AS

The solute carrier 22-member 18 at chromosome 11p15.5 is a NAT to the *SLC18A18* and overlap with the 5′ regions. Different imprinting was found in malignant and benign breast tissues (*SLC22A18-AS* was 36% heterozygote in malignant and 54% heterozygote in benign breast tissues) [100].

6.2.3.4 RPS6KA2-AS

The ribosomal protein S6 kinase, 90 kDa, polypeptide 2 (*RPS6KA2*) on chromosome 6q27 is a member of the ribosomal S6 kinases, a serine/threonine kinase. Phosphorylating various proteins, *RPS6KA2* has been implicated in controlling cell growth and differentiation. An antisense transcript of *RPS6KA2* was assigned to chromosome 6q27 (*cis*-NAT) as well as to 15q25.2 (*trans*-NAT). An analysis of the sense/antisense transcript ratio of *RPS6KA2* showed a significant inverse correlation in 42 of 59 ER-positive breast cancer patients [101].

6.2.3.5 ZFAS1

ZNFX1 antisense RNA 1 (*ZFAS1*) is located on chromosome 20q13.13. *ZFAS1* is transcribed antisense to the 5′ end of *ZNFX1*, and the *ZFAS1* transcripts were located in ducts and alveoli of the mammary gland. Knockdown of the murine

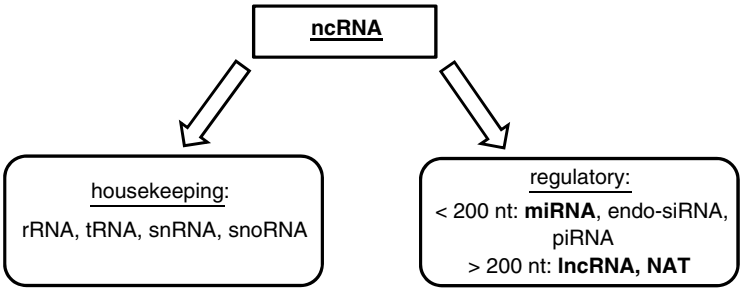


Figure 6.2 The division and differentiation of ncRNAs into housekeeping (ribosomal, transfer, small nuclear, and small nucleolar) ncRNAs and regulatory (micro, endogenous small interfering, PIWI-interacting, and long intergenic non-coding) ncRNAs.

Zfas1 resulted in increased proliferation and differentiation of murine mammary cells [102].

6.2.4 miRNAs

miRNAs are a class of short, single-stranded, ncRNA molecules of 17–24 nucleotides in length (Figure 6.2). These RNAs post-transcriptionally modulate gene expression by either translational repression or degradation of target messenger RNAs (mRNAs) [106–109]. Due to their important regulatory functions, miRNAs modulate a number of cellular processes, such as cell differentiation, growth, and apoptosis. These highly conserved endogenous RNAs have been implicated in cancer onset and progression [108–111]. More than half of the human miRNA-encoding loci reside in fragile chromosomal regions and their expression is often altered in human cancer [112,113].

6.2.4.1 Tissue-Based miRNA Profiling in Breast Cancer

The first evidence for miRNA expression differences between breast cancer tissue and normal breast tissue was found by Iorio *et al.*, identifying significant expression differences for miR-145 and miR-155 [114]. Since then, aberrant miRNA expression profiles in different cancer subtypes have been characterized and the function of miRNAs as tumor suppressors or oncogenes (oncomiRs) has been established.

Many recent studies identified tissue-based miRNA profiles correlating with breast cancer subtype, histological grade, hormone receptor status, and so on [115–117].

Examples of Important Breast Cancer-Related Tumor Suppressor miRNAs

miR-200 Family The miR-200 family consists of five members: miR-200a, b, c, miR-141, and miR-429. It is assumed that this miRNA family regulates genes

involved in breast cancer initiation, aggressiveness, and metastasis [118]. This is presumably mediated by induction of the epithelial phenotype and suppression of epithelial-to-mesenchymal transition (EMT) by miR-200 family members [119]. The loss of the tumor suppressor function results in malignant transformation, clonal expansion, as well as invasion and metastasis by increased tumor cell mobility [120,121].

miR-125 Another important miRNA for breast cancer is miR-125 with two known isoforms in humans (miR-125a and miR-125b), which are both significantly downregulated in breast cancer patients and in HER2-overexpressing breast cancer tissues [122–124]. Overexpression of miR-125a resulted in suppression of cell growth and reduced cell migration, while miR-125b was upregulated in paclitaxel-resistant breast cancer cells *in vitro* [122,125].

Lethal-7 Family The Lethal-7 (let-7) family, consisting of 12 members, was one of the first miRNA families identified in *Caenorhabditis elegans* [106]. The precise role of let-7 in tumorigenesis is still a matter of debate, but it is assumed that let-7 levels are reduced in invasive cancer cells compared with normal cells. Abrogation of let-7 g expression in non-metastatic breast cancer cells resulted in rapid metastasis *in vitro* and *in vivo* [126]. Let-7 correlated with differentiation and mesenchymal phenotype *in vitro*, suggesting that loss of let-7 expression could be used as a biomarker for aggressive, less-differentiated cancers with poor outcome [127,128]. A stimulating effect of estradiol (E_2) on several let-7 family members and regulation of ERα signaling by let-7 in hormone receptor-positive breast cancer cells *in vitro* connected let-7 to hormone receptor signaling [117,129]. Overexpression of let-7 family members led to a downregulation of ERα activity, reduced cell proliferation, and induced apoptosis in hormone receptor-positive breast cancer cells.

miR-205 miR-205 has been shown to be significantly underexpressed in breast tumor tissue compared with matched normal breast tissue. In addition, miR-205 suppressed the aggressive phenotype in triple-negative breast cancer cells *in vitro* [130]. Similar results were found in the comparison of breast cancer cell lines with non-malignant cell culture cells [131]. Moreover, in the latter study ErbB3 (HER3) and vascular endothelial growth factor A (VEGF-A) were identified as direct targets for miR-205. The regulation of HER3, one of the four structurally related HER receptors of the epidermal growth factor receptor (EGFR) family, by miR-205 was confirmed in a second study [132].

miR-34a Kato *et al.* investigated the effect of DNA-damaging agents on miR-34a expression, and found that miR-34 expression was upregulated by p53 and required for a normal cellular response to DNA damage *in vivo* [133]. This miRNA was also downregulated in many different breast cancer cell lines and modulated chemosensitivity. Due to the inhibition of proliferation and migration of breast cancer cells, miR-34a functions as a metastasis suppressor [134–136].

Serum concentrations of circulating miR-34a were shown to be significantly different between early-stage (adjuvant) breast cancer patients and healthy individuals. Similar differences in serum concentrations have also been shown for miR-93 and miR-373 in the same study [137].

Examples of Important Breast Cancer-Related Oncogenic miRNAs

miR-21 Expression profiling of miRNAs in breast cancer tissue and matched normal breast tissue revealed that miR-21 was highly overexpressed in tumors and could be an indicator of an aggressive breast cancer phenotype. In addition, miR-21 suppressed tumor cell growth *in vitro* and in xenograft models [138,139]. Important tumor suppressors like PTEN (phosphatase and tensin homolog) and PDCD4 (programmed cell death-4) have been identified as targets of miR-21 [140,141]. Transforming growth factor (TGF)-β upregulated the expression of miR-21, whose expression has been found to be associated with high tumor grade, negative hormone receptor status, and poor disease-free survival in early-stage breast cancer patients [142,143]. A small study also suggested miR-21 and miR-146a as potential plasma-based diagnostic biomarkers in breast cancer [144].

miR-10b and miR-520/373 Family miR-10b was found increased in lymphatic metastatic tissues, and could be a good candidate for the prediction of breast cancer metastasis and treatment response [145–147]. A first, small pilot study also suggested that circulating miR-10b and miR-373 could be potential blood-based biomarkers for the preoperative detection of breast cancer lymph node status [146]. Decreased expression of miR-520c correlated with lymph node metastasis in ER-negative tumors and the miR-520/373 family seems to have a metastasis-suppressive role in breast cancer [148]. The relative serum concentrations of miR-373, miR-34a, and miR-93 showed significant differences between early-stage breast cancer patients and healthy women [137].

miR-155 Functional *in vitro* analyses identified miR-155 as a TGF-β-regulated suppressor of apoptosis effecting cell migration, differentiation, angiogenesis through targeting of von Hippel–Lindau, proliferation, chemosensitivity, invasion, and EMT [149–152]. The expression of miR-155 has been shown to be induced by the loss of p53 function, connecting this miRNA with breast cancer subtypes bearing a high p53 mutation rate, such as triple-negative breast cancer [153]. Different studies found upregulation of miR-155 in the majority of analyzed primary breast cancer tumors, and this miRNA was also correlated with higher tumor grade, advanced tumor stage, lymph node metastasis, and poor prognosis [150,154,155].

miR-155 was one of the first miRNAs suspected as a potential serum-based biomarker for breast cancer detection [156]. This miRNA showed a fivefold overexpression in breast cancer tissue compared with non-cancerous breast tissue, a three- to fivefold higher plasma and serum level in breast cancer patients,

and was inversely correlated with ER and PR expression [157,158]. Interestingly, decreased levels of serum miR-155 were found after breast cancer surgery, underlining the potential role of miR-155 as a non-invasive breast cancer biomarker [157]. Significant differences in serum concentrations of miR-155 and miR-17 have also been shown between early-stage (M0) and metastasized (M1) breast cancer patients [137].

6.2.4.2 Circulating miRNAs

The origin of circulating miRNAs and their precise functional role in the bloodstream is still not completely understood. Nevertheless, circulating miRNAs represent an easy accessible biomarker derived from blood or other body fluids with a potential role in non-invasive breast cancer detection. As miRNAs are bound to proteins like Argonaute 2 (Ago2) or lipoproteins and encapsulated in vesicles like exosomes, they are highly resistant to RNase activity and remarkably stable in the bloodstream [152,159–162]. Even repeated freeze–thawing cycles and boiling do not markedly alter expression levels of endogenous miRNAs [161,163,164]. The first pilot studies analyzed different panels of candidate miRNA in serum, plasma or whole blood of breast cancer patients and controls. Zhu *et al.* performed the first study analyzing miRNAs in serum and demonstrated a higher miR-155 expression in PR-positive tumors [156]. In another of these initial studies, miR-155, miR-10b, and miR-34a were found to be overexpressed in the blood circulation of breast cancer patients compared with healthy individuals [165]. Thereby, the single circulating serum miRNA miR-155 significantly distinguished primary breast cancer patients from healthy individuals [165]. Recently, miR-155 expression in the serum of breast cancer patients was found to be elevated, and positively correlated with clinical stage, molecular subtype, and Ki-67 (Mib-1), as well as negatively correlated with p53 status [166].

Another study using a candidate gene approach found significantly higher levels of miR-195 and let-7a in whole blood of breast cancer patients compared with controls [167]. This study analyzed the levels of seven circulating candidate miRNAs (in 83 breast cancer cases and 44 controls), and found correlations with nodal status and ER status. Moreover, the expression level of miR-195 decreased post-operatively after complete breast cancer resection to levels comparable with controls [167].

The analyses of 10 candidate miRNAs in 100 serum samples and 48 tissue samples of patients with primary breast cancer revealed significant lower miR-92a expression levels and higher miR-21 levels in tissue and serum of breast cancer patients compared with healthy individuals. These two miRNAs were also associated with tumor size and positive lymph node status in this study [168]. These data confirmed a previous study investigating the diagnostic potential of miR-21 in the serum of breast cancer patients showing significant expression differences compared with healthy controls and a correlation of high circulating miR-21 levels with visceral metastasis [169]. Zhao *et al.* used microarray-based expression profiling followed by quantitative real-time polymerase chain reaction (qRT-PCR) for the comparison of plasma levels of circulating miRNAs [170].

They analyzed 20 early-stage breast cancer patients and 20 healthy individuals, subdivided into Caucasian-Americans and African-Americans, and found marked expression differences between the two ethnic groups. In the Caucasian-American subgroup, 31 miRNAs were differentially expressed between cases and controls, while in the African-American subgroup, only 18 miRNAs showed significant expression differences with at least twofold expression changes. Different candidate miRNAs like let-7c (for the Caucasian-American subgroup) were successfully validated by qRT-PCR in an independent set of 30 participants. Let-7c predicted disease onset with an area under the curve (AUC) of 0.99 in African-Americans, while the AUC value in the Caucasian-Americans cohort was only 0.78 [170]. The best predictor for the Caucasian-American subgroup was miR-589 with an AUC of 0.85; however, the number of patients was very small in this study [170].

In our own study we performed microarray-based miRNA profiling on whole-blood samples of 48 early-stage breast cancer patients at diagnosis and used 57 healthy individuals as controls. This initial "screening step" with a microarray panel of 1100 miRNAs identified 59 miRNAs with significantly different expression between breast cancer patients and controls (13 upregulated and 46 downregulated). Using a set of 240 miRNAs, we were able to differentiate breast cancer patients from healthy individuals with a specificity of 78.8% and a sensitivity of 92.5%. The accuracy of this miRNA set was 85.6% and from this set two miRNAs were validated by qRT-PCR in a completely independent screening cohort ($n = 48$) [171].

A similar approach with an initial screening and further validation cohort was used by two other studies. Cuk *et al.* identified a panel of three miRNAs (miR-148b, miR-409-3p, and miR-801) in plasma of early-stage breast cancer patients with a reasonable discriminatory power (AUC = 0.69) [172]. Chan *et al.* analyzed paired breast cancer tissue and serum samples in a screening cohort of 32 Asian-Chinese breast cancer patients and validated a panel of four miRNAs (miR-1, miR-92a, miR-133a, and miR-133b) with an independent set of 233 serum samples (breast cancer patients and controls) [173]. Similar to our results, they found that the majority ($n = 13$) of the 20 most significant differentially expressed miRNAs were downregulated. Moreover, the miRNA profiles between serum and the corresponding tumor tissue were largely dissimilar, suggesting a selective release mechanism for miRNAs. Nevertheless, receiver operating characteristic curve analyses revealed promising results using different miRNA combinations of the miRNAs named above (AUC = 0.9) [173].

6.3 Molecular Breast Cancer Subtypes and Prognostic/Predictive Molecular Biomarkers

Molecular analysis of the tumor tissue is one of the most accepted methods for retrieving prognostic information about the natural course of breast cancer. The assessment of ER status, PR status, nuclear grading, and HER2 status within the

tumor tissue are examples of well-established prognostic factors that are important for future treatments [2].

Recent molecular profiling studies have shown that not only these established factors but also the expression levels of a variety of other genes characterize different molecular subtypes of breast cancer with therapeutic implications. Gene expression-based classification of breast cancer into "intrinsic subtypes" has confirmed the concept of breast cancer as a heterogeneous disease with individualized treatment strategies [174,175]. These molecular intrinsic breast cancer subtypes (luminal A, luminal B, HER2-enriched = *ERBB2*+, and basal-like) are similar but not identical to histopathological classifications adapted from analyzes of hormone receptor status, HER2 status, and the proliferation factor Ki-67 (Mib-1) [176]. The luminal A subtype is characterized by high expression levels of ER-related genes, while luminal B tumors possess higher levels of genes associated with proliferation [174,175]. Genes of the HER2 signaling pathway dominate the HER2-enriched (*ERBB2*+) subtype, while the basal-like subtype shows strong overlap with a triple-negative breast cancer subtype (negative for ER, PR, and HER2) [177]. Classifiers for intrinsic subtyping like PAM50 (50-gene subtype predictor) were designed as supervised risk predictors of breast cancer using a standardized gene set to identify the molecular subtype with a robust and reliable assay [178]. Other current tests to determine intrinsic subtypes are the MammaTyper (STRATIFYER) and BluePrint (Agendia).

Prognostic and/or predictive tissue-based multigene assays utilize fresh frozen tumor tissue (Amsterdam 70-gene breast cancer gene signature; MammaPrint; Agendia) or formalin-fixed paraffin-embedded (FFPE) tissue (21-gene recurrence score; Oncotype DX breast cancer assay; Genomic Health) and have recently been approved by the US Food and Drug Administration [179]. These tests have emerged as promising prognostic markers of local and distant recurrence, and are currently being investigated in prospective clinical trials (MINDACT and TAILORx Trial) [180]. The MammaPrint test can be used for breast cancer patients with hormone receptor-positive and -negative tumors with up to three involved lymph nodes, while the Oncotype DX breast cancer assay is approved for ER-positive, node-negative tumors. More recently, the EndoPredict assay (Sividon Diagnostics) was launched to divide the group of ER-positive, HER2-negative breast cancer patients into high- and low-risk subgroups. The Genomic Grade Index (GGI) intends to subdivide the broad group of tumors with intermediate histological grading (G2) using a 97-gene expression signature. This reclassification was shown to be predictive of recurrence in endocrine-treated patients and associated with response to chemotherapy [181,182].

Expression profiles and levels of circulating miRNAs were also reported to be associated with breast cancer treatment response. Multivariate analyses, corrected for the traditional predictive factors, showed that higher expression levels of miR-30c were associated with a benefit to anti-hormonal breast cancer treatment with tamoxifen and with longer progression-free survival [183]. miRNA-342, which is known to be associated with ERα expression, contributed to tamoxifen resistance and could be used in combination with other miRNAs for

the future prediction of anti-hormonal treatment response [184,185]. High levels of miR-125b were associated with poor response to paclitaxel-based breast cancer treatment *in vitro*, while miR-205 expression was connected to the responsiveness of tyrosine kinase inhibitors [125,132]. Another study using a candidate miRNA approach with four miRNA (miR-10b, miR-34a, miR-125b, and miR-155) was able to correlate miR-125b serum levels to neoadjuvant chemotherapy response in advanced breast cancer [186]. Single miRNAs, such as miR-22, miR-126, miR-155, miR-200b, and miR-335, as well as groups of miRNAs were identified as promising prognostic markers for progression-free, distant metastasis-free, and overall survival in breast cancer patients [155,187–189].

Despite these first encouraging results of miRNAs as prognostic and predictive biomarkers, this area is dominated by the aforementioned tissue-based mRNA expression profiling assays already commercially available. However, the concept of circulating miRNAs as potential biomarkers for breast cancer detection seems to be one of the very promising applications for these ncRNA molecules.

References

1 Youlden, D.R., Cramb, S.M., Dunn, N.A., Muller, J.M., Pyke, C.M., and Baade, P.D. (2012) The descriptive epidemiology of female breast cancer: an international comparison of screening, incidence, survival and mortality. *Cancer Epidemiology*, **36**, 237–248.

2 Fasching, P.A., Lux, M.P., Beckmann, K., Strick, R., and Beckmann, M.W. (2005) Prognose- und Prädiktivfaktoren – Entscheidungshilfen bei der Therapiewahl für Patientinnen mit Mammakarzinom. *Gynäkologe*, **38**, 388–397.

3 Early Breast Cancer Trialists' Collaborative, Group (1998) Tamoxifen for early breast cancer: an overview of the randomised trials. *Lancet*, **351**, 1451–1467.

4 Early Breast Cancer Trialists' Collaborative, Group (2005) Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet*, **365**, 1687–1717.

5 Tabar, L., Duffy, S.W., Vitak, B., Chen, H. H., and Prevost, T.C. (1999) The natural history of breast carcinoma: what have we learned from screening? *Cancer*, **86**, 449–462.

6 Tabar, L., Smith, R.A., and Duffy, S.W. (2002) Update on effects of screening mammography. *Lancet*, **360**, 337, author reply 9–40.

7 Gotzsche, P.C. and Nielsen, M. (2011) Screening for breast cancer with mammography. *Cochrane Database of Systematic Reviews*, **19**, CD001877.

8 deKoning, H.J. (2003) Mammographic screening: evidence from randomised controlled trials. *Annals of Oncology*, **14**, 1185–1189.

9 Allgood, P.C., Warwick, J., Warren, R.M., Day, N.E., and Duffy, S.W. (2008) A case-control study of the impact of the East Anglian breast screening programme on breast cancer mortality. *British Journal of Cancer*, **98**, 206–209.

10 Taplin, S., Abraham, L., Barlow, W.E., Fenton, J.J., Berns, E.A., Carney, P.A. *et al.* (2008) Mammography facility characteristics associated with interpretive accuracy of screening mammography. *Journal of the National Cancer Institute*, **100**, 876–887.

11 deGelder, R., Heijnsdijk, E.A., vanRavesteyn, N.T., Fracheboud, J., Draisma, G., and deKoning, H.J. (2011) Interpreting overdiagnosis estimates in population-based mammography

screening. *Epidemiologic Reviews*, **33**, 111–121.
12 vanGils, C.H., Otten, J.D., Verbeek, A.L., Hendriks, J.H., and Holland, R. (1998) Effect of mammographic breast density on breast cancer screening performance: a study in Nijmegen, The Netherlands. *Journal of Epidemiology and Community Health*, **52**, 267–271.
13 Pinsky, R.W. and Helvie, M.A. (2010) Mammographic breast density: effect on imaging and breast cancer risk. *Journal of the National Comprehensive Cancer Network*, **8**, 1157–1164; quiz 65.
14 Boyd, N.F., Guo, H., Martin, L.J., Sun, L., Stone, J., Fishell, E. *et al.* (2007) Mammographic density and the risk and detection of breast cancer. *New England Journal of Medicine*, **356**, 227–236.
15 Broca, P. (1866) *Traite Des Tumers*, Asselin, Paris.
16 Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P.A., Harshman, K., Tavtigian, S. *et al.* (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*, **266**, 66–71.
17 Wooster, R., Neuhausen, S.L., Mangion, J., Quirk, Y., Ford, D., Collins, N. *et al.* (1994) Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12–13. *Science*, **265**, 2088–2090.
18 Corsetti, V., Houssami, N., Ghirardi, M., Ferrari, A., Speziani, M., Bellarosa, S. *et al.* (2011) Evidence of the effect of adjunct ultrasound screening in women with mammography-negative dense breasts: interval breast cancers at 1 year follow-up. *European Journal of Cancer*, **47**, 1021–1026.
19 Michell, M., Wasan, R., Whelehan, P., Iqbal, A., Lawinski, C., Donaldson, A. *et al.* (2009) Digital breast tomosynthesis: a comparison of the accuracy of digital breast tomosynthesis, two-dimensional digital mammography and two-dimensional screening mammography (film-screen). *Breast Cancer Research*, **11** (Suppl 2), O1.
20 Hakim, C.M., Chough, D.M., Ganott, M. A., Sumkin, J.H., Zuley, M.L., and Gur, D. (2010) Digital breast tomosynthesis in the diagnostic environment: a subjective side-by-side review. *American Journal of Roentgenology*, **195**, W172–W176.
21 D'Orsi, C.J. and Newell, M.S. (2011) On the frontline of screening for breast cancer. *Seminars in Oncology*, **38**, 119–127.
22 Hooley, R.J., Andrejeva, L., and Scoutt, L. M. (2011) Breast cancer screening and problem solving using mammography, ultrasound, and magnetic resonance imaging. *Ultrasound Quarterly*, **27**, 23–47.
23 Rack, B.K., Schindlbeck, C., Schneeweiss, A., Schrader, I., Lorenz, R., Beckmann, M. *et al.* (2009) Persistance of circulating tumor cells (CTCs) in peripheral blood of breast cancer (BC) patients two years after primary diagnosis. *Journal of Clinical Oncology*, **27**, 15s.
24 Schindlbeck, C., Rack, B., Jueckstock, J., Schneeweiss, A., Thurner-Herrmanns, E., Schneider, A. *et al.* (2009) Prognostic relevance of circulating tumor cells (CTCs) in peripheral blood of breast cancer patients before and after adjuvant chemotherapy – translational research program of the German SUCCESS trial. *Cancer Research*, **69** (2 Suppl.), 303.
25 Leon, S.A., Shapiro, B., Sklaroff, D.M., and Yaros, M.J. (1977) Free DNA in the serum of cancer patients and the effect of therapy. *Cancer Research*, **37**, 646–650.
26 Stroun, M., Anker, P., Lyautey, J., Lederrey, C., and Maurice, P.A. (1987) Isolation and characterization of DNA from the plasma of cancer patients. *European Journal of Cancer & Clinical Oncology*, **23**, 707–712.
27 Catarino, R., Ferreira, M.M., Rodrigues, H., Coelho, A., Nogal, A., Sousa, A. *et al.* (2008) Quantification of free circulating tumor DNA as a diagnostic marker for breast cancer. *DNA and Cell Biology*, **27**, 415–421.
28 Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y. *et al.* (2004) Detection of large-scale variation in the human genome. *Nature Genetics*, **36**, 949–951.
29 Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P. *et al.* (2004) Large-scale copy number

polymorphism in the human genome. *Science*, **305**, 525–528.
30 Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D. *et al.* (2006) Global variation in copy number in the human genome. *Nature*, **444**, 444–454.
31 Higgins, M.J., Jelovac, D., Barnathan, E., Blair, B., Slater, S., Powers, P. *et al.* (2012) Detection of tumor PIK3CA status in metastatic breast cancer using peripheral blood. *Clinical Cancer Research*, **18**, 3462–3469.
32 Board, R.E., Wardley, A.M., Dixon, J.M., Armstrong, A.C., Howell, S., Renshaw, L. *et al.* (2010) Detection of PIK3CA mutations in circulating free DNA in patients with breast cancer. *Breast Cancer Research and Treatment*, **120**, 461–467.
33 Silva, J.M., Gonzalez, R., Dominguez, G., Garcia, J.M., Espana, P., and Bonilla, F. (1999) TP53 gene mutations in plasma DNA of cancer patients. *Genes Chromosomes Cancer*, **24**, 160–161.
34 Schwarzenbach, H., Eichelser, C., Kropidlowski, J., Janni, W., Rack, B., and Pantel, K. (2012) Loss of heterozygosity at tumor suppressor genes detectable on fractionated circulating cell-free tumor DNA as indicator of breast cancer progression. *Clinical Cancer Research*, **18**, 5719–5730.
35 Dawson, S.J., Tsui, D.W., Murtaza, M., Biggs, H., Rueda, O.M., Chin, S.F. *et al.* (2013) Analysis of circulating tumor DNA to monitor metastatic breast cancer. *New England Journal of Medicine*, **368**, 1199–1209.
36 Dawson, S.J., Rosenfeld, N., and Caldas, C. (2013) Circulating tumor DNA to monitor metastatic breast cancer. *New England Journal of Medicine*, **369**, 93–94.
37 Laird, P.W. (2003) The power and the promise of DNA methylation markers. *Nature Reviews Cancer*, **3**, 253–266.
38 Chen, W.Y. and Baylin, S.B. (2005) Inactivation of tumor suppressor genes: choice between genetic and epigenetic routes. *Cell Cycle*, **4**, 10–12.
39 Liggett, T.E., Melnikov, A.A., Marks, J.R., and Levenson, V.V. (2011) Methylation patterns in cell-free plasma DNA reflect removal of the primary tumor and drug treatment of breast cancer patients. *International Journal of Cancer*, **128**, 492–499.
40 Hoque, M.O., Feng, Q., Toure, P., Dem, A., Critchlow, C.W., Hawes, S.E. *et al.* (2006) Detection of aberrant methylation of four genes in plasma DNA for the detection of breast cancer. *Journal of Clinical Oncology*, **24**, 4262–4269.
41 Radpour, R., Barekati, Z., Kohler, C., Lv, Q., Burki, N., Diesch, C. *et al.* (2011) Hypermethylation of tumor suppressor genes involved in critical regulatory pathways for developing a blood-based test in breast cancer. *PLoS One*, **6**, e16080.
42 Brooks, J.D., Cairns, P., Shore, R.E., Klein, C.B., Wirgin, I., and Afanasyeva, Y. *et al.* (2010) DNA methylation in pre-diagnostic serum samples of breast cancer cases: results of a nested case-control study. *Cancer Epidemiology*, **34**, 717–723.
43 Kloten, V., Becker, B., Winner, K., Schrauder, M.G., Fasching, P.A., Anzeneder, T. *et al.* (2013) Promoter hypermethylation of the tumor-suppressor genes ITIH5, DKK3, and RASSF1A as novel biomarkers for blood-based breast cancer screening. *Breast Cancer Research*, **15**, R4.
44 Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G. *et al.* (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*, **477**, 295–300.
45 Guttman, M. and Rinn, J.L. (2012) Modular regulatory principles of large non-coding RNAs. *Nature*, **482**, 339–346.
46 Mattick, J.S. and Gagen, M.J. (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**, 1611–1630.
47 Mattick, J.S., Amaral, P.P., Dinger, M.E., Mercer, T.R., and Mehler, M.F. (2009) RNA regulation of epigenetic processes. *BioEssays*, **31**, 51–59.
48 Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D. *et al.* (2009) Many human large

intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 11667–11672.

49 Dinger, M.E., Amaral, P.P., Mercer, T.R., Pang, K.C., Bruce, S.J., Gardiner, B.B. *et al.* (2008) Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Research*, **18**, 1433–1445.

50 Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H. *et al.* (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Research*, **22**, 1775–1789.

51 Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, **147**, 789–802.

52 Guttman, M., Russell, P., Ingolia, N.T., Weissman, J.S., and Lander, E.S. (2013) Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell*, **154**, 240–251.

53 Brown, J.D., Mitchell, S.E., and O'Neill, R. J. (2012) Making a long story short: noncoding RNAs and chromosome change. *Heredity*, **108**, 42–49.

54 Brown, C.J., Ballabio, A., Rupert, J.L., Lafreniere, R.G., Grompe, M., Tonlorenzi, R. *et al.* (1991) A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature*, **349**, 38–44.

55 Bartolomei, M.S., Zemel, S., and Tilghman, S.M. (1991) Parental imprinting of the mouse H19 gene. *Nature*, **351**, 153–155.

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

57 Chen, G., Wang, Z., Wang, D., Qiu, C., Liu, M., Chen, X. *et al.* (2013) LncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic Acids Research*, **41**, D983–D.986.

58 Gutschner, T. and Diederichs, S. (2012) The hallmarks of cancer: a long non-coding RNA point of view. *RNA Biology*, **9**, 703–719.

59 Shore, A.N., Herschkowitz, J.I., and Rosen, J.M. (2012) Noncoding RNAs involved in mammary gland development and tumorigenesis: there's a long way to go. *Journal of Mammary Gland Biology and Neoplasia*, **17**, 43–58.

60 Piao, H.L. and Ma, L. (2012) Non-coding RNAs as regulators of mammary development and breast cancer. *Journal of Mammary Gland Biology and Neoplasia*, **17**, 33–42.

61 Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J. *et al.* (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*, **464**, 1071–1076.

62 Tsai, M.C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F. *et al.* (2010) Long noncoding RNA as modular scaffold of histone modification complexes. *Science*, **329**, 689–693.

63 Matouk, I.J., DeGroot, N., Mezan, S., Ayesh, S., Abu-lail, R., Hochberg, A. *et al.* (2007) The H19 non-coding RNA is essential for human tumor growth. *PloS one*, **2**, e845.

64 Barsyte-Lovejoy, D., Lau, S.K., Boutros, P. C., Khosravi, F., Jurisica, I., Andrulis, I.L. *et al.* (2006) The c-Myc oncogene directly induces the H19 noncoding RNA by allele-specific binding to potentiate tumorigenesis. *Cancer Research*, **66**, 5330–5337.

65 Kino, T., Hurt, D.E., Ichijo, T., Nader, N., and Chrousos, G.P. (2010) Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Science Signaling*, **3**, ra8.

66 Mourtada-Maarabouni, M., Pickard, M. R., Hedge, V.L., Farzaneh, F., and Williams, G.T. (2009) GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. *Oncogene*, **28**, 195–208.

67 Silva, J.M., Boczek, N.J., Berres, M.W., Ma, X., and Smith, D.I. (2011) LSINCT5 is over expressed in breast and ovarian cancer and affects cellular proliferation. *RNA Biology*, **8**, 496–505.

68 Augoff, K., McCue, B., Plow, E.F., and Sossey-Alaoui, K. (2012) miR-31 and its host gene lncRNA LOC554202 are regulated by promoter hypermethylation in triple-negative breast cancer. *Molecular Cancer*, **11**, 5.

69 Richardson, A.L., Wang, Z.C., De Nicolo, A., Lu, X., Brown, M., Miron, A. *et al.* (2006) X chromosomal abnormalities in basal-like human breast cancer. *Cancer Cell*, **9**, 121–132.

70 Cooper, C., Guo, J., Yan, Y., Chooniedass-Kothari, S., Hube, F., Hamedani, M.K. *et al.* (2009) Increasing the relative expression of endogenous non-coding Steroid Receptor RNA Activator (SRA) in human breast cancer cells using modified oligonucleotides. *Nucleic Acids Research*, **37**, 4518–4531.

71 Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A. *et al.* (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, **129**, 1311–1323.

72 Kaneko, S., Li, G., Son, J., Xu, C.F., Margueron, R., Neubert, T.A. *et al.* (2010) Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA. *Genes & Development*, **24**, 2615–2620.

73 Wang, L., Zeng, X., Chen, S., Ding, L., Zhong, J., Zhao, J.C. *et al.* (2013) BRCA1 is a negative modulator of the PRC2 complex. *EMBO Journal*, **32**, 1584–1597.

74 Gibb, E.A., Brown, C.J., and Lam, W.L. (2011) The functional role of long non-coding RNA in human carcinomas. *Molecular Cancer*, **10**, 38.

75 Chisholm, K.M., Wan, Y., Li, R., Montgomery, K.D., Chang, H.Y., and West, R.B. (2012) Detection of long non-coding RNA in archival tissue: correlation with polycomb protein expression in primary and metastatic breast carcinoma. *PloS One*, **7**, e47998.

76 Adriaenssens, E., Dumont, L., Lottin, S., Bolle, D., Lepretre, A., Delobelle, A. *et al.* (1998) H19 overexpression in breast adenocarcinoma stromal cells is associated with tumor values and steroid receptor status but independent of p53 and Ki-67 expression. *American Journal of Pathology*, **153**, 1597–1607.

77 Gabory, A., Jammes, H., and Dandolo, L. (2010) The H19 locus: role of an imprinted non-coding RNA in growth and development. *BioEssays*, **32**, 473–480.

78 Valastyan, S., Reinhardt, F., Benaich, N., Calogrias, D., Szasz, A.M., Wang, Z.C. *et al.* (2009) A pleiotropically acting microRNA, miR-31, inhibits breast cancer metastasis. *Cell*, **137**, 1032–1046.

79 Valastyan, S., Chang, A., Benaich, N., Reinhardt, F., and Weinberg, R.A. (2011) Activation of miR-31 function in already-established metastases elicits metastatic regression. *Genes & Development*, **25**, 646–659.

80 Chooniedass-Kothari, S., Hamedani, M. K., Auge, C., Wang, X., Carascossa, S., Yan, Y. *et al.* (2010) The steroid receptor RNA activator protein is recruited to promoter regions and acts as a transcriptional repressor. *FEBS Letters*, **584**, 2218–2224.

81 Leygue, E., Dotzlaw, H., Watson, P.H., and Murphy, L.C. (1999) Expression of the steroid receptor RNA activator in human breast tumors. *Cancer Research*, **59**, 4190–4193.

82 Hendrich, B.D., Brown, C.J., and Willard, H.F. (1993) Evolutionary conservation of possible functional domains of the human and murine XIST genes. *Human Molecular Genetics*, **2**, 663–672.

83 Sirchia, S.M., Tabano, S., Monti, L., Recalcati, M.P., Gariboldi, M., Grati, F.R. *et al.* (2009) Misbehaviour of XIST RNA in breast cancer cells. *PloS One*, **4**, e5559.

84 Kawakami, T., Zhang, C., Taniguchi, T., Kim, C.J., Okada, Y., Sugihara, H. *et al.* (2004) Characterization of loss-of-inactive X in Klinefelter syndrome and female-derived cancer cells. *Oncogene*, **23**, 6163–6169.

85 Rottenberg, S., Vollebergh, M.A., deHoon, B., deRonde, J., Schouten, P.C., Kersbergen, A. *et al.* (2012) Impact of intertumoral heterogeneity on predicting

chemotherapy response of BRCA1-deficient mammary tumors. *Cancer Research*, **72**, 2350–2361.

86 Ganesan, S., Silver, D.P., Greenberg, R.A., Avni, D., Drapkin, R., Miron, A. *et al.* (2002) BRCA1 supports XIST RNA concentration on the inactive X chromosome. *Cell*, **111**, 393–405.

87 Xiao, C., Sharp, J.A., Kawahara, M., Davalos, A.R., Difilippantonio, M.J., Hu, Y. *et al.* (2007) The XIST noncoding RNA functions independently of BRCA1 in X inactivation. *Cell*, **128**, 977–989.

88 Rosok, O. and Sioud, M. (2004) Systematic identification of sense–antisense transcripts in mammalian cells. *Nature Biotechnology*, **22**, 104–108.

89 Lapidot, M. and Pilpel, Y. (2006) Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. *EMBO Reports*, **7**, 1216–1222.

90 Chen, J., Sun, M., Kent, W.J., Huang, X., Xie, H., Wang, W. *et al.* (2004) Over 20% of human transcripts might form sense–antisense pairs. *Nucleic Acids Research*, **32**, 4812–4820.

91 Yelin, R., Dahary, D., Sorek, R., Levanon, E.Y., Goldstein, O., Shoshan, A. *et al.* (2003) Widespread occurrence of antisense transcription in the human genome. *Nature Biotechnology*, **21**, 379–386.

92 He, Y., Vogelstein, B., Velculescu, V.E., Papadopoulos, N., and Kinzler, K.W. (2008) The antisense transcriptomes of human cells. *Science*, **322**, 1855–1857.

93 Faghihi, M.A. and Wahlestedt, C. (2009) Regulatory roles of natural antisense transcripts. *Nature Reviews Molecular Cell Biology*, **10**, 637–643.

94 Grigoriadis, A., Oliver, G.R., Tanney, A., Kendrick, H., Smalley, M.J., Jat, P. *et al.* (2009) Identification of differentially expressed sense and antisense transcript pairs in breast epithelial tissues. *BMC Genomics*, **10**, 324.

95 Maruyama, R., Shipitsin, M., Choudhury, S., Wu, Z., Protopopov, A., Yao, J. *et al.* (2012) Altered antisense-to-sense transcript ratios in breast cancer. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 2820–2824.

96 Yu, W., Gius, D., Onyango, P., Muldoon-Jacobs, K., Karp, J., Feinberg, A.P. *et al.* (2008) Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature*, **451**, 202–206.

97 Tzadok, S., Caspin, Y., Hachmo, Y., Canaani, D., and Dotan, I. (2013) Directionality of noncoding human RNAs: how to avoid artifacts. *Analytical Biochemistry*, **439**, 23–29.

98 Cayre, A., Rossignol, F., Clottes, E., and Penault-Llorca, F. (2003) aHIF but not HIF-1alpha transcript is a poor prognostic marker in human breast cancer. *Breast Cancer Research*, **5**, R223–R230.

99 Berteaux, N., Aptel, N., Cathala, G., Genton, C., Coll, J., Daccache, A. *et al.* (2008) A novel H19 antisense RNA overexpressed in breast cancer contributes to paternal IGF2 expression. *Molecular and Cellular Biology*, **28**, 6731–6745.

100 Gallagher, E., Mc Goldrick, A., Chung, W.Y., Mc Cormack, O., Harrison, M., Kerin, M. *et al.* (2006) Gain of imprinting of SLC22A18 sense and antisense transcripts in human breast cancer. *Genomics*, **88**, 12–17.

101 Monti, L., Cinquetti, R., Guffanti, A., Nicassio, F., Cremona, M., Lavorgna, G. *et al.* (2009) *In silico* prediction and experimental validation of natural antisense transcripts in two cancer-associated regions of human chromosome 6. *International Journal of Oncology*, **34**, 1099–1108.

102 Askarian-Amiri, M.E., Crawford, J., French, J.D., Smart, C.E., Smith, M.A., Clark, M.B. *et al.* (2011) SNORD-host RNA Zfas1 is a regulator of mammary development and a potential marker for breast cancer. *RNA*, **17**, 878–891.

103 Thrash-Bingham, C.A. and Tartof, K.D. (1999) aHIF: a natural antisense transcript overexpressed in human renal cancer and during hypoxia. *Journal of the National Cancer Institute*, **91**, 143–151.

104 Lottin, S., Adriaenssens, E., Dupressoir, T., Berteaux, N., Montpellier, C., Coll, J. *et al.* (2002) Overexpression of an ectopic

H19 gene enhances the tumorigenic properties of breast cancer cells. *Carcinogenesis*, **23**, 1885–1895.

105 Tran, V.G., Court, F., Duputie, A., Antoine, E., Aptel, N., Milligan, L. *et al.* (2012) H19 antisense RNA can up-regulate Igf2 transcription by activation of a novel promoter in mouse myoblasts. *PloS One*, **7**, e37923.

106 Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*, **75**, 843–854.

107 Croce, C.M. and Calin, G.A. (2005) miRNAs, cancer, and stem cell division. *Cell*, **122**, 6–7.

108 Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.

109 Lowery, A.J., Miller, N., McNeill, R.E., and Kerin, M.J. (2008) MicroRNAs as prognostic indicators and therapeutic targets: potential effect on breast cancer management. *Clinical Cancer Research*, **14**, 360–365.

110 Bartels, C.L. and Tsongalis, G.J. (2009) MicroRNAs: novel biomarkers for human cancer. *Clinical Chemistry*, **55**, 623–631.

111 Iorio, M.V. and Croce, C.M. (2012) MicroRNA dysregulation in cancer: diagnostics, monitoring and therapeutics. A comprehensive review. *EMBO Molecular Medicine*, **4**, 143–159.

112 Croce, C.M. (2009) Causes and consequences of microRNA dysregulation in cancer. *Nature Reviews Genetics*, **10**, 704–714.

113 Calin, G.A., Sevignani, C., Dumitru, C.D., Hyslop, T., Noch, E., Yendamuri, S. *et al.* (2004) Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 2999–3004.

114 Iorio, M.V., Ferracin, M., Liu, C.G., Veronese, A., Spizzo, R., Sabbioni, S. *et al.* (2005) MicroRNA gene expression deregulation in human breast cancer. *Cancer Research*, **65**, 7065–7070.

115 Foekens, J.A., Sieuwerts, A.M., Smid, M., Look, M.P., deWeerd, V., Boersma, A.W. *et al.* (2008) Four miRNAs associated with aggressiveness of lymph node-negative, estrogen receptor-positive human breast cancer. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 13021–13026.

116 Blenkiron, C., Goldstein, L.D., Thorne, N. P., Spiteri, I., Chin, S.F., Dunning, M.J. *et al.* (2007) MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. *Genome Biology*, **8**, R214.

117 Zhao, Y., Deng, C., Wang, J., Xiao, J., Gatalica, Z., Recker, R.R. *et al.* (2011) Let-7 family miRNAs regulate estrogen receptor alpha signaling in estrogen receptor positive breast cancer. *Breast Cancer Research and Treatment*, **127**, 69–80.

118 Uhlmann, S., Zhang, J.D., Schwager, A., Mannsperger, H., Riazalhosseini, Y., Burmester, S. *et al.* (2010) miR-200bc/429 cluster targets PLCgamma1 and differentially regulates proliferation and EGF-driven invasion than miR-200a/141 in breast cancer. *Oncogene*, **29**, 4297–4306.

119 Iliopoulos, D., Polytarchou, C., Hatziapostolou, M., Kottakis, F., Maroulakou, I.G., Struhl, K. *et al.* (2009) MicroRNAs differentially regulated by Akt isoforms control EMT and stem cell renewal in cancer cells. *Science Signaling*, **2**, ra62.

120 Shimono, Y., Zabala, M., Cho, R.W., Lobo, N., Dalerba, P., Qian, D. *et al.* (2009) Downregulation of miRNA-200c links breast cancer stem cells with normal stem cells. *Cell*, **138**, 592–603.

121 Barh, D., Malhotra, R., Ravi, B., and Sindhurani, P. (2010) MicroRNA let-7: an emerging next-generation cancer therapeutic. *Current Oncology*, **17**, 70–80.

122 Guo, X., Wu, Y., and Hartley, R.S. (2009) MicroRNA-125a represses cell growth by targeting HuR in breast cancer. *RNA Biology*, **6**, 575–583.

123 Gaur, A., Jewell, D.A., Liang, Y., Ridzon, D., Moore, J.H., Chen, C. *et al.* (2007) Characterization of microRNA expression levels and their biological

correlates in human cancer cell lines. *Cancer Research*, **67**, 2456–2468.

124 Mattie, M.D., Benz, C.C., Bowers, J., Sensinger, K., Wong, L., and Scott, G.K. *et al.* (2006) Optimized high-throughput microRNA expression profiling provides novel biomarker assessment of clinical prostate and breast cancer biopsies. *Molecular Cancer*, **5**, 24.

125 Zhou, M., Liu, Z., Zhao, Y., Ding, Y., Liu, H., Xi, Y. *et al.* (2010) MicroRNA-125b confers the resistance of breast cancer cells to paclitaxel through suppression of pro-apoptotic Bcl-2 antagonist killer 1 (Bak1) expression. *Journal of Biological Chemistry*, **285**, 21496–21507.

126 Qian, P., Zuo, Z., Wu, Z., Meng, X., Li, G., Wu, Z. *et al.* (2011) Pivotal role of reduced let-7g expression in breast cancer invasion and metastasis. *Cancer Research*, **71**, 6463–6474.

127 Shell, S., Park, S.M., Radjabi, A.R., Schickel, R., Kistner, E.O., Jewell, D.A. *et al.* (2007) Let-7 expression defines two differentiation stages of cancer. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 11400–11405.

128 Dangi-Garimella, S., Yun, J., Eves, E.M., Newman, M., Erkeland, S.J., Hammond, S.M. *et al.* (2009) Raf kinase inhibitory protein suppresses a metastasis signalling cascade involving LIN28 and let-7. *EMBO Journal*, **28**, 347–358.

129 Bhat-Nakshatri, P., Wang, G., Collins, N. R., Thomson, M.J., Geistlinger, T.R., Carroll, J.S. *et al.* (2009) Estradiol-regulated microRNAs control estradiol response in breast cancer cells. *Nucleic Acids Research*, **37**, 4850–4861.

130 Piovan, C., Palmieri, D., Di Leva, G., Braccioli, L., Casalini, P., and Nuovo, G. *et al.* (2012) Oncosuppressive role of p53-induced miR-205 in triple negative breast cancer. *Molecular Oncology*, **6**, 458–472.

131 Wu, H., Zhu, S., and Mo, Y.Y. (2009) Suppression of cell growth and invasion by miR-205 in breast cancer. *Cell Research*, **19**, 439–448.

132 Iorio, M.V., Casalini, P., Piovan, C., Di Leva, G., Merlo, A., Triulzi, T. *et al.* (2009) microRNA-205 regulates HER3 in human breast cancer. *Cancer Research*, **69**, 2195–2200.

133 Kato, M., Paranjape, T., Muller, R.U., Nallur, S., Gillespie, E., Keane, K. *et al.* (2009) The mir-34 microRNA is required for the DNA damage response *in vivo* in *C. elegans* and *in vitro* in human breast cancer cells. *Oncogene*, **28**, 2419–2424.

134 Li, L., Yuan, L., Luo, J., Gao, J., Guo, J., and Xie, X. (2013) MiR-34a inhibits proliferation and migration of breast cancer through down-regulation of Bcl-2 and SIRT1. *Clinical and Experimental Medicine*, **13**, 109–117.

135 Li, X.J., Ji, M.H., Zhong, S.L., Zha, Q.B., Xu, J.J., Zhao, J.H. *et al.* (2012) MicroRNA-34a modulates chemosensitivity of breast cancer cells to adriamycin by targeting Notch1. *Archives of Medical Research*, **43**, 514–521.

136 Yang, S., Li, Y., Gao, J., Zhang, T., Li, S., Luo, A. *et al.* (2013) MicroRNA-34 suppresses breast cancer invasion and metastasis by directly targeting Fra-1. *Oncogene*, **32**, 4294–4303.

137 Eichelser, C., Flesch-Janys, D., Chang-Claude, J., Pantel, K., and Schwarzenbach, H. (2013) Deregulated serum concentrations of circulating cell-free microRNAs miR-17, miR-34a, miR-155, and miR-373 in human breast cancer development and progression. *Clinical Chemistry*, **59**, 1489–1496.

138 Ozgun, A., Karagoz, B., Bilgi, O., Tuncel, T., Baloglu, H., and Kandemir, E.G. (2013) MicroRNA-21 as an indicator of aggressive phenotype in breast cancer. *Onkologie*, **36**, 115–118.

139 Si, M.L., Zhu, S., Wu, H., Lu, Z., Wu, F., and Mo, Y.Y. (2007) miR-21-mediated tumor growth. *Oncogene*, **26**, 2799–2803.

140 Qi, L., Bart, J., Tan, L.P., Platteel, I., Sluis, T., Huitema, S. *et al.* (2009) Expression of miR-21 and its targets (PTEN, PDCD4, TM1) in flat epithelial atypia of the breast in relation to ductal carcinoma in situ and invasive carcinoma. *BMC Cancer*, **9**, 163.

141 Frankel, L.B., Christoffersen, N.R., Jacobsen, A., Lindow, M., Krogh, A., and Lund, A.H. (2008) Programmed cell death 4 (PDCD4) is an important functional target of the microRNA

miR-21 in breast cancer cells. *Journal of Biological Chemistry*, **283**, 1026–1033.
142 Corcoran, C., Friel, A.M., Duffy, M.J., Crown, J., and O'Driscoll, L. (2011) Intracellular and extracellular microRNAs in breast cancer. *Clinical Chemistry*, **57**, 18–32.
143 Qian, B., Katsaros, D., Lu, L., Preti, M., Durando, A., Arisio, R. *et al.* (2009) High miR-21 expression in breast cancer associated with poor disease-free survival in early stage disease and high TGF-beta1. *Breast Cancer Research and Treatment*, **117**, 131–140.
144 Kumar, S., Keerthana, R., Pazhanimuthu, A., and Perumal, P. (2013) Overexpression of circulating miRNA-21 and miRNA-146a in plasma samples of breast cancer patients. *Indian Journal of Biochemistry and Biophysics*, **50**, 210–214.
145 Manikandan, J., Aarthi, J.J., Kumar, S.D., and Pushparaj, P.N. (2008) Oncomirs: the potential role of non-coding microRNAs in understanding cancer. *Bioinformation*, **2**, 330–334.
146 Chen, W., Cai, F., Zhang, B., Barekati, Z., and Zhong, X.Y. (2013) The level of circulating miRNA-10b and miRNA-373 in detecting lymph node metastasis of breast cancer: potential biomarkers. *Tumour Biology*, **34**, 455–462.
147 Biagioni, F., Bossel Ben-Moshe, N., Fontemaggi, G., Yarden, Y., Domany, E., and Blandino, G. (2013) The locus of microRNA-10b: a critical target for breast cancer insurgence and dissemination. *Cell Cycle*, **12**, 2371–2375.
148 Keklikoglou, I., Koerner, C., Schmidt, C., Zhang, J.D., Heckmann, D., Shavinskaya, A. *et al.* (2012) MicroRNA-520/373 family functions as a tumor suppressor in estrogen receptor negative breast cancer by targeting NF-kappaB and TGF-beta signaling pathways. *Oncogene*, **31**, 4150–4163.
149 Ovcharenko, D., Kelnar, K., Johnson, C., Leng, N., and Brown, D. (2007) Genome-scale microRNA and small interfering RNA screens identify small RNA modulators of TRAIL-induced apoptosis pathway. *Cancer Research*, **67**, 10782–10788.
150 Kong, W., Yang, H., He, L., Zhao, J.J., Coppola, D., Dalton, W.S. *et al.* (2008) MicroRNA-155 is regulated by the transforming growth factor beta/Smad pathway and contributes to epithelial cell plasticity by targeting RhoA. *Molecular and Cellular Biology*, **28**, 6773–6784.
151 Mattiske, S., Suetani, R.J., Neilsen, P.M., and Callen, D.F. (2012) The oncogenic role of miR-155 in breast cancer. *Cancer Epidemiology, Biomarkers & Prevention*, **21**, 1236–1243.
152 Volinia, S., Calin, G.A., Liu, C.G., Ambs, S., Cimmino, A., Petrocca, F. *et al.* (2006) A microRNA expression signature of human solid tumors defines cancer gene targets. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 2257–2261.
153 Neilsen, P.M., Noll, J.E., Mattiske, S., Bracken, C.P., Gregory, P.A., Schulz, R.B. *et al.* (2013) Mutant p53 drives invasion in breast tumors through up-regulation of miR-155. *Oncogene*, **32**, 2992–3000.
154 Kong, W., He, L., Coppola, M., Guo, J., Esposito, N.N., Coppola, D. *et al.* (2010) MicroRNA-155 regulates cell survival, growth, and chemosensitivity by targeting FOXO3a in breast cancer. *Journal of Biological Chemistry*, **285**, 17869–17879.
155 Chen, J., Wang, B.C., and Tang, J.H. (2012) Clinical significance of microRNA-155 expression in human breast cancer. *European Journal of Surgical Oncology*, **106**, 260–266.
156 Zhu, W., Qin, W., Atasoy, U., and Sauter, E.R. (2009) Circulating microRNAs in breast cancer and healthy subjects. *BMC Research Notes*, **2**, 89.
157 Sun, Y., Wang, M., Lin, G., Sun, S., Li, X., Qi, J. *et al.* (2012) Serum microRNA-155 as a potential biomarker to track disease in breast cancer. *PLoS One*, **7**, e47003.
158 Lu, Z., Ye, Y., Jiao, D., Qiao, J., Cui, S., and Liu, Z. (2012) miR-155 and miR-31 are differentially expressed in breast cancer patients and are correlated with the estrogen receptor and progesterone receptor status. *Oncology Letters*, **4**, 1027–1032.
159 Lawrie, C.H., Gal, S., Dunlop, H.M., Pushkaran, B., Liggins, A.P., Pulford, K. *et al.* (2008) Detection of elevated levels

of tumour-associated microRNAs in serum of patients with diffuse large B-cell lymphoma. *British Journal of Haematology*, **141**, 672–675.

160 Arroyo, J.D., Chevillet, J.R., Kroh, E.M., Ruf, I.K., Pritchard, C.C., Gibson, D.F. *et al.* (2011) Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 5003–5008.

161 Mitchell, P.S., Parkin, R.K., Kroh, E.M., Fritz, B.R., Wyman, S.K., Pogosova-Agadjanyan, E.L. *et al.* (2008) Circulating microRNAs as stable blood-based markers for cancer detection. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 10513–10518.

162 Turchinovich, A. and Burwinkel, B. (2012) Distinct AGO1 and AGO2 associated miRNA profiles in human cells and blood plasma. *RNA Biology*, **9**, 1066–1075.

163 Chen, X., Ba, Y., Ma, L., Cai, X., Yin, Y., Wang, K. *et al.* (2008) Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Research*, **18**, 997–1006.

164 Boeri, M., Verri, C., Conte, D., Roz, L., Modena, P., Facchinetti, F. *et al.* (2011) MicroRNA signatures in tissues and plasma predict development and prognosis of computed tomography detected lung cancer. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 3713–3718.

165 Roth, C., Rack, B., Muller, V., Janni, W., Pantel, K., and Schwarzenbach, H. (2010) Circulating microRNAs as blood-based markers for patients with primary and metastatic breast cancer. *Breast Cancer Research*, **12**, R90.

166 Liu, J., Mao, Q., Liu, Y., Hao, X., Zhang, S., and Zhang, J. (2013) Analysis of miR-205 and miR-155 expression in the blood of breast cancer patients. *Chinese Journal of Cancer Research*, **25**, 46–54.

167 Heneghan, H.M., Miller, N., Lowery, A.J., Sweeney, K.J., Newell, J., and Kerin, M.J. (2010) Circulating microRNAs as novel minimally invasive biomarkers for breast cancer. *Annals of Surgery*, **251**, 499–505.

168 Si, H., Sun, X., Chen, Y., Cao, Y., Chen, S., Wang, H. *et al.* (2013) Circulating microRNA-92a and microRNA-21 as novel minimally invasive biomarkers for primary breast cancer. *Journal of Cancer Research and Clinical Oncology*, **139**, 223–229.

169 Asaga, S., Kuo, C., Nguyen, T., Terpenning, M., Giuliano, A.E., and Hoon, D.S. (2011) Direct serum assay for microRNA-21 concentrations in early and advanced breast cancer. *Clinical Chemistry*, **57**, 84–91.

170 Zhao, H., Shen, J., Medico, L., Wang, D., Ambrosone, C.B., and Liu, S. (2010) A pilot study of circulating miRNAs as potential biomarkers of early stage breast cancer. *PLoS One*, **5**, e13735.

171 Schrauder, M.G., Strick, R., Schulz-Wendtland, R., Strissel, P.L., Kahmann, L., Loehberg, C.R. *et al.* (2012) Circulating micro-RNAs as potential blood-based markers for early stage breast cancer detection. *PLoS One*, **7**, e29770.

172 Cuk, K., Zucknick, M., Heil, J., Madhavan, D., Schott, S., Turchinovich, A. *et al.* (2013) Circulating microRNAs in plasma as early detection markers for breast cancer. *International Journal of Cancer*, **132**, 1602–1612.

173 Chan, M., Liaw, C.S., Ji, S.M., Tan, H.H., Wong, C.Y., Thike, A.A. *et al.* (2013) Identification of circulating MicroRNA signatures for breast cancer detection. *Clinical Cancer Research*, **19**, 4477–4487.

174 Perou, C.M., Sorlie, T., Eisen, M.B., van deRijn, M., Jeffrey, S.S., Rees, C.A. *et al.* (2000) Molecular portraits of human breast tumours. *Nature*, **406**, 747–752.

175 Sotiriou, C., Neo, S.Y., McShane, L.M., Korn, E.L., Long, P.M., Jazaeri, A. *et al.* (2003) Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 10393–10398.

176 Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H. *et al.*

(2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 10869–10874.

177 Liedtke, C., Mazouni, C., Hess, K.R., Andre, F., Tordai, A., Mejia, J.A. *et al.* (2008) Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. *Journal of Clinical Oncology*, **26**, 1275–1281.

178 Parker, J.S., Mullins, M., Cheang, M.C., Leung, S., Voduc, D., Vickery, T. *et al.* (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of Clinical Oncology*, **27**, 1160–1167.

179 Arpino, G., Generali, D., Sapino, A., Lucia del, M., Frassoldati, A., deLaurentis, M. *et al.* (2013) Gene expression profiling in breast cancer: a clinical perspective. *Breast*, **22**, 109–120.

180 Harbeck, N., Salem, M., Nitz, U., Gluz, O., and Liedtke, C. (2010) Personalized treatment of early-stage breast cancer: present concepts and future directions. *Cancer Treatment Reviews*, **36**, 584–594.

181 Desmedt, C., Giobbie-Hurder, A., Neven, P., Paridaens, R., Christiaens, M.R., Smeets, A. *et al.* (2009) The Gene Expression Grade Index: a potential predictor of relapse for endocrine-treated breast cancer patients in the BIG 1–98 trial. *BMC Medical Genomics*, **2**, 40.

182 Naoi, Y., Kishi, K., Tanei, T., Tsunashima, R., Tominaga, N., Baba, Y. *et al.* (2011) Prediction of pathologic complete response to sequential paclitaxel and 5-fluorouracil/epirubicin/cyclophosphamide therapy using a 70-gene classifier for breast cancers. *Cancer*, **117**, 3682–3690.

183 Rodriguez-Gonzalez, F.G., Sieuwerts, A. M., Smid, M., Look, M.P., Meijer-van Gelder, M.E., deWeerd, V. *et al.* (2011) MicroRNA-30c expression level is an independent predictor of clinical benefit of endocrine therapy in advanced estrogen receptor positive breast cancer. *Breast Cancer Research and Treatment*, **127**, 43–51.

184 He, Y.J., Wu, J.Z., Ji, M.H., Ma, T., Qiao, E.Q., Ma, R. *et al.* (2013) miR-342 is associated with estrogen receptor-alpha expression and response to tamoxifen in breast cancer. *Experimental and therapeutic medicine*, **5**, 813–818.

185 Lowery, A.J., Miller, N., Devaney, A., McNeill, R.E., Davoren, P.A., Lemetre, C. *et al.* (2009) MicroRNA signatures predict oestrogen receptor, progesterone receptor and HER2/*neu* receptor status in breast cancer. *Breast Cancer Research*, **11**, R27.

186 Wang, H., Tan, G., Dong, L., Cheng, L., Li, K., Wang, Z. *et al.* (2012) Circulating MiR-125b as a marker predicting chemoresistance in breast cancer. *PLoS One*, **7**, e34210.

187 Tavazoie, S.F., Alarcon, C., Oskarsson, T., Padua, D., Wang, Q., Bos, P.D. *et al.* (2008) Endogenous human microRNAs that suppress breast cancer metastasis. *Nature*, **451**, 147–152.

188 Song, S.J., Poliseno, L., Song, M.S., Ala, U., Webster, K., Ng, C. *et al.* (2013) MicroRNA-antagonism regulates breast cancer stemness and metastasis via TET-family-dependent chromatin remodeling. *Cell*, **154**, 311–324.

189 Madhavan, D., Zucknick, M., Wallwiener, M., Cuk, K., Modugno, C., Scharpff, M. *et al.* (2012) Circulating miRNAs as surrogate markers for circulating tumor cells and prognostic markers in metastatic breast cancer. *Clinical Cancer Research*, **18**, 5972–5982.

7
Nucleic Acid-Based Diagnostics in Gynecological Malignancies

Sebastian F.M. Häusler, Johannes Dietl, and Jörg Wischhusen

7.1
Introduction

Nucleic acid-based diagnostics has become indispensable for gynecological malignancies, as evidenced by the fluorescence *in situ* hybridization (FISH)-based evaluation of the HER2/*neu* status in breast cancer or human papilloma virus (HPV) diagnostics during cervical cancer screening. While this chapter does not attempt to be exhaustive, these standard applications will be discussed, and further examples will be cited to illustrate the present and emerging role of DNA-, RNA-, or microRNA (miRNA)-based diagnostics in gynecological oncology. As breast cancer is mostly treated by gynecological oncologists in our country (Germany), it is also included in this context. Nevertheless, we are aware that this disease is not universally regarded as a gynecological disease and is thus left to other clinical disciplines in other countries. No such controversy exists with regard to the five most frequent genuinely gynecological malignancies – neoplasms of the cervix, the corpus uteri, the vagina, the vulva, and the adnexa – which will all be addressed in this chapter.

7.2
Cervix, Vulva, and Vaginal Carcinoma

7.2.1
Background

With an estimated 530 000 new diagnoses and 290 000 deaths per year, cervical cancer is the third most common malignant disease worldwide, and the fourth most frequent cause of cancer-related death after lung, breast, and colorectal cancer [1]. Differences between industrialized and developing countries are quite striking. In highly developed countries, cervical cancer is not among the 10 most frequent causes of malignancy-associated death. For less developed regions, however, statistics show cervical cancer to be the malignant disease causing the

Nucleic Acids as Molecular Diagnostics, First Edition. Edited by Andreas Keller and Eckart Meese.

second highest number of casualties [1], making it a most relevant public health problem for these countries. Put another way, the risk of dying from cervical cancer is three times higher for women in developing countries than for women in Europe or the United States [1]. Considering that late-stage cervical cancer is essentially incurable, this difference may be attributed largely to better screening programs and the potential for early diagnostics in the first-world countries. The introduction of suitable vaccines will likely lower the incidence further. In fact, nucleic acid-based diagnostics could already prove the efficacy of the vaccines [2,3].

Due to their viral etiology, cervical neoplasias are very accessible to modern diagnostics. Oncogenic HPVs, which are the underlying cause of the disease, can be detected in more than 99% of affected women [4]. Consequently, risk factors for cervical cancer are mostly associated with an increased risk of virus transmission or with compromised immune function. This includes early onset of sexual activity, promiscuity, high-risk partners (e.g., HPV-infected individuals), positivity for other sexually transmitted diseases (e.g., *Chlamydia* infection), or immunosuppression (e.g., due to HIV infection) [5–8]. Abuse of nicotine, use of oral contraceptives, and low socio-economic status are also considered to increase the respective risks [9,10].

Infection with oncogenic HPV strains likely precedes tumor development by many years, often decades. Initially, the infection is thought to occur in metaplastic epithelial cells from the transformational zone of the cervix uteri (i.e., the tissue at the transition between the outer squamous epithelial cell layer covering the portio and the inner glandular mucosa). Whereas 90% of the infections can be cleared by the immune system within 2 years, viral persistence is likely to occur in the remainder [11]. This enables clonal transformation until an affected cell forms a pre-invasive lesion, which can later give rise to invasive cervical cancer cells that penetrate the basal membrane and form metastases [12]. Histologically, the most frequent subtypes are squamous epithelial or adenomatous carcinomas (69 versus 25%) [13].

Therapeutic options depend on the stage of the disease. HPV-dependent pre-invasive alterations of the cervical mucosa (so-called cervical intraepithelial neoplasias (CINs) can be observed and locally excised. Likewise, early stages of cervical cancer can be cured via local resection. Later-stage cervical cancer, however, requires extensive surgery (removal of the cervix, the uterus, parametrial tissue, pelvic, and often para-aortal lymph nodes according to Wertheim–Meigs–Obayashi or Ruthledge–Piver), which is associated with significant comorbidity. In addition, combined radiochemotherapy is recommended [14]. Thus, both medical and socio-economic considerations stress the importance of early diagnostics for cervical cancer or for high-risk HPV infection.

HPV infection is also considered to be causative for about 60% of all vulvar carcinomas. With about two new diagnoses per 100 000 women per year, these mostly squamous epithelial tumors represent approximately 5% of all genuinely gynecological malignancies (excluding breast cancer) [15,16]. The major difference between HPV-related cervical and HPV-related vulvar carcinoma appears

to be the initial site of infection. Risk factors are thus also similar: associations were found with smoking, with HPV infection, with cervical carcinoma, and with immune deficiencies [17,18]. The pathogenesis of HPV-unrelated vulvar carcinoma, in contrast, was linked to chronic inflammation and to local autoimmune processes such as lichen sclerosus [17]. Therapeutic strategies, again, depend on the stage of disease and range from local excision to radical vulvectomy with evaluation of local lymphatic drainage and radiochemotherapy [19]. Accordingly, early diagnostics can again provide the key to reduce mortality and morbidity.

A further HPV-associated gynecological malignancy is vaginal carcinoma, which also tends to display a squamous epithelial phenotype. With an annual incidence of 1: 100 000 [20], it represents less than 3% of all gynecological malignomas [16]. Risk factors are again linked to HPV infection (promiscuity, early onset of sexual activity, and abuse of nicotine). A previous history of cervical cancer was also more frequent in patients with vaginal carcinoma than in age- and sex-matched controls [20]. Due to its orphan status, there are only few standardized recommendations for the therapeutic management of this disease. In clinical practice, vaginal carcinomas are usually excised, which may require a complete colpectomy and adjuvant radio- and platinum-based chemotherapy [21].

7.2.2 Routine Diagnostics for HPV Infection

Two approaches can be combined for screening and diagnostics of pre-invasive or invasive forms of cervical, vulvar, and vaginal carcinomas. The traditional method developed by George Papanicolaou in 1928 [22] is based on cytologic analysis of a cervical smear in which cells are visualized by staining with tinctorial dyes and acids ("PAP smear"). Microscopic inspection may then reveal cells with atypical morphology and thereby reveal the need for further examination. A complementary approach is the nucleic acid-based search for HPVs, as these are intrinsically involved in the pathogenesis of cervical, vaginal and vulvar cancers (see Section 7.2.1). With the exception of some vulvar cancers and less than 1% of cervical carcinomas, HPV diagnosis can thus detect all patients who are at risk for these malignancies.

HPVs represent a class of DNA viruses which comprises over 100 subtypes. To be classified as a new subtype, an isolate must show differences in greater than 90% of its typically 7900 bp [23]. Four main groups can be distinguished [24]. The first contains the so-called *low-risk* viruses which cause genital warts (condylomata acuminata, benign fibroepithelial tumors). Typical representatives of this group are HPV-6 and -11. The second group encompasses *high-risk* viruses known to induce cervical, vulvar, and vaginal carcinomas. Paradigmatic examples are HPV-16 and -18. Then there are viruses that are classified as "probable/possible high-risk" HPVs (HPV-26, -53, or -66) and subtypes for which no clear risk assessment is available [24].

The HPV genome consists of genes for six "early" (E) and two "late" (L) proteins with E proteins being responsible for viral gene regulation and cell transformation, whereas L proteins form the viral capsid. HPV further contains a "long control region" which is thought to exert regulatory functions on DNA [25,26]. The genetic regions encoding E6 and E7 proteins are essential for the molecular pathogenesis of epithelial carcinomas. Hence, they can always be detected in the respective lesions and their gene products play a seminal role in immortalization of the host cell [27]. Consequently, these regions are primary diagnostic targets.

Strategies for HPV detection use different approaches (which may also be related to patent issues) [28]. In routine diagnostics, tests are performed for HPV DNA. Likewise, the presence of HPV RNA (in particular of E6 and E7 mRNA) can be analyzed. Alternative screening strategies aim at the detection of HPV-associated cellular markers like increased $p16^{INK4a}$ expression. However, as tests based on the first two principles have been approved by the US Food and Drug Administration (FDA), we will focus on these.

The first screening procedures that have gained approval rely on DNA–DNA hybridization. This enables detection of viral nucleic acids via specific probes that can be visualized via signal-amplified chemiluminescence. Then there are polymerase chain reaction (PCR)-based techniques that amplify HPV-specific DNA or RNA segments to enable their detection. In the following, commercially available HPV tests are listed and briefly described [29].

7.2.2.1 Digene Hybrid Capture 2 High-Risk HPV DNA Test (Qiagen)

This test has been FDA-approved since 2003, and is thus the best-established and most widely used nucleic acid-based HPV test in routine clinical use. It detects 13 high-risk HPV subtypes (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68) as well as five low-risk types (6, 11, 42, 43, and 44) without discriminating between them. Signal amplification is achieved via hybridization of HPV DNA with separate complementary, synthetic RNA probes for high- and low-risk HPV subtypes. The resulting DNA–RNA hybrids are captured by immobilized antibodies before alkaline phosphatase-conjugated anti-hybrid antibodies are added. For detection, a dioxetane-based chemiluminescent substrate is used. The signal is proportional to the amount of HPV DNA with a detection limit of 4500–8000 copies of the virus DNA per sample. According to a recommendation by the FDA, a relative light intensity (RLU) ≥ 1 should be considered positive, which corresponds to 1 pg HPV DNA or 16 500 copies.

7.2.2.2 Cervista HPV HR (Hologics)

The Cervista HPV HR test also uses hybridization techniques to detect DNA of the high-risk HPV subtypes 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68 without subtyping. While the Hybrid Capture 2" test is based on base complementarity, this test detects atypical DNA triplex structures that are characteristic for HPV carriers and that can be cleaved by a specific enzyme. This leads to dissociation of the overlapping structure ("5′ flap"). Using fluorescent resonance

energy transfer (FRET), these flaps can generate a fluorescent signal that can be amplified and detected. As the number of flaps depends on the initial amount of HPV DNA, quantification is possible via the intensity of the FRET signal. The analytical detection limit for this test lies between 1250 and 7500 copies of the viral genome per sample. For further subtyping, the Cervista HPV 16/18 test can be used, which only recognizes the high-risk HPV subtypes 16 and 18. Both tests were approved by the FDA in 2009.

7.2.2.3 cobas 4800 System (Roche)

This quantitative real-time (qRT)-PCR-based method coamplifies DNA of the high-risk HPV subtypes 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68 in a single reaction. Using specifically labeled probes, it allows either subtyping for HPV-16 and -18 or a more general quantification of high-risk HPV gene segments. Advantages are automated sample preparation and qRT-PCR (which minimizes the risk of contamination), and the fairly precise quantification of small amounts of template. The analytical cutoff is indicated as 150–2400 virus copies per sample. To obtain clinically meaningful results, however, it was adjusted and is now similar to the detection limit of the Hybrid Capture 2 test. FDA approval was granted in 2011.

7.2.2.4 APTIMA HPV (Gen-Probe)

The diagnostic principle underlying this test is the detection of mRNA from the viral oncoproteins E6 and E7. This does not only cover HPV subtypes 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68, but also includes genotyping. In a first step (target capture), magnetic bead-coupled DNA probes are used to pull down HPV mRNA from the sample. Importantly, the capture oligomers (primer 1) contain a T7 promoter site and can be elongated by reverse transcription using the captured RNA as template. While the initial template RNA is digested by RNase H activity, further amplification (which requires primer 2) leads to formation of double-stranded DNA. The DNA strand containing the T7 promoter site then enables the T7 RNA polymerase to synthesize further RNA copies resembling the original HPV RNA sequence. At the same time, reverse transcription of the complementary DNA strand re-initiates the synthesis of double-stranded cDNA templates, which include the T7 promoter sequence. In a last step, the obtained amplicons are hybridized with a labeled probe and detected via chemiluminescence. As the level of amplification far exceeds that of conventional PCR-based methods, as few as 20–500 virus copies can be detected in a sample. For clinical diagnostics, however, the level of sensitivity has been reduced and now resembles that achieved by other tests. FDA approval was also granted in 2011.

7.2.2.5 Abbot RealTime High Risk HPV Assay (Abbot)

As the name indicates, this assay is based on RT-PCR. It enables the detection of 14 high-risk HPV subtypes (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68), and the genotyping of HPV-16 and -18. Spiking with an internal control

allows quantification of the initial HPV copy number and also the monitoring of the qRT-PCR reaction for efficiency and detection quality. Abbot recommends that signals based on amplification of HPV are rated as positive when they are detected with less than 32 cycles of amplification. While this assay has not yet been approved by the FDA, studies have shown that its sensitivity and specificity resemble those of the Hybrid Capture 2 test [30].

7.2.2.6 PapilloCheck Genotyping Assay (Greiner BioOne)

The PapilloCheck genotyping assay from Greiner BioOne is a microarray-based method for the detection of 24 different HPV subtypes (6, 11, 16, 18, 31, 33, 35, 39, 40, 42, 43, 44/45, 45, 51, 52, 53, 56, 58, 59, 66, 68, 70, 73, and 82). First, fragments of the HPV-E1 protein are PCR-amplified. Next, PCR products are fluorescently labeled and hybridized to DNA probes that have been spotted on a microarray. Detection occurs via the fluorescent signal. This enables computer-aided discrimination between a wide range of HPV genotypes, which represents an asset of the approach. Analytically, 20–750 virus copies (depending on the respective subtype) are required for detection. Clinically, the cutoff was adjusted to the standard set by the Hybrid Capture 2 test. In studies, this assay has shown similar sensitivity and specificity as other assays, but it is not yet approved by the FDA [31].

7.2.2.7 INNO-LiPA HPVG enotyping Extra (Innogenetics)

In this test, the highly conserved L1 gene is PCR-amplified, which enables the detection and genotyping of HPV subtypes 6, 11, 16, 18, 31, 33, 35, 39, 40, 42, 43, 44, 45, 51, 52, 53, 56, 58, 59, 66, 68, 69, 70, 71, 73, 74, and 82. Addition of a primer mix to control for sample and extraction quality and the presence of uracil-*N*-glycosylase contribute to a well-controlled test that can detect 20–700 virus copies. However, there is no defined cutoff for a clinically relevant signal, which compromises the clinical usability of this highly sensitive method.

7.2.2.8 Linear Array (Roche)

Roche's Linear Array is also based on the PCR amplification of L1 gene fragments. In addition, a region of the β-globin gene gets amplified. In a second step, genotyping of 36 HPV subtypes (6, 11, 16, 18, 26, 31, 33, 35, 39, 40, 42, 44, 45, 51–54, 56, 58, 59, 61, 62, 64, 66–73, 81–84, and 89) is achieved on a test strip that has been covered with sequence-specific probes. Sensitivity is very high, but the classical Hybrid Capture 2 test appears to be more specific [32].

7.2.2.9 Recommendations for Clinical Use

Apart from the technical and socio-economic feasibility, some further considerations should guide the implementation of HPV tests in clinical practice. Purely screening for HPV may not be advisable for patients who are less than 30 years old. In this younger collective, infection even with high-risk HPV is fairly common, but progressing higher-stage intraepithelial lesions are still rare. At around

30 years of age, the link between a positive HPV diagnosis and cervical neoplasia becomes more apparent [33]. Accordingly, an inconspicuous PAP smear might then be complemented by a further nucleic acid-based HPV test. A further application is the quest for HPV in the case of a conspicuous cytological finding (PAP class IIw or, for the first time, PAP IIID) [34]. In case of a positive test, further diagnostics via differential colposcopy is recommended; a negative test result would allow for a reduced frequency of controls [34]. HPV diagnostics is also advisable during the follow-up after excision of HPV-positive neoplasia [35]. For risk assessment, genotyping is also relevant as HPV-16- or -18-positive cervical neoplasias are associated with a higher risk of progression towards CIN III [36]. Further indications include the follow-up after a suspicious finding from colposcopy or cases where anatomical peculiarities prevent a proper inspection of the cervix [37].

7.2.3 Outlook – DNA Methylation Patterns

For cervical cancer, interesting approaches like further gene expression or miRNA analyses are still experimental, and it cannot be predicted whether they will transfer to clinical diagnostics. However, analysis of DNA methylation patterns has already reached a fairly advanced stage and should thus also be discussed. As DNA methylation can prevent transcription of specific genes, it represents an important epigenetic regulation mechanism [38]. Methylation typically occurs at GC base pairs hat can be bridged via phosphodiesters to yield methylated CpG sites. Upon HPV infection, methylation changes occur that can inactivate tumor suppressor genes in the afflicted epithelial cells and thereby promote cellular immortalization [39]. Suitable PCR protocols or microarray techniques allow the detection of such aberrant methylations: pre-treatment of the DNA samples with bisulfite converts methylated cytosine into uracil, which marks a methylation site. Subsequent “bisulfite sequencing” then allows the sequence-specific detection of these epigenetically modified sites and hence of methylation patterns [40]. More restricted information is obtained by using methylation-specific primers which are complementary either to methylated or unmethylated cytosine residues (“methylation-specific PCR”) [41]. While such primer pairs can also be used in multiplex combinations, the logical next step was the development of microarray-based methylation analysis. This employs pairs of oligonucleotide probes that are either specific for the methylated or for the non-methylated CpG site. As methylation-specific primer pairs for all possible methylation sites can be spotted on a microarray, genome-wide methylation patterns can be analyzed on a chip [42]. For cervical cancer, methylation changes induced by HPV are of particular interest. Clinical studies showed a possible correlation between hypermethylation of viral L1, L2, E2, and E3 genes and high-grade cervical intraepithelial neoplasia or cervical cancer [39,38]. Thus, the method might be suitable to assess the risk for cervical neoplasia more precisely within the group of HPV-positive women.

7.3 Endometrial Carcinoma (Carcinoma Corpus Uteri)

7.3.1 Background

About 287 000 women are diagnosed annually with endometrial carcinoma [1]. With a peak incidence between 65 and 80 years of age, the disease affects mainly elderly women [43]. Clinically, two distinct subtypes are known. Based on histology, the more frequent type 1 tumors are endometrial adenocarcinomas. They arise from atypical hyperplasias of the endometrium and depend on estrogen for proliferation. Type 2 carcinomas comprise serous or clear cell subtypes as well as carcinosarcomas and have a less favorable prognosis [44]. Known risk factors, especially for type 1 carcinomas, are increased age, hormonal replacement therapy, especially when estrogen is not balanced by progesterone, late menopause, early menarche, nulliparity, polycystic ovarian syndrome, adipositas, type 2 diabetes, use of tamoxifen, or familial predispositions, such as Lynch syndrome [45]. Many of these factors are thus associated with increased exposure to estrogen. Clinically, vaginal bleeding indicates the presence of an endometrial carcinoma mostly at an early stage, which leads to a rather favorable prognosis. In 2006, the Fédération Internationale de Gynécologie et d'Obstétrique (FIGO) reported a 5-year survival of 80% over all stages [43,46]. The therapy relies on surgical resection, which in some cases requires radical hysterectomy, and, depending on the stage, pelvic and para-aortal lymphadenectomy. Adjuvant treatments include irradiation and platinum- or anthracycline-containing chemotherapy, which have been reported to be at least as effective as the previous standard of radiotherapy [44].

7.3.2 Routine Diagnostics – Microsatellite Instability

In clinical practice, nucleic acid-based diagnostics plays only a minor role for endometrioid cancer. Nevertheless, as Lynch syndrome leads to a familial predisposition for endometrial cancer, the related diagnostics should briefly be mentioned. This condition, which is also known as hereditary non-polyposis colorectal cancer syndrome (HNPCC), is passed on in an autosomal-dominant manner and 60–80% of all carriers will eventually develop a carcinoma corpus uteri, mostly at a relatively young age [44]. The underlying mutation occurs typically in one of the DNA mismatch repair genes *MLH1*, *MSH2*, *MSH6*, or *PMS2* [47]. Diagnostically, an algorithm has been proposed [47] that starts by screening for microsatellite instability (MSI) in neoplastic tissue from patients with an appropriate familial history. Alternatively, expression of HNPCC-associated genes can be analyzed via immunohistochemistry (IHC). Microsatellites are repetitive DNA sequences consisting of 1–6 bp that are spread throughout the genome. Due to their highly repetitive nature, they are prone to errors during

DNA replication. This can lead to accidental duplication or loss of such repeats [48]. To analyze microsatellite structures, patient DNA is obtained (e.g., from lymphocytes) and from tumor material. Next, gene loci that are known to be susceptible to MSI (e.g., *AT-25*, *BAT-26*, *D2S123*, *D5S346*, and *D17S250*) are PCR-amplified, and the respective length of the products is compared between the tumor and the healthy tissue [49,50]. If this leads to the detection of MSI in the tumor, sequencing of HNPCC genes is recommended. If no mutations can be detected, methylation analysis (compare Section 7.2.3) of the respective promoters and genes should be considered [51]. This stepwise procedure takes into account that MSI occurs in 25% of type 1 endometrial carcinomas, while mutations in *PTEN*, *KRAS*, *PIK3CA*, and *CTNNB1* (β-catenin) can also occur. Even in patients with Lynch syndrome, MSIs are not pathognomic, although they can be found in 75% of all cases [52].

7.3.3
Emerging Diagnostics – miRNA Markers

miRNAs consist of 21–23 bp. They represent non-coding RNA (ncRNA) segments that can regulate gene expression and thus cell function in a highly specific manner [53]. Over 2000 of these basic gene regulators have now been discovered and each one can influence the expression of several hundreds of target genes. During tumorigenesis, miRNAs are thought to play a multifaceted role, as they can act as oncogenes (e.g., miR-155) or as tumor suppressors (e.g., miR-15a). Deregulated miRNA expression may be caused by mutations, but also by genetic and epigenetic alterations affecting their transcriptional control. Altered expression of individual miRNAs can be shown using PCR protocols that often generate a cDNA with a suitable adapter to enable later detection by qRT-PCR [54]. Larger groups of miRNAs can be analyzed by microarrays like the microfluidic primer extension assay (MPEA) [55], which does not require pre-amplification of miRNAs since hybridization of miRNAs with immobilized probes is followed by enzymatic primer extension.

Major advantages of miRNAs as diagnostic targets are their relatively high stability, the ease of detection, and the possibility to compute multiparametric expression patterns. In endometrial carcinoma, both overexpressed (miR-185, miR-106a, miR-181a, miR-210, miR-423, miR-103, miR-107, miR-Let7c, miR-205, miR-200c, miR-449, miR-429, miR-650, miR-183, miR-572, miR-200a, miR-182, miR-622, miR-34a, and miR-205) and downregulated miRNAs (miR-Let7e, miR-221, miR-30c, miR-152, miR-193, miR-204, miR-99b, miR-193b, miR-204, miR-99b, miR-193b, miR-411, miR-133, miR-203, miR-10a, miR-31, miR-141, miR-155, miR-200b, and miR-487b) were found. Target genes of deregulated miRNAs include a number of candidates that are known to be involved in tumor development. Examples include *KCNMB1*, *IGFBP-6*, *ENPP2*, *TBL1X*, *CNN1*, *MYH11*, *KLF2*, *TGFB1/1*, *MYL9*, *SNCAIP*, *RAMP1*, *FOXO1*, *FOXC1 E2F3*, *MET*, and *Rictor* [52]. Investigations regarding the prognostic relevance of the various miRNA deregulations are ongoing. In 2012, Karaayvaz *et al.*

presented study data that proposed that miR-205 may represent a new prognostic marker in endometrial cancer [56].

Another class of ncRNAs are so-called "long ncRNAs" (lncRNA), which are usually longer than 200 bp. They are thought to modulate the recruitment of histone-modifying complexes, and to influence transcription and splicing [52]. In spite of our restricted knowledge on lncRNAs, candidates which may be characteristically deregulated have already been identified in endometrial cancer. In particular, the lncRNA NC25 seems to be mutated in about half of all endometrial carcinomas [57]. Accordingly, NC25 might soon become a diagnostic target. In addition, further remarkable findings relating to epigenetic regulatory mechanisms might well become future diagnostic targets in endometrial carcinoma.

7.4 Ovarian Carcinoma

7.4.1 Background

With 225 000 new diagnoses per year, ovarian carcinoma is the eighth most frequent carcinoma in women. Moreover, the disease is associated with an extremely high rate of mortality evidenced by the annual rate of 140 000 deaths from ovarian cancer in industrialized countries [1,16]. The outcome depends largely on the tumor stage at diagnosis. When the cancer is still locally confined to the ovaries, 5-year survival rates exceed 90%, whereas only 10% of the patients with disseminated intraperitoneal or distant metastasis will still be alive after the same period. Unfortunately, more than two-thirds of all ovarian cancers are only detected at such a late stage [58], which can be attributed to the absence of early symptoms and to the lack of a suitable screening test. Consequently, development of a reliable procedure for early diagnostics is of paramount importance in this disease.

More than 95% of all ovarian malignancies originate from epithelial cells ("epithelial ovarian cancers"), the remainder are formed by other cell types in the ovary, such as germ cells or stromal cells ("germ cell tumors" or "sex cord stromal tumors") [59]. Epithelial ovarian carcinomas can be further subdivided histologically into serous-papillary ("serous"), mucinous, endometrioid, clear cell, and some exceedingly rare tumor types. By far the most common histological diagnosis is serous ovarian carcinoma. According to their biological behavior and their genetic alterations, these malignancies can be further subclassified as high- or low-grade serous ovarian carcinomas [59]. Nucleic acid-based diagnostics could thus further refine the histological assessment and possibly guide therapeutic decisions.

To explain the origin of serous ovarian carcinomas, two hypotheses dominated the literature. In the 1970s, studies showed that repeated ovulation – causing

tissue rupture and repair processes – may increase the number of mutations in ovarian cells, which, in turn, could promote malignant transformation [60,61]. Further, it was postulated that excessive secretion of gonadotropin combined with high levels of estrogen would increase the proliferation of ovarian epithelial cells, which might also promote malignant progression [62]. Based on recent immunohistochemical and genetic assessments, however, evidence is accumulating that high-grade serous ovarian cancer, tubal carcinoma, and primary peritoneal cancer all develop from precursor lesions in a transitional zone close to the fimbriated end of the fallopian tube and thus represent the same disease [63–65]. Importantly, presumed precursor lesions detected in this compartment where found to display the same characteristic mutations (e.g. "p53 signatures") as the corresponding invasive tumors in the ovaries and the peritoneum. Accordingly, attempts to develop suitable screening strategies for precancerous lesions may well have focused on the wrong location, which likely contributed to the numerous failures.

Empirically determined risk factors for ovarian cancer are in line with the presumed hormone-dependent, ovulation-associated pathogenesis. This includes age, early menarche, late menopause, nulliparity/infertility, polycystic ovarian syndrome, hormonal treatment, endometriosis (as risk factor for endometrioid carcinomas), and familial predisposition for gynecological malignancies [66–74]. While 85% or more of all ovarian cancers occur sporadically without familial history, germline mutations in *BRCA1/2* and some unknown genes lead to an inheritable risk for ovarian cancer. *BRCA* diagnostics, however, will be discussed in the section on breast cancer (Section 7.5.2) where mutations in these breast cancer-associated genes are also of enormous relevance.

Adjuvant treatment of ovarian cancer is mainly based on surgical resection complemented by chemotherapy and specific antibody therapy. Surgical interventions in ovarian cancer patients are among the most extended gynecological operations. As reduction of the tumor load below a certain threshold clearly improves the individual prognosis, there is a strong rationale for extended multivisceral interventions in patients with disseminated disease spreading throughout the abdomen [75]. Subsequent chemotherapeutic regimens are typically based on platinum-containing compounds and taxanes. The most recent addition to the treatment protocol is the anti-vascular endothelial growth factor antibody bevacizumab [76,77]. In addition, there is a strong rationale for T cell-based immunotherapies in ovarian cancer [78,79].

7.4.2 Routine Diagnostics

Most screening or follow-up/after-care approaches in ovarian cancer are based on the detection of fairly unspecific tumor and inflammation markers like CA125 or on ultrasound diagnostics, or on biomathematically computed combinations of both [58]. Nucleic acid-based diagnostics is not routinely used in this indication. The only notable exception is the DNA sequencing-based search for

rare hereditary cancer markers, which include mostly mutations in *BRCA1/2*. Even less frequent genetic risk factors [80] like a familial predisposition for Lynch syndrome (HNPCC) or for Peutz–Jeghers syndrome (also known as hereditary intestinal polyposis syndrome) can also be analyzed by DNA sequencing analysis. With a lifetime risk for ovarian cancer of 46% or less in *BRCA* mutation carriers (depending on the respective mutation) and of 12% in women with Lynch syndrome, there is certainly a strong rationale for testing. Technical details of the test for Lynch syndrome (which is typically caused by a mutation in one of the genes for either *MLH1*, *MSH2*, *MSH6*, *PMS2*, or *EPCAM*) have already been described in Section 7.3.2. *BRCA* diagnostics will be further explained in the section on nucleic acid-based diagnostics in breast cancer (Section 7.5.2).

Peutz–Jeghers syndrome is an autosomal dominant genetic disease characterized by hyperpigmented patches on the skin and by the development of benign gastrointestinal polyps. It is caused by a mutation in the gene encoding the serine/threonine kinase *STK11/LKB1* on chromosome 19p13.3. While this condition causes a strong disposition for gastrointestinal tumors, associations with further cancer types including ovarian carcinoma could also be shown. For detection, the following approaches have been reported: conformational sensitive gel electrophoresis, single-strand conformational polymorphism, denaturing high-performance liquid chromatography, denaturing gradient gel electrophoresis, and direct sequencing of exons. The Institute of Cancer Research and the Mayo Clinic have additionally made use of long-range PCR, and the Institute of Cancer Research and the VU University Medical Center have used multiplex ligation-dependent probe amplification to screen larger patient collectives [81].

7.4.3 Emerging Diagnostics/Perspective – miRNA Profiling

miRNAs are important regulators of post-transcriptional gene expression implicated in numerous diseases (compare Section 7.3.3). Deregulated miRNA expression has also been reported for ovarian carcinoma. Importantly, miRNA expression can easily and comprehensively be quantified via next-generation sequencing, via miRNA arrays, or via MPEA [55]. Data on various miRNAs can then be combined to calculate a multimodal marker profile. The use of such miRNA panels for the early diagnostics of ovarian cancer thus appears promising. Using a blood-based test on patients with confirmed diagnosis of ovarian cancer, altered miRNA expression patterns were already shown relative to age- and sex-matched healthy controls and to patients with unrelated diseases [82,83]. The strength of the approach lies in the ability to simultaneously measure and evaluate a number of parameters, which means that the diagnosis does not rely on any individual parameter but rather on a selected subset of markers ($n = 147$ in the aforementioned studies) that can be computed via unbiased statistical learning methods [84]. The appropriate bioinformatic processing of these data then leads to a greatly enhanced sensitivity and specificity as compared with

individual markers. Such a marker profile should also be less prone to outliers, which may be due to specific conditions of the individual patient (e.g., CA125 is upregulated in most inflammatory conditions). A blood test as used for the proof-of-principle study should also be a very suitable tool for screening. It remains to be determined whether free miRNAs from serum or a profile that includes the cellular components from whole blood is preferable. Cell-free miRNAs should contain a higher proportion of tumor (or tumor exosome)-derived miRNAs. However, inclusion of blood cells allows detection of immune responses to the underlying disease. Thus, significant efforts will be required for further optimization of the assay and for the establishment of a diagnostic test. Nevertheless, as the data obtained during these early stages compared favorably with most other further developed screening approaches, there is at least some hope that multiparametric miRNA-based diagnostics could help to overcome the biggest challenge in ovarian cancer, which is still the development of a suitable screening test for early detection.

7.5 Breast Cancer

7.5.1 Background

With over 1.38 million new diagnoses per year, breast cancer is by far the most frequent carcinoma in women [1]. Diagnostic and clinical management have constantly been improved and new therapies are emerging. Still, the disease takes a death toll of almost 500 000 women per year and thus remains the most frequent cause of cancer-related death in women [1].

Known risk factors can be genetic, such as mutations in the breast cancer-associated *BRCA* genes, or cancer predispositions, such as Li–Fraumeni (caused by mutant p53) or Cowden (mutant PTEN) syndrome. Mutations in genes with low or moderate penetrance have also been described [85]. A second set of risk factors is associated with increased exposure to estrogen and other hormones. Hormonal replacement therapy, early menarche, late menopause, and obesity are thus risk-promoting, whereas pregnancies and breast feeding are protective [86–89].

Adjuvant therapy of mammary carcinoma is based on surgery, irradiation, chemotherapy, endocrine therapy, and antibodies [90,91]. The combination of various treatments nowadays enables less radical procedures during surgery. In fact, breast-preserving surgery can be as effective as mastectomy provided that it is followed by irradiation. Likewise, the previous strategy of removing all axillary lymph nodes is challenged by recent studies [92]. Nevertheless, this is individually decided according to the histopathological assessment. The use of chemotherapeutics like anthracyclines or taxane is indicated for high-grade tumors or when the lymphatic drainage is affected or for HER2/*neu*-positive tumors.

Endocrine therapy, which can be quite effective, is applied whenever a tumor is hormone receptor-positive. For *HER2*-overexpressing tumors, the anti-HER2 antibody trastuzumab has become the standard-of-care [93] and a new antibody (pertuzumab [94]) as well as an antibody–drug conjugate (trastuzumab emtansine [95]) have been approved by the FDA. Palliative treatments are mainly endocrine based or guided by the respective symptoms.

7.5.2 Routine Diagnostics

The term "breast cancer" comprises subtypes whose histopathological and molecular characteristics are so different that they should be seen as different diseases unified by a common localization. Accordingly, an aggressive treatment which may be life-saving for one patient might be an absolute overtreatment for another patient whose quality of life would be unnecessarily diminished [90]. Further subclassification of the disease is thus required to allow for individually adapted treatment schemes. Nucleic acid-based diagnostics appear to be a most promising tool to enable the necessary discrimination, even though classical pathological tools, such as assessment of hormone receptor status, tumor grading, and nodal status [90], will likely retain their place in breast cancer diagnostics.

7.5.2.1 HER2 Diagnostics

The oncogene *HER2* (human epidermal growth factor receptor 2, erb-B2, c-erbB2, Her2/*neu*, *ERBB2*) encodes a 185-kDa transmembrane receptor for epidermal growth factor that possesses intracellular tyrosine kinase activity [96,97]. Ligand-dependent and ligand-independent activation of this receptor drives carcinogenic processes such as proliferation and angiogenesis [98]. In breast cancer patients, *HER2* expression and potential gene amplification was found to be predictive for at least three reasons. When *HER2* is overexpressed in a specific tumor, it represents a suitable target for the antibodies trastuzumab or pertuzumab, for the small-molecule receptor inhibitor lapatinib, and for combinations thereof [94,99,100]. The *HER2* status further appears to correlate with the response to chemotherapy and to hint at a likely resistance towards endocrine therapies [101,102]. HER2-overexpressing tumors used to relapse more frequently and to show a shorter survival. However, due to the availability of targeted therapeutics, the prognostic value of *HER2* expression is no longer as clear as it used to be [103].

Currently, three approaches for the determination of the *HER2* status are clinically accepted:

- Amplification of the *HER2* gene can be detected by FISH, chromogenic *in situ* hybridization (CISH), silver-enhanced *in situ* hybridization (SISH), or via differential PCR.
- At the RNA level, *HER2* overexpression can be assessed by Northern blotting or by reverse transcription PCR.

- At the protein level, HER2 can be visualized by Western blotting, by enzyme-linked immunosorbent assay (ELISA) or IHC. Recent studies have also analyzed the relevance of circulating HER2 extracellular domain (c-ECD) protein levels for the response to therapy (e.g., [104]).

During the first efficacy studies with trastuzumab, *HER2* overexpression had been assessed in paraffin sections using immunohistochemical staining and a grading system from 0 (no expression) to 3+ (strong overexpression) [105]. Subsequent studies showed that all patients who had a clear benefit from trastuzumab displayed either an IHC overexpression score of 3+ or a positive FISH test for gene amplification combined with an IHC score of 2+ or more [106]. IHC scores of 0 and 1 strongly correlated with a lack of amplification according to *HER2* FISH analyses. In practice, most diagnostic laboratories now use the FDA-approved, IHC-based HercepTest from Dako [107]. In addition, there are FDA-approved reagents that enable numerous other *HER2* analysis tests, which are all based on the above-mentioned techniques. The American Society of Clinical Oncology recommends that all breast cancers should be tested for *HER2* overexpression. Criteria for a positive test result would be an IHC score of 3+, or a FISH test indicating more than six copies of *HER2* per nucleus, or a FISH *HER2* to chromosome 17 ratio of 2.2 or more (FISH ratio). Patients with an IHC score of 1+ or less, less than four copies of *HER2* per nucleus, or a FISH ratio of 1.8 or less are unlikely to benefit from trastuzumab and thus should not be treated. Values in between the positive and the negative cutoff indicate the need for further testing. New tests should be accepted provided they are at least 95% congruent with established validated tests [108]. As the *HER2* status can change when an initially hormone receptor-positive, *HER2*-negative breast cancer forms distant metastases, further histological *HER2* testing is recommended upon metastasis. Patients with *HER2*-positive metastases are also likely to benefit from *HER2*-directed targeted therapy [109].

7.5.2.2 Gene Expression Profiling

While different subtypes of breast cancer are known to display different histopathological morphologies and variable clinical behavior, microarray-based gene expression profiling revealed huge genetic differences between various tumors. The term breast cancer thus describes the initial site of tumor formation (without further refinement). Genetically, however, this tumor entity comprises very different diseases. In 2011, an international conference dedicated towards building an expert consensus on the primary therapy of early breast cancer [110] came to the conclusion that at least four different groups need to be distinguished: luminal A and B, basal-like, and *HER2*-positive breast cancer. Further genetically distinct subgroups (e.g., "normal breast like," "Claudin low," and "molecular apocrine") are exceedingly rare and thus clinically less relevant. For implementation in clinical diagnostics, approximate descriptions of the four genetically distinct subgroups were based on routinely assessed histological parameters. Luminal A and B tumors both show high estrogen and progesterone

receptor levels in the absence of *HER2* overexpression (even though a subgroup of luminal B tumors may express *HER2*). These two subtypes are differentiated via their proliferative behavior (as measured by Ki-67 staining). Basal-like carcinomas show neither hormone receptor nor *HER2* expression and are thus also called "triple negative." *HER2*-positive carcinomas likewise lack hormone receptor expression, but are defined via high levels of *HER2*. Importantly, these subtypes show different responses towards therapeutic treatment. Whereas slowly proliferating luminal A tumors respond well to endocrine therapy, chemotherapy is less effective. Luminal B tumors, instead, proliferate much more rapidly and are thus more sensitive towards chemotherapeutic drugs (and in some cases towards targeted anti-*HER2* treatments). *HER2*-specific treatments are indicated for *HER2*-positive tumors. Triple-negative carcinomas, in contrast, cannot be effectively treated with targeted therapies and are consequently afflicted with a worse prognosis than all other subtypes [110].

While these guidelines appear rather simple, they need to be adapted to the real life situation, which does not always fit into the scheme. Instead, clinicians often encounter intermediate-risk situations (e.g., tumors with a luminal phenotype and an intermediate rate of proliferation). In these cases, treatment decisions should be based on additional information that can be provided by nucleic acid-based diagnostic tests. These include fully automated, commercially available tests that provide individualized prognostic assessments based on a combination of gene expression parameters. Some of these tests (like the Prosigna tests) have implemented the above-described genetic subtypes into their analysis. A selection of available tests and the underlying principles is provided in Table 7.1. Except for the Prosigna, all other tests draw a correlation between poor prognosis and response to chemotherapy, which implies a certain predictive power. According to the recommendations of the latest consensus conference, multigene analyses are mostly recommended for luminal tumors as these tests may indicate whether a patient is likely to benefit from chemotherapy or not. Clinically, the 21-gene recurrence score (Oncotype DX) is currently favored, but the PAM50 signature (Prosigna) and the 70-gene signature (MammaPrint) or the EPClin score (EndoPredict) appear to be valid alternatives [111].

7.5.2.3 Hereditary Breast Cancer/BRCA Diagnostics

Only 10% of all breast and 15% of all ovarian cancers can be attributed to inheritable germline mutations. Most frequently afflicted are the genes for BRCA1 and BRCA2. Both encode enzymes that can repair DNA double-strand breaks via homologous recombination. *BRCA1/2* mutations are both autosomal dominant, though their level of penetrance is different. *BRCA1* mutation carriers have a 50–85% lifetime risk for breast and a 39–46% risk for ovarian cancer, whereas *BRCA2* mutations are associated with a 40–85% risk for breast and a 10–20% risk for ovarian carcinoma [116]. *BRCA1* mutations further predispose for cervical, uterine, pancreatic, colon, and prostate cancer [117]. *BRCA2* mutation carriers are more likely to develop cholangiocarcinoma or gall bladder, prostate, pancreatic, gastric cancer, or malignant melanoma [118]. Heterozygous carriers

Table 7.1 Gene expression profiling tests (modified from [112–115]).Gene expression profiling tests (modified from [112–115]).

	Test						
	Oncotype DX	**EndoPredict**	**MammaPrint**	**Prosigna**	**Veridex 76-Gene**	**Breast Cancer Index**	**MapQuant DX/simplified**
Manufacturer	Genomic Health, USA	Sividon Diagnostics, Germany	Agendia, Netherlands	NanoString Technologies, USA	currently not available	bio Theranostics, USA	Ipsogen, France
Method	qRT-PCR	qRT-PCR	microarray	microarray	microarray	qRT-PCR	microarray/ qRT-PCR
Coverage	21-gene recurrence score	11-gene score combined with tumor size and nodal status	70-gene signature	PAM50 50-gene signature combined with tumor size and proliferation	76-gene signature	Two-gene *HOXB13: IL17R*/molecular grade index	97-gene signature or eight-gene PCR
Tissue	FFPE	FFPE	frozen or stabilized RNA	FFPE	frozen	FFPE	frozen or FFPE
Outcome parameter	10-year risk of recurrence	10-year risk of recurrence	5-year risk for distant metastasis	10-year risk of recurrence	5-year risk for distant metastasis	disease-free/overall survival	good versus poor prognosis
Output statement	continuous score	continuous score	dichotomous: good versus poor prognosis	continuous score	dichotomous	continuous score	dichotomous (genomic grade index (GGI))
Main area of application	ER-positive, HER2-negative, 0–3 affected lymph nodes	ER-positive patients	N0, tumor size ≤5 cm, stage I/II, <61 years, 1–3 affected lymph nodes	N0	N0	ER-positive, prediction of response to tamoxifen therapy	ER-positive, G2

FFPE: formalin-fixed paraffin-embedded; ER: estrogen receptor; N0: nodal negative, G: WHO grading.

of the mutation are more likely to suffer a complete loss of this tumor suppressor mechanism upon cell division. To compensate for a loss of BRCA activity, cells often hyperactivate alternative DNA repair mechanism like poly(ADP-ribose) polymerase (PARP)-dependent repair of single-strand breaks. Accordingly, there is some hope that PARP inhibitors, as an emerging class of new drugs, may be most toxic for BRCA-deficient carcinoma cells and thus reasonably tumor specific [116]. Carcinomas with *BRCA* mutations typically arise as high-grade tumors that show a medullary phenotype, and lack estrogen and progesterone receptors as well as *HER2* amplifications ("triple negative") [119].

Testing for *BRCA* mutations can be done from blood or saliva. As several hundred possible mutations have been identified in these genes, sequencing of all exons of both genes and of the flanking regions is advisable to cover known and potentially new mutations. Further, testing for genetic rearrangements is recommended for women with a family history of breast and ovarian cancer (familial high-risk patients) who did not show any mutation upon sequencing [120]. Biochemical techniques used include denaturing high-performance liquid chromatography, DNA sequencing, multiplex ligation-dependent probe amplification, and high-resolution melting analysis performed on a LightCycler [121]. As these tests are somewhat laborious and expensive, they are mostly used for selected patients who are likely to carry genetic traits associated with hereditary carcinoma development. According to the US Preventive Services Task Force, BRCA testing should be offered to patients meeting any of the following criteria [122]:

- Two first-degree relatives with breast cancer, one of whom received the diagnosis at age 50 years or younger.
- A combination of three or more first- or second-degree relatives with breast cancer regardless of age at diagnosis.
- A combination of both breast and ovarian cancer among first- and second-degree relatives.
- A first-degree relative with bilateral breast cancer.
- A combination of two or more first- or second-degree relatives with ovarian cancer regardless of age at diagnosis.
- A first- or second-degree relative with both breast and ovarian cancer at any age.
- A history of breast cancer in a male relative.
- For women of Ashkenazi Jewish heritage: any first-degree relative (or two second-degree relatives on the same side of the family) with breast or ovarian cancer.

There are also mathematical modeling approaches that estimate the risk for *BRCA* mutations according to anamnestic data. BRCA testing is recommended for patients whose risk assessment exceeds 10% in these tests. Examples are:

- IBIS Breast Cancer Risk Evaluation tool [123].
- Penn II Risk model [124].
- BRCAPRO [125].

In genetic counseling for hereditary breast or ovarian carcinoma, the IBIS Breast Cancer Risk Evaluation tool is particularly popular as it indicates the general probability of a genetic predisposition without being confined to *BRCA1/2* mutations. This is of particular relevance as an increased breast cancer risk can also be due to mutations in genes that are unrelated to *BRCA1/2*. These comparatively rare alterations include mutations in *TP53* (Li–Fraumeni syndrome, 70% of all carriers of the syndrome show mutant p53 which gives rise to various malignancies), *PTEN* (Cowden syndrome), STK11 (Peutz–Jeghers syndrome), and *CDH1* (hereditary diffuse gastric cancer). In addition, various laboratories can now also test for numerous genetic risk factors showing low-to-moderate penetrance with regard to breast cancer. These can be subdivided in three larger groups. Genes from the first group are functionally associated with *BRCA1/2*. Examples are *ATM* (ataxia telangiectasia mutated), *BARD1* (BRCA1-associated RING domain 1), *CHEK2* (cell cycle checkpoint kinase 2), *MRE11A* (meiotic recombination 11 homolog A), NBN (nibrin), and *RAD50* and *RAD51D*. The second group contains genes from the "Fanconi anemia pathway:" *BRIP1* (BRCA-interacting protein C terminal helicase 1, *FANCJ*), *PALB2* (partner and localizer of *BRCA2*, *FANCN*), and *RAD51C* (*FANC0*). The third group finally comprises genes that are linked to hereditary colorectal carcinomas like *MLH1*, *MSH2*, *MSH6*, *PMS2*, *EPCAM* (also involved in some cases of Lynch syndrome, see above) and *MYH* (MYH-associated polyposis) [119].

Finally, Easton *et al.* showed in 2007 that single nucleotide polymorphisms (SNPs) play a role in breast cancer. In magnitude, their contribution appears similar that of low-penetrance mutations [126].

7.5.3
Emerging Diagnostics/Perspectives

Due to its high incidence and mortality, breast cancer is a central topic of oncological research. Consequently, it is impossible to cover all the manifold approaches for nucleic acid-based diagnostics in this chapter. Regarding gene signatures, tests like Oncotype DX or MammaPrint may only stand at the beginning of the development. Thus, improved prognostic tools will likely become available. Second-generation tests for gene expression profiling are already awaiting market introduction for clinical routine diagnostics. To reduce the heterogeneity within the defined groups, several independent teams of researchers are searching for further distinct subtypes within the defined categories of luminal A and B, basal-like (triple-negative), and *HER2*-overexpressing breast cancer. Additional parameters that correlate with the course of the disease also likely deserve more attention. These include the expression of cell cycle regulatory proteins, the presence or absence of B cell or plasma cell metagenes in highly proliferative tumors, as well as immune cell infiltrates in the tumor microenvironment and prognostically relevant gene expression patterns in the tumor stroma [114]. Consideration of these further aspects should help in tailoring the respective therapy to the needs of the individual patient.

Little is known with regard to predictive genetic parameters. There is no validated test available that could predict the response to a specific treatment. Empirically based attempts at selecting the most suitable chemotherapeutics for individual tumors were, at best, moderately successful – with a few exceptions like the "Topoisomerase II Alpha Gene Amplification and Protein Overexpression Predicting Efficacy of Epirubicin" (TOP) trial. In that study, Desmedt *et al.* [127] used three separate expression signatures (a *TOP2A* gene expression module and two previously published expression modules related to tumor microenvironment) to establish a numerical anthracycline-based score (A-Score) which could reliably predict a lack of sensitivity towards anthracycline-containing chemotherapy. The failure of other predictive algorithms might be due to the influence of further genetic discriminators that are not covered by the present signatures. Deep sequencing approaches thus hold considerable promise to overcome this limitation.

Many tests also neglect epigenetic modifications that can also affect the expression of drug targets. Accordingly, drugs that would seem immensely suited to target the genetic alterations in a specific tumor may become ineffective due to epigenetic regulations. Attempts to use epigenetic markers or modifiers to obtain prognostic predictions are thus ongoing. Based on data from 1302 breast cancer patients, miRNA profiling appears suitable to identify prognostic subgroups of breast cancer patients [128]. This subclassification was particularly useful for standard breast cancer subtypes without detectable copy number events. In this collective, epigenetic deregulations on the miRNA level appeared to be characteristic for oncogenic transformation and for the shaping of the immune contexture [128]. A further epigenetic cause for altered gene expression in breast cancer is DNA methylation (see Section 7.2.3). As several studies showed, *BRCA1* is inactivated by methylation in 10–15% of sporadic breast cancers. Consequently, these tumors resemble those caused by *BRCA* mutations and may also be sensitive to PARP inhibitors. Other breast cancer genes known to be susceptible to methylation-dependent regulation include *RAD51C*, *PTEN*, and *TP53* [129]. Accordingly, DNA methylation or miRNA expression patterns might also help to select the most appropriate treatment strategies.

Given the progress in the field, more complete coverage of the tumor genome and consideration of epigenetic changes no longer represent technical challenges, and will likely be implemented into clinical routine diagnostics. A problem that will be much harder to overcome, however, is tumor heterogeneity. Clinically, many treatments show good effects on the bulk of the tumor cell population, but spare individual clones that can then give rise to recurrence. Considering that these resistant subclones contribute little to the overall gene signature [114], they may not become immediately apparent upon gene expression profiling. Nevertheless, these cells, which likely represent the so-called tumor-initiating, metastasis-initiating, or cancer stem cells, might be decisive for the individual outcome. Using an 11-gene signature based on stem cell-associated genes, a "death from cancer" signature has thus been proposed [130]. Still, an

improved biological understanding of these cells will be required to guide the development of suitable prognostic tests that could add to the existing repertoire of nucleic acid-based diagnostics.

7.6 Conclusion

Nucleic acid-based diagnostics is well established in breast and gynecological cancers. This includes testing for genetic cancer risk and for HPV infection, and for patient stratification according to the presence of therapeutic targets such as *HER2*. Other tests for screening or treatment selection are still at an experimental stage, but are emerging. Finally, the relative ease of obtaining information from cancer-associated nucleic acid sequences predestines this approach for tackling the major diagnostic challenges of the future. In our opinion, these will be the establishment or optimization of screening tests for early detection, the assessment of the required treatment intensity, and the individualized prediction of response to a specific treatment. From both a medical and socio-economic perspective, an immense benefit would result if over- and undertreatment could be minimized and if patients could be spared from futile therapeutic approaches with all the associated side-effects. Moreover, early diagnosis and the selection of an individually optimized treatment scheme could be life-saving. Thus, there is ample justification for further developments in this promising field.

References

1 Jemal, A., Bray, F., Center, M.M., Ferlay, J., Ward, E., and Forman, D. (2011) Global cancer statistics. *CA: A Cancer Journal for Clinicians*, **61**, 69–90.

2 Lehtinen, M., Paavonen, J., Wheeler, C. M., Jaisamrarn, U., Garland, S.M., Castellsague, X., Skinner, S.R., Apter, D., Naud, P., Salmeron, J. *et al.* (2012) Overall efficacy of HPV-16/18 AS04-adjuvanted vaccine against grade 3 or greater cervical intraepithelial neoplasia: 4-year end-of-study analysis of the randomised, double-blind PATRICIA trial. *Lancet Oncology*, **13**, 89–99.

3 Wheeler, C.M., Castellsague, X., Garland, S.M., Szarewski, A., Paavonen, J., Naud, P., Salmeron, J., Chow, S.N., Apter, D., Kitchener, H. *et al.* (2012) Cross-protective efficacy of HPV-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by non-vaccine oncogenic HPV types: 4-year end-of-study analysis of the randomised, double-blind PATRICIA trial. *Lancet Oncology*, **13**, 100–110.

4 Walboomers, J.M., Jacobs, M.V., Manos, M.M., Bosch, F.X., Kummer, J.A., Shah, K.V., Snijders, P.J., Peto, J., Meijer, C.J., and Munoz, N. (1999) Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *Journal of Pathology*, **189**, 12–19.

5 Castellsague, X., Bosch, F.X., Munoz, N., Meijer, C.J., Shah, K.V., de Sanjose, S., Eluf-Neto, J., Ngelangel, C.A., Chichareon, S., Smith, J.S. *et al.* (2002) Male circumcision, penile human papillomavirus infection, and cervical cancer in female partners. *New England Journal of Medicine*, **346**, 1105–1112.

6 International Collaboration of Epidemiological Studies of Cervical

Cancer (2007) Comparison of risk factors for invasive squamous cell carcinoma and adenocarcinoma of the cervix: collaborative reanalysis of individual data on 8,097 women with squamous cell carcinoma and 1,374 women with adenocarcinoma from 12 epidemiological studies. *International Journal of Cancer*, **120**, 885–891.

7 Matorras, R., Ariceta, J.M., Rementeria, A., Corral, J., Gutierrez de Teran, G., Diez, J., Montoya, F., and Rodriguez-Escudero, F.J. (1991) Human immunodeficiency virus-induced immunosuppression: a risk factor for human papillomavirus infection. *American Journal of Obstetrics and Gynecology*, **164**, 42–44.

8 Wallin, K.L., Wiklund, F., Luostarinen, T., Angstrom, T., Anttila, T., Bergman, F., Hallmans, G., Ikaheimo, I., Koskela, P., Lehtinen, M. *et al.* (2002) A population-based prospective study of *Chlamydia trachomatis* infection and cervical carcinoma. *International Journal of Cancer*, **101**, 371–374.

9 Jemal, A., Simard, E.P., Dorell, C., Noone, A.-M., Markowitz, L.E., Kohler, B., Eheman, C., Saraiya, M., Bandi, P., Saslow, D. *et al.* (2013) Annual Report to the Nation on the Status of Cancer, 1975–2009, featuring the burden and trends in human papillomavirus (HPV)-associated cancers and HPV vaccination coverage levels. *Journal of the National Cancer Institute*, **105**, 175–201.

10 Singh, G., Miller, B., Hankey, B., and Edwards, B. (2003) *Area Socioeconomic Variations in US Cancer Incidence, Mortality, Stage, Treatment, and Survival, 1975–1999*, National Cancer Institute, Bethesda, MD.

11 Cubie, H.A. (2013) Diseases associated with human papillomavirus infection. *Virology*, **445**, 21–34.

12 Schiffman, M., Castle, P.E., Jeronimo, J., Rodriguez, A.C., and Wacholder, S. (2007) Human papillomavirus and cervical cancer. *Lancet*, **370**, 890–907.

13 Ries, L.A.G., Krapcho, M.D., Mariotto, M., Miller, A., Feuer, B.A., Clegg, E.J., Horner, L., Howlader, M.J., Eisner, N., Reichman, M.P., and Edwards BK, M. (eds) (2007) *SEER Cancer Statistics Review*, National Cancer Institute, Bethesda, MD.

14 Deutsche Krebsgesellschaft eV and Deutsche Gesellschaft für Gynäkologie und Geburtshilfe (2008) *2k-Leitlinie – Diagnostik und Therapie des Zervixkarzinoms (Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften)*, Zuckschwerdt Verlag, Munich.

15 Bodelon, C., Madeleine, M.M., Voigt, L. F., and Weiss, N.S. (2009) Is the incidence of invasive vulvar cancer increasing in the United States? *Cancer Causes & Control*, **20**, 1779–1782.

16 Siegel, R., Naishadham, D., and Jemal, A. (2013) Cancer statistics, 2013. *CA: A Cancer Journal for Clinicians*, **63**, 11–30.

17 de Koning, M.N., Quint, W.G., and Pirog, E.C. (2008) Prevalence of mucosal and cutaneous human papillomaviruses in different histologic subtypes of vulvar carcinoma. *Modern Pathology*, **21**, 334–344.

18 Monk, B.J., Burger, R.A., Lin, F., Parham, G., Vasilev, S.A., and Wilczynski, S.P. (1995) Prognostic significance of human papillomavirus DNA in vulvar carcinoma. *Obstetrics and Gynecology*, **85**, 709–715.

19 Deutsche Krebsgesellschaft eV and Deutsche Gesellschaft für Gynäkologie und Geburtshilfe (2009) *Interdisziplinäre S2k-Leitlinie für die Diagnostik und Therapie des Vulvakarzinoms und seiner Vorstufen*, Zuckschwerdt Verlag, Munich.

20 Daling, J.R., Madeleine, M.M., Schwartz, S.M., Shera, K.A., Carter, J.J., McKnight, B., Porter, P.L., Galloway, D.A., McDougall, J.K., and Tamimi, H. (2002) A population-based study of squamous cell vaginal cancer: HPV and cofactors. *Gynecologic Oncology*, **84**, 263–270.

21 Tjalma, W.A., Monaghan, J.M., de Barros Lopes, A., Naik, R., Nordin, A.J., and Weyler, J.J. (2001) The role of surgery in invasive squamous carcinoma of the vagina. *Gynecologic Oncology*, **81**, 360–365.

22 Papanicolaou, G.N. and Traut, H.F. (1997) The diagnostic value of vaginal smears in carcinoma of the uterus. 1941.

Archives of Pathology & Laboratory Medicine, **121**, 211–224.

23 Tilston, P. (1997) Anal human papillomavirus and anal cancer. *Journal of Clinical Pathology*, **50**, 625–634.

24 Muñoz, N., Bosch, F.X., de Sanjosé, S., Herrero, R., Castellsagué, X., Shah, K.V., Snijders, P.J.F., and Meijer, C.J.L.M. (2003) Epidemiologic classification of human papillomavirus types associated with cervical cancer. *New England Journal of Medicine*, **348**, 518–527.

25 Palefsky, J.M. (1994) Anal human papillomavirus infection and anal cancer in HIV-positive individuals: an emerging problem. *AIDS*, **8**, 283–295.

26 Palefsky, J.M. and Holly, E.A. (1995) Molecular virology and epidemiology of human papillomavirus and cervical cancer. *Cancer Epidemiology, Biomarkers & Prevention*, **4**, 415–428.

27 Munger, K., Phelps, W.C., Bubb, V., Howley, P.M., and Schlegel, R. (1989) The E6 and E7 genes of the human papillomavirus type 16 together are necessary and sufficient for transformation of primary human keratinocytes. *Journal of Virology*, **63**, 4417–4421.

28 Schiffman, M., Wentzensen, N., Wacholder, S., Kinney, W., Gage, J.C., and Castle, P.E. (2011) Human papillomavirus testing in the prevention of cervical cancer. *Journal of the National Cancer Institute*, **103**, 368–383.

29 Iftner, T.V. (2013) *Liste kommerzieller HPV-Tests*, Projektgruppe Zervita GbR, Tübingen.

30 Poljak, M., Ostrbenk, A., Seme, K., Ucakar, V., Hillemanns, P., Bokal, E.V., Jancar, N., and Klavs, I. (2011) Comparison of clinical and analytical performance of the Abbott RealTime High Risk HPV test to the performance of hybrid capture 2 in population-based cervical cancer screening. *Journal of Clinical Microbiology*, **49**, 1721–1729.

31 Jones, J., Powell, N.G., Tristram, A., Fiander, A.N., and Hibbitts, S. (2009) Comparison of the PapilloCheck DNA micro-array Human Papillomavirus detection assay with Hybrid Capture II and PCR-enzyme immunoassay using the GP5/6+ primer set. *Journal of Clinical Virology*, **45**, 100–104.

32 Stevens, M.P., Garland, S.M., Rudland, E., Tan, J., Quinn, M.A., and Tabrizi, S.N. (2007) Comparison of the Digene Hybrid Capture 2 assay and Roche AMPLICOR and LINEAR ARRAY human papillomavirus (HPV) tests in detecting high-risk HPV genotypes in specimens from women with previous abnormal Pap smear results. *Journal of Clinical Microbiology*, **45**, 2130–2137.

33 Kjaer, S., Hogdall, E., Frederiksen, K., Munk, C., van den Brule, A., Svare, E., Meijer, C., Lorincz, A., and Iftner, T. (2006) The absolute risk of cervical abnormalities in high-risk human papillomavirus-positive, cytologically normal women over a 10-year period. *Cancer Research*, **66**, 10630–10636.

34 Wright, T.C.Jr., Cox, J.T., Massad, L.S., Twiggs, L.B., Wilkinson, E.J., and Conference, A.S.-S.C. (2002) 2001 Consensus Guidelines for the management of women with cervical cytological abnormalities. *JAMA*, **287**, 2120–2129.

35 Nobbenhuis, M.A., Meijer, C.J., van den Brule, A.J., Rozendaal, L., Voorhorst, F.J., Risse, E.K., Verheijen, R.H., and Helmerhorst, T.J. (2001) Addition of high-risk HPV testing improves the current guidelines on follow-up after treatment for cervical intraepithelial neoplasia. *British Journal of Cancer*, **84**, 796–801.

36 Khan, M.J., Castle, P.E., Lorincz, A.T., Wacholder, S., Sherman, M., Scott, D.R., Rush, B.B., Glass, A.G., and Schiffman, M. (2005) The elevated 10-year risk of cervical precancer and cancer in women with human papillomavirus (HPV) type 16 or 18 and the possible utility of type-specific HPV testing in clinical practice. *Journal of the National Cancer Institute*, **97**, 1072–1079.

37 Deutsche Gesellschaft für Gynäkologie und Geburtshilfe, Arbeitsgemeinschaft Infektiologie und Infektimmunologie der DGGG, Berufsverband der Frauenärzte eV, Deutsche Gesellschaft für Pathologie eV, Deutsche Gesellschaft für Urologie eV, Deutsche

Krebsgesellschaft eV, Deutsche STD-Gesellschaft eV. and Frauenselbsthilfe nach Krebs eV (2010) *Prävention, Diagnostik und Therapie der HPV-Infektion und präinvasiver Läsionen des weiblichen Genitales (Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF))*, Zuckschwerdt Verlag, Munich.

38 Cuzick, J., Bergeron, C., von Knebel Doeberitz, M., Gravitt, P., Jeronimo, J., Lorincz, A.T., Meijer, C., Sankaranarayanan, R., JF Snijders, P., and Szarewski, A. (2012) New technologies and procedures for cervical cancer screening. *Vaccine*, **30** (Suppl. 5), F107–F116.

39 Clarke, M.A., Wentzensen, N., Mirabello, L., Ghosh, A., Wacholder, S., Harari, A., Lorincz, A., Schiffman, M., and Burk, R. D. (2012) Human papillomavirus DNA methylation as a potential biomarker for cervical cancer. *Cancer Epidemiology, Biomarkers & Prevention*, **21**, 2125–2137.

40 Krueger, F., Kreck, B., Franke, A., and Andrews, S.R. (2012) DNA methylome analysis using short bisulfite sequencing data. *Nature Methods*, **9**, 145–151.

41 Herman, J.G., Graff, J.R., Myohanen, S., Nelkin, B.D., and Baylin, S.B. (1996) Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 9821–9826.

42 Adorjan, P., Distler, J., Lipscher, E., Model, F., Muller, J., Pelet, C., Braun, A., Florl, A.R., Gutig, D., Grabs, G. *et al.* (2002) Tumour class prediction and discovery by microarray-based DNA methylation analysis. *Nucleic Acids Research*, **30**, e21.

43 Denschlag, D., Ulrich, U., and Emons, G. (2010) The diagnosis and treatment of endometrial cancer: progress and controversies. *Deutsches Ärzteblatt International*, **108**, 571–577.

44 Wright, J.D., Barrena Medel, N.I., Sehouli, J., Fujiwara, K., and Herzog, T.J. (2012) Contemporary management of endometrial cancer. *Lancet*, **379**, 1352–1360.

45 Smith, R.A., von Eschenbach, A.C., Wender, R., Levin, B., Byers, T., Rothenberger, D., Brooks, D., Creasman, W., Cohen, C., Runowicz, C. *et al.* (2001) American Cancer Society guidelines for the early detection of cancer: update of early detection guidelines for prostate, colorectal, and endometrial cancers. Also: update 2001 – testing for early lung cancer detection. *CA: A Cancer Journal for Clinicians*, **51**, 38–75, quiz 77–80.

46 Creasman, W.T., Odicino, F., Maisonneuve, P., Quinn, M.A., Beller, U., Benedet, J.L., Heintz, A.P., Ngan, H.Y., and Pecorelli, S. (2006) Carcinoma of the corpus uteri. FIGO 26th Annual Report on the Results of Treatment in Gynecological Cancer. *International Journal of Gynaecology and Obstetrics*, **95** (Suppl. 1), S105–143.

47 Lindor, N.M., Petersen, G.M., Hadley, D. W., Kinney, A.Y., Miesfeldt, S., Lu, K.H., Lynch, P., Burke, W., and Press, N. (2006) Recommendations for the care of individuals with an inherited predisposition to Lynch syndrome: a systematic review. *JAMA*, **296**, 1507–1517.

48 Oki, E., Oda, S., Maehara, Y., and Sugimachi, K. (1999) Mutated gene-specific phenotypes of dinucleotide repeat instability in human colorectal carcinoma cell lines deficient in DNA mismatch repair. *Oncogene*, **18**, 2143–2147.

49 Goel, A., Nagasaka, T., Hamelin, R., and Boland, C.R. (2010) An optimized pentaplex PCR for detecting DNA mismatch repair-deficient colorectal cancers. *PLoS One*, **5**, e9393.

50 Jass, J.R., Cottier, D.S., Jeevaratnam, P., Pokos, V., Holdaway, K.M., Bowden, M. L., Van de Water, N.S., and Browett, P.J. (1995) Diagnostic use of microsatellite instability in hereditary non-polyposis colorectal cancer. *Lancet*, **346**, 1200–1201.

51 Lindor, N.M., Petersen, G.M., Hadley, D. W. *et al.* (2006) Recommendations for the care of individuals with an inherited predisposition to Lynch syndrome: a systematic review. *JAMA* **296**, 1507–1517.

52 Matias-Guiu, X. and Prat, J. (2013) Molecular pathology of endometrial carcinoma. *Histopathology*, **62**, 111–123.

53 He, L. and Hannon, G.J. (2004) MicroRNAs: small RNAs with a big role in gene regulation. *Nature Reviews Genetics*, **5**, 522–531.

54 Chen, C., Ridzon, D.A., Broomer, A.J., Zhou, Z., Lee, D.H., Nguyen, J.T., Barbisin, M., Xu, N.L., Mahuvakar, V.R., Andersen, M.R. *et al.* (2005) Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Research*, **33**, e179.

55 Vorwerk, S., Ganter, K., Cheng, Y., Hoheisel, J., Stahler, P.F., and Beier, M. (2008) Microfluidic-based enzymatic on-chip labeling of miRNAs. *New Biotechnology*, **25**, 142–149.

56 Karaayvaz, M., Zhang, C., Liang, S., Shroyer, K.R., and Ju, J. (2012) Prognostic significance of miR-205 in endometrial cancer. *PLoS One*, **7**, e35158.

57 Perez, D.S., Hoage, T.R., Pritchett, J.R., Ducharme-Smith, A.L., Halling, M.L., Ganapathiraju, S.C., Streng, P.S., and Smith, D.I. (2008) Long, abundantly expressed non-coding transcripts are altered in cancer. *Human Molecular Genetics*, **17**, 642–655.

58 Clarke-Pearson, D.L. (2009) Clinical practice. Screening for ovarian cancer. *New England Journal of Medicine*, **361**, 170–177.

59 Kurman, R., Ellenson, L.H., and Ronnett, B.M. (2011) *Blaustein's Pathology of the Female Genital Tract*, 6th edn, Springer, Berlin.

60 Casagrande, J.T., Louie, E.W., Pike, M.C., Roy, S., Ross, R.K., and Henderson, B.E. (1979) "Incessant ovulation" and ovarian cancer. *Lancet*, **2**, 170–173.

61 Fathalla, M.F. (1971) Incessant ovulation – a factor in ovarian neoplasia? *Lancet*, **2**, 163.

62 Cramer, D.W. and Welch, W.R. (1983) Determinants of ovarian cancer risk. II. Inferences regarding pathogenesis. *Journal of the National Cancer Institute*, **71**, 717–721.

63 Dietl, J. and Wischhusen, J. (2011) The forgotten fallopian tube. *Nature Reviews Cancer*, **11**, 227, author reply 227.

64 Dietl, J., Wischhusen, J., and Häusler, S.F. (2011) The post-reproductive Fallopian tube: better removed? *Human Reproduction*, **26**, 2918–2924.

65 Flesken-Nikitin, A., Hwang, C.I., Cheng, C.Y., Michurina, T.V., Enikolopov, G., and Nikitin, A.Y. (2013) Ovarian surface epithelium at the junction area contains a cancer-prone stem cell niche. *Nature*, **495**, 241–245.

66 Chen, S. and Parmigiani, G. (2007) Meta-analysis of BRCA1 and BRCA2 penetrance. *Journal of Clinical Oncology*, **25**, 1329–1333.

67 Chittenden, B.G., Fullerton, G., Maheshwari, A., and Bhattacharya, S. (2009) Polycystic ovary syndrome and the risk of gynaecological cancer: a systematic review. *Reproductive Biomedicine Online*, **19**, 398–405.

68 Garg, P.P., Kerlikowske, K., Subak, L., and Grady, D. (1998) Hormone replacement therapy and the risk of epithelial ovarian carcinoma: a meta-analysis. *Obstetrics and Gynecology*, **92**, 472–479.

69 Gates, M.A., Rosner, B.A., Hecht, J.L., and Tworoger, S.S. (2010) Risk factors for epithelial ovarian cancer by histologic subtype. *American Journal of Epidemiology*, **171**, 45–53.

70 Ness, R.B., Cramer, D.W., Goodman, M. T., Kjaer, S.K., Mallin, K., Mosgaard, B.J., Purdie, D.M., Risch, H.A., Vergona, R., and Wu, A.H. (2002) Infertility, fertility drugs, and ovarian cancer: a pooled analysis of case-control studies. *American Journal of Epidemiology*, **155**, 217–224.

71 Stewart, L.M., Holman, C.D., Aboagye-Sarfo, P., Finn, J.C., Preen, D.B., and Hart, R. (2013) *In vitro* fertilization, endometriosis, nulliparity and ovarian cancer risk. *Gynecologic Oncology*, **128**, 260–264.

72 Sueblinvong, T. and Carney, M.E. (2009) Current understanding of risk factors for ovarian cancer. *Current Treatment Options in Oncology*, **10**, 67–81.

73 Tsilidis, K.K., Allen, N.E., Key, T.J., Dossus, L., Lukanova, A., Bakken, K., Lund, E., Fournier, A., Overvad, K., Hansen, L. *et al.* (2011) Oral contraceptive use and reproductive factors and risk of ovarian cancer in the

European Prospective Investigation into Cancer and Nutrition. *British Journal of Cancer*, **105**, 1436–1442.

74 Van Gorp, T., Amant, F., Neven, P., Vergote, I., and Moerman, P. (2004) Endometriosis and the development of malignant tumours of the pelvis. A review of literature. *Best Practice & Research Clinical Obstetrics & Gynaecology*, **18**, 349–371.

75 Goff, B.A., Matthews, B.J., Wynn, M., Muntz, H.G., Lishner, D.M., and Baldwin, L.M. (2006) Ovarian cancer: patterns of surgical care across the United States. *Gynecologic Oncology*, **103**, 383–390.

76 Burger, R.A., Brady, M.F., Bookman, M. A., Fleming, G.F., Monk, B.J., Huang, H., Mannel, R.S., Homesley, H.D., Fowler, J., Greer, B.E. *et al.* (2011) Incorporation of bevacizumab in the primary treatment of ovarian cancer. *New England Journal of Medicine*, **365**, 2473–2483.

77 Hall, M., Gourley, C., McNeish, I., Ledermann, J., Gore, M., Jayson, G., Perren, T., Rustin, G., and Kaye, S. (2013) Targeted anti-vascular therapies for ovarian cancer: current evidence. *British Journal of Cancer*, **108**, 250–258.

78 Curiel, T.J., Coukos, G., Zou, L., Alvarez, X., Cheng, P., Mottram, P., Evdemon-Hogan, M., Conejo-Garcia, J.R., Zhang, L., Burow, M. *et al.* (2004) Specific recruitment of regulatory T cells in ovarian carcinoma fosters immune privilege and predicts reduced survival. *Nature Medicine*, **10**, 942–949.

79 Zhang, L., Conejo-Garcia, J.R., Katsaros, D., Gimotty, P.A., Massobrio, M., Regnani, G., Makrigiannakis, A., Gray, H., Schlienger, K., Liebman, M.N. *et al.* (2003) Intratumoral T cells, recurrence, and survival in epithelial ovarian cancer. *New England Journal of Medicine*, **348**, 203–213.

80 Weissman, S.M., Weiss, S.M., and Newlin, A.C. (2012) Genetic testing by cancer site: ovary. *Cancer Journal*, **18**, 320–327.

81 Hearle, N., Schumacher, V., Menko, F.H., Olschwang, S., Boardman, L.A., Gille, J.J., Keller, J.J., Westerman, A.M., Scott, R.J., Lim, W. *et al.* (2006) Frequency and spectrum of cancers in the Peutz–Jeghers syndrome. *Clinical Cancer Research*, **12**, 3209–3215.

82 Häusler, S.F., Keller, A., Chandran, P.A., Ziegler, K., Zipp, K., Heuer, S., Krockenberger, M., Engel, J.B., Honig, A., Scheffler, M. *et al.* (2010) Whole blood-derived miRNA profiles as potential new tools for ovarian cancer screening. *British Journal of Cancer*, **3**, 3.

83 Keller, A., Leidinger, P., Bauer, A., Elsharawy, A., Haas, J., Backes, C., Wendschlag, A., Giese, N., Tjaden, C., Ott, K. *et al.* (2011) Toward the blood-borne miRNome of human diseases. *Nature Methods*, **8**, 841–843.

84 R Development Core Team (2008) *R: A Language and Environment for Statistical Computing*, vol. **3**, R Foundation for Statistical Computing, Vienna.

85 Foulkes, W.D. (2008) Inherited susceptibility to common cancers. *New England Journal of Medicine*, **359**, 2143–2153.

86 Beral, V. (2003) Breast cancer and hormone-replacement therapy in the Million Women Study. *Lancet*, **362**, 419–427.

87 Collaborative Group on Hormonal Factors in Breast Cancer (2002) Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50302 women with breast cancer and 96973 women without the disease. *Lancet*, **360**, 187–195.

88 Horn, J., Asvold, B.O., Opdahl, S., Tretli, S., and Vatten, L.J. (2013) Reproductive factors and the risk of breast cancer in old age: a Norwegian cohort study. *Breast Cancer Research and Treatment*, **139**, 237–243.

89 Sheean, P.M., Hoskins, K., and Stolley, M. (2012) Body composition changes in females treated for breast cancer: a review of the evidence. *Breast Cancer Research and Treatment*, **135**, 663–680.

90 Benson, J.R., Jatoi, I., Keisch, M., Esteva, F.J., Makris, A., and Jordan, V.C. (2009) Early breast cancer. *Lancet*, **373**, 1463–1479.

91 Deutsche Krebsgesellschaft eV and Deutsche Gesellschaft für Gynäkologie und Geburtshilfe (2012) *Interdisziplinäre*

S3-Leitlinie für die Diagnostik, Therapie und Nachsorge des Mammakarzinoms (Arbeitsgemeinschaft der wissenschaftlichen medizinischen Fachgesellschaften (AWMF), Deutsche Krebsgesellschaft eV, Deutsche Krebshilfe), Zuckschwerdt Verlag, Munich.

92 Giuliano, A.E., Hunt, K.K., Ballman, K.V., Beitsch, P.D., Whitworth, P.W., Blumencranz, P.W., Leitch, A.M., Saha, S., McCall, L.M., and Morrow, M. (2011) Axillary dissection vs no axillary dissection in women with invasive breast cancer and sentinel node metastasis: a randomized clinical trial. *JAMA*, **305**, 569–575.

93 Reim, F., Dombrowski, Y., Ritter, C., Buttmann, M., Hausler, S., Ossadnik, M., Krockenberger, M., Beier, D., Beier, C.P., Dietl, J. *et al.* (2009) Immunoselection of breast and ovarian cancer cells with trastuzumab and natural killer cells: selective escape of $CD44^{high}/CD24^{low}/HER2^{low}$ breast cancer stem cells. *Cancer Research*, **69**, 8058–8066.

94 Swain, S.M., Kim, S.B., Cortes, J., Ro, J., Semiglazov, V., Campone, M., Ciruelos, E., Ferrero, J.M., Schneeweiss, A., Knott, A. *et al.* (2013) Pertuzumab, trastuzumab, and docetaxel for HER2-positive metastatic breast cancer (CLEOPATRA study): overall survival results from a randomised, double-blind, placebo-controlled, phase 3 study. *Lancet Oncology*, **14**, 461–471.

95 Verma, S., Miles, D., Gianni, L., Krop, I. E., Welslau, M., Baselga, J., Pegram, M., Oh, D.Y., Dieras, V., Guardino, E. *et al.* (2012) Trastuzumab emtansine for HER2-positive advanced breast cancer. *New England Journal of Medicine*, **367**, 1783–1791.

96 King, C.R., Kraus, M.H., and Aaronson, S. A. (1985) Amplification of a novel v-erbB-related gene in a human mammary carcinoma. *Science*, **229**, 974–976.

97 Owens, M.A., Horten, B.C., and Da Silva, M.M. (2004) HER2 amplification ratios by fluorescence *in situ* hybridization and correlation with immunohistochemistry in a cohort of 6556 breast cancer tissues. *Clinical Breast Cancer*, **5**, 63–69.

98 Petit, A.M., Rak, J., Hung, M.C., Rockwell, P., Goldstein, N., Fendly, B., and Kerbel, R.S. (1997) Neutralizing antibodies against epidermal growth factor and ErbB-2/neu receptor tyrosine kinases down-regulate vascular endothelial growth factor production by tumor cells *in vitro* and *in vivo*: angiogenic implications for signal transduction therapy of solid tumors. *American Journal of Pathology*, **151**, 1523–1530.

99 Geyer, C.E., Forster, J., Lindquist, D., Chan, S., Romieu, C.G., Pienkowski, T., Jagiello-Gruszfeld, A., Crown, J., Chan, A., Kaufman, B. *et al.* (2006) Lapatinib plus capecitabine for HER2-positive advanced breast cancer. *New England Journal of Medicine*, **355**, 2733–2743.

100 Hudis, C.A. (2007) Trastuzumab – mechanism of action and use in clinical practice. *New England Journal of Medicine*, **357**, 39–51.

101 Muss, H.B., Thor, A.D., Berry, D.A., Kute, T., Liu, E.T., Koerner, F., Cirrincione, C. T., Budman, D.R., Wood, W.C., Barcos, M. *et al.* (1994) c-erbB-2 expression and response to adjuvant therapy in women with node-positive early breast cancer. *New England Journal of Medicine*, **330**, 1260–1266.

102 Pritchard, K.I., Levine, M.N., and Tu, D. (2003) Neu/erbB-2 overexpression and response to hormonal therapy in premenopausal women in the adjuvant breast cancer setting: will it play in Peoria? Part II. *Journal of Clinical Oncology*, **21**, 399–400.

103 Harris, L., Fritsche, H., Mennel, R., Norton, L., Ravdin, P., Taube, S., Somerfield, M.R., Hayes, D.F., and Bast, R.C.Jr. (2007) American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *Journal of Clinical Oncology*, **25**, 5287–5312.

104 Ali, S.M., Carney, W.P., Esteva, F.J., Fornier, M., Harris, L., Kostler, W.J., Lotz, J.P., Luftner, D., Pichon, M.F., and Lipton, A. (2008) Serum HER-2/*neu* and relative resistance to trastuzumab-based therapy in patients with metastatic breast cancer. *Cancer*, **113**, 1294–1301.

105 Baselga, J., Tripathy, D., Mendelsohn, J., Baughman, S., Benz, C.C., Dantis, L., Sklarin, N.T., Seidman, A.D., Hudis, C.A., Moore, J. *et al.* (1996) Phase II study of weekly intravenous recombinant humanized anti-p185^{HER2} monoclonal antibody in patients with HER2/*neu*-overexpressing metastatic breast cancer. *Journal of Clinical Oncology*, **14**, 737–744.

106 Yaziji, H., Goldstein, L.C., Barry, T.S., Werling, R., Hwang, H., Ellis, G.K., Gralow, J.R., Livingston, R.B., and Gown, A.M. (2004) HER-2 testing in breast cancer using parallel tissue-based methods. *JAMA*, **291**, 1972–1977.

107 Tubbs, R.R., Pettay, J.D., Roche, P.C., Stoler, M.H., Jenkins, R.B., and Grogan, T.M. (2001) Discrepancies in clinical laboratory testing of eligibility for trastuzumab therapy: apparent immunohistochemical false-positives do not get the message. *Journal of Clinical Oncology*, **19**, 2714–2721.

108 Wolff, A.C., Hammond, M.E., Schwartz, J.N., Hagerty, K.L., Allred, D.C., Cote, R. J., Dowsett, M., Fitzgibbons, P.L., Hanna, W.M., Langer, A. *et al.* (2007) American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *Journal of Clinical Oncology*, **25**, 118–145.

109 Lipton, A., Leitzel, K., Ali, S.M., Demers, L., Harvey, H.A., Chaudri-Ross, H.A., Evans, D., Lang, R., Hackl, W., Hamer, P. *et al.* (2005) Serum HER-2/neu conversion to positive at the time of disease progression in patients with breast carcinoma on hormone therapy. *Cancer*, **104**, 257–263.

110 Goldhirsch, A., Wood, W.C., Coates, A. S., Gelber, R.D., Thurlimann, B., and Senn, H.J. (2011) Strategies for subtypes – dealing with the diversity of breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Annals of Oncology*, **22**, 1736–1747.

111 Goldhirsch, A., Winer, E.P., Coates, A.S., Gelber, R.D., Piccart-Gebhart, M., Thurlimann, B., and Senn, H.J. (2013) Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Annals of Oncology*, **24**, 2206–2223.

112 Dowsett, M., Sestak, I., Lopez-Knowles, E., Sidhu, K., Dunbier, A.K., Cowens, J. W., Ferree, S., Storhoff, J., Schaper, C., and Cuzick, J. (2013) Comparison of PAM50 risk of recurrence score with Oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. *Journal of Clinical Oncology*, **31**, 2783–2790.

113 Muller, B.M., Keil, E., Lehmann, A., Winzer, K.J., Richter-Ehrenstein, C., Prinzler, J., Bangemann, N., Reles, A., Stadie, S., Schoenegg, W. *et al.* (2013) The EndoPredict gene-expression assay in clinical practice – performance and impact on clinical decisions. *PLoS One*, **8**, e68252.

114 Reis-Filho, J.S. and Pusztai, L. (2011) Gene expression profiling in breast cancer: classification, prognostication, and prediction. *Lancet*, **378**, 1812–1823.

115 Schem, C., Wenners, A.S., Mackelenbergh, M.T., Heilmann, T., Mathiak, M., Jonat, W., and Mundhenke, C. (2013) Prädiktion beim mammakarzinom. *Gynäkologe*, **46**, 377–381.

116 Schorge, J.O., Modesitt, S.C., Coleman, R. L., Cohn, D.E., Kauff, N.D., Duska, L.R., and Herzog, T.J. (2010) SGO White Paper on ovarian cancer: etiology, screening and surveillance. *Gynecologic Oncology*, **119**, 7–17.

117 Thompson, D. and Easton, D.F. (2002) Cancer Incidence in BRCA1 mutation carriers. *Journal of the National Cancer Institute*, **94**, 1358–1365.

118 Cancer Research Campaign (CRC) Genetic Epidemiology Unit (1999) Cancer risks in BRCA2 mutation carriers. The Breast Cancer Linkage Consortium. *Journal of the National Cancer Institute*, **91**, 1310–1316.

119 Shannon, K.M. and Chittenden, A. (2012) Genetic testing by cancer site: breast. *Cancer Journal*, **18**, 310–319.

120 Walsh, T., Casadei, S., Coats, K.H., Swisher, E., Stray, S.M., Higgins, J., Roach, K.C., Mandell, J., Lee, M.K., Ciernikova, S. *et al.* (2006) Spectrum of mutations in BRCA1, BRCA2, CHEK2, and TP53 in families at high risk of breast cancer. *JAMA*, **295**, 1379–1388.

121 Wieacker, P. and Welling, B. (2008) *Indication Criteria for Diseases: Familial Breast/Ovary Cancer [BRCA1/BRCA2]: Indication Criteria for Genetic Testing – Evaluation of Validity and Clinical Utility*, German Society of Human Genetics, Münster.

122 US Preventive Services Task Force (2005) Genetic risk assessment and BRCA mutation testing for breast and ovarian cancer susceptibility: recommendation statement. *Annals of Internal Medicine*, **143**, 355–361.

123 Tyrer, J., Duffy, S.W., and Cuzick, J. (2004) A breast cancer prediction model incorporating familial and personal risk factors. *Statistics in Medicine*, **23**, 1111–1130.

124 Lindor, N.M., Johnson, K.J., Harvey, H., Shane Pankratz, V., Domchek, S.M., Hunt, K., Wilson, M., Cathie Smith, M., and Couch, F. (2010) Predicting BRCA1 and BRCA2 gene mutation carriers: comparison of PENN II model to previous study. *Familial Cancer*, **9**, 495–502.

125 Berry, D.A., Iversen, E.S.Jr., Gudbjartsson, D.F., Hiller, E.H., Garber, J.E., Peshkin, B. N., Lerman, C., Watson, P., Lynch, H.T., Hilsenbeck, S.G. *et al.* (2002) BRCAPRO validation, sensitivity of genetic testing of BRCA1/BRCA2, and prevalence of other breast cancer susceptibility genes. *Journal of Clinical Oncology*, **20**, 2701–2712.

126 Easton, D.F., Pooley, K.A., Dunning, A. M., Pharoah, P.D., Thompson, D., Ballinger, D.G., Struewing, J.P., Morrison, J., Field, H., Luben, R. *et al.* (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*, **447**, 1087–1093.

127 Desmedt, C., Di Leo, A., de Azambuja, E., Larsimont, D., Haibe-Kains, B., Selleslags, J., Delaloge, S., Duhem, C., Kains, J.P., Carly, B. *et al.* (2011) Multifactorial approach to predicting resistance to anthracyclines. *Journal of Clinical Oncology*, **29**, 1578–1586.

128 Dvinge, H., Git, A., Graf, S., Salmon-Divon, M., Curtis, C., Sottoriva, A., Zhao, Y., Hirst, M., Armisen, J., Miska, E.A. *et al.* (2013) The shaping and functional consequences of the microRNA landscape in breast cancer. *Nature*, **497**, 378–382.

129 Stefansson, O.A. and Esteller, M. (2013) Epigenetic modifications in breast cancer and their role in personalized medicine. *American Journal of Pathology*, **183**, 1052–1063.

130 Glinsky, G.V., Berezovska, O., and Glinskii, A.B. (2005) Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer. *Journal of Clinical Investigation*, **115**, 1503–1521.

8
Nucleic Acids as Molecular Diagnostics in Hematopoietic Malignancies – Implications in Diagnosis, Prognosis, and Therapeutic Management

Janine Schwamb and Christian P. Pallasch

8.1
Introduction

Hematological malignancies were among the first diseases in humans to be approached by molecular diagnostics using nucleic acids. In particular, disorders showing a leukemic phenotype were the first area to establish and introduce molecular diagnostics into clinical practice. Since the acquisition of leukocytes delivers directly the cellular correlate of the disease, isolation of nucleic acids for diagnostic purposes is easily accessible in terms of sufficient quantity and quality of samples. DNA originating from the altered cell populations delivers directly a full molecular representation of the origin of the disease – easily accessible by blood withdrawal from a peripheral vein. The field of hematological malignancies was a vanguard particularly in preclinical research and led to the identification of major insights in cancer biology. Similarly, translation of molecular findings into therapeutic principles led to significant improvements in clinical practice. Specific diseases such as chronic myeloid leukemia (CML) caused by one defined genetic lesion became model diseases for diagnostics and therapy development in the fields of hematology and oncology.

Molecular diagnostics has been established as a steadily growing pillar of diagnostics, replacing established classifications and criteria defined by morphology, cytochemistry, and cytogenetic aberrations at the chromosome level. All of these diagnostic tools were based on microscopic evaluation of transformed cells. However, in recent years the advent of a multitude of molecular biology approaches (as discussed in other chapters in this volume) has led to the introduction of DNA-based read-out systems assessing specific mutations, defined gene fusion products, and overexpression and amplification of malignancy-specific alterations.

Nucleic Acids as Molecular Diagnostics, First Edition. Edited by Andreas Keller and Eckart Meese.

8.2 Methodological Approaches

A multitude of molecular diagnostics can be performed in hematopoietic malignancies based on nucleic acids isolated from peripheral blood. However, this requires the presence of sufficient transformed cells in the peripheral blood. EDTA or heparin are usually used as anti-coagulants; samples can usually be shipped overnight at room temperature.

A more representative sample can be acquired by aspiration from bone marrow. This will, however, require more invasive specimen withdrawal from either the sternum or the iliac crest.

8.3 Cytogenetic Analysis to Molecular Diagnostics

Cytogenetic investigation plays an important role in the diagnostic and prognostic assessment of not only acute but also chronic leukemia. By means of classical chromosomal analysis as well as additional fluorescence *in situ* hybridization (FISH), structural as well as numeric chromosomal abnormalities can be detected. Metaphase cytogenetic analysis usually requires viable proliferating cells in order to obtain metaphase spreads, while uncultured cells can be used in interphase FISH. Similarly, in molecular diagnostics, cells do not have to be vital to be propagated *in vitro*. However, RNA-based diagnostics requires rapid processing of the samples.

The principle methods used for assessing molecular genetic alterations are polymerase chain reaction (PCR)-based detection, DNA Sanger sequencing, and, increasingly, next-generation sequencing (NGS)-based methods. In contrast to classical cytogenetic analysis, PCR is suitable for the detection of specific, already known changes in nucleic acids and not for any screening approach. Apart from that, quantitative real-time PCR (RT-PCR) represents the second part of molecular diagnostics, particularly for the detection and confirmation of already verified genetic changes. Treatment response as well as residual disease can be evaluated by quantification of particular fusion proteins via RT-PCR.

8.4 Minimal Residual Disease

Minimal residual disease (MRD) represents a major achievement in monitoring acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL). Quantification of malignant cells, mainly by PCR and RT-PCR, is based on the detection of molecular fusion proteins or, for ALL, verification of clonal rearrangement of immunoglobulin and T cell receptor genes in patient samples. Less commonly, MRD is performed using specific surface marker combinations

defining the transformed cellular population via flow cytometry. Molecular detection of specific markers enables determination of the leukemic burden at initial diagnosis, monitoring response during treatment, and, finally and almost most importantly, recognition of early relapse while clinical symptoms are not present or blood count is not yet heavily altered.

8.5 Chronic Myeloid Leukemia

The first established and most prominent molecular aberration was identified in CML. The discovery of the Philadelphia chromosome was a break-through as a model disease first rendering a specific cytogenetic aberration – the translocation t(9;22)(q34;q11) with the typical Philadelphia chromosome 22q– [1]. Due to this translocation, the gene on chromosome 9 coding for the ABL (Abelson murine leukemia viral oncogene) tyrosine kinase combines with the BCR (breakpoint cluster region) gene on chromosome 22, which finally results in expression of the fusion protein BCR–ABL1 with consecutive activity of tyrosine kinase [2–4]. Expression of the BCR–ABL1 protein directly propagates cellular proliferation and disrupts physiological differentiation in the hematopoietic stem cell, and is the driving force of oncogenic transformation towards a leukemic phenotype.

Assessment of BCR–ABL1 via PCR and initial quantification by RT-PCR is essential for the initial diagnosis of CML. Beyond that, subsequent monitoring of the BCR–ABL1 burden from peripheral blood forms part of the standard procedures in CML management. According to the European Leukemia Network, quantification of BCR–ABL1 is recommended every 3 months until a major molecular remission (BCR–ABL1/ABL1 $\leq$ 0.1) has been achieved. Later on, the frequency of measurements can reduced to every 6 months [5].

Specific genetic alterations affecting "druggable" genes offer novel options for targeted therapy in leukemic patients. Again, in CML as a model disease, the most remarkable change in treatment over the last decade was achieved by the introduction of the tyrosine kinase inhibitor imatinib. As mentioned above, due to translocation 9;22, the Philadelphia chromosome produces the oncogenic fusion protein BCR–ABL1 with subsequent increased activity of tyrosine kinase. Imatinib inhibits the overexpressed tyrosine kinase and consequently disables growth of malignant cells without affecting normal cells [6]. Therefore, imatinib has replaced the former standard therapy that used interferon or cytarabine [6]. Nevertheless, imatinib is not limited to BCR–ABL1 specificity, it also affects platelet-derived growth factor (PDGF) and c-kit tyrosine kinase receptors [6,7]. Following the great success of targeted therapy with imatinib, second-generation tyrosine kinase inhibitors were developed for treatment of CML with even greater efficacy. Nilotinib and dasatinib were shown to be equally effective in the treatment of CML [8,9], and to also act on PDGF and c-kit receptors [10]. Both substances

have been approved for first-line therapy of CML patients, but are also frequently used in relapsed disease.

Despite the fact that targeted therapy is gaining in importance in the treatment of leukemia, patients continue to relapse due to drug resistance. Thus, continuous monitoring of BCR–ABL in the peripheral blood is performed in order to detect relapse.

Treatment failure occurs due to acquisition of resistance towards imatinib, particularly in patients with advanced-stage disease [10]. Mechanisms of resistance are basically [11] point mutations in the kinase domain of the BCR–ABL enzyme that reduce sensitivity to imatinib [6,10,12]. Second-generation tyrosine kinase inhibitors such as dasatinib or nilotinib are usually able to overcome imatinib-mediated drug resistance [13]. Acquisition of resistance mutations is, however, associated with an unfavorable clinical course [6]. Recently presented studies with newly designed agents, in particular ponatinib, revealed good efficacy against BCR–ABL kinase with additional activity against T315I mutants [6]. Accordingly, ponatinib offers new therapeutic options beyond the current concepts, including allogeneic stem cell transplantation.

The development of specific mutations in treated CML enables the specific detection of the respective mutations. This has been extensively done accompanying clinical trials and in the follow-up of tyrosine kinase inhibitor-treated CML patients.

The routine use of mutational screening analysis by conventional Sanger sequencing is currently recommended in standard clinical use in the case of progression, failure, and clinical warning signs. NGS methods still are under evaluation such that conventional Sanger sequencing still remains the method of choice [5]. In particular, mutational analysis should be performed after failure of initial treatment, decreasing response to tyrosine kinase inhibitors, or during disease progression [5].

BCR–ABL1 kinase domain point mutations evolve in around 50% of patients showing treatment failure and progression [14]. The multitude of mutations has been assessed using Sanger sequencing. The presence of further mutations at lower levels can be identified using more sensitive techniques, such as mass spectrometry or ultra-deep sequencing; here, the relevance of rare clones and sample heterogeneity still has to be evaluated [15]. About 80 amino acid substitutions have been identified to be associated with the development of resistance to imatinib [14]. Dasatinib and nilotinib show less frequent acquisition of resistance mutations; however, they fail to inhibit the most prominent alteration resulting in the T315I mutation. Patients taking nilotinib were most frequently found to have acquired Y253H, E255K/V, F359V/C/I, or T315I mutations [16], whereas patients taking dasatinib were most frequently found to have acquired V299L, F317L/V/I/C, T315A, or T315I mutations [17].

Beyond that, BCR–ABL-independent resistance to imatinib by several heterogeneous mechanisms has also been described [6,10]. While additional cytogenetic alterations occur in imatinib-resistant patients [18], increased BCR–ABL levels due to either genomic amplification of the tyrosine kinase or overexpression of

BCR–ABL transcripts cause treatment failure [19]. In such cases, patients mostly respond to increased imatinib dosage [19]. Activation of alternative signaling pathways and overexpression of glycoproteins such as P-glycoprotein, leading to increased efflux of the drug or decreased imatinib uptake due to impaired expression of the drug transporter, also contribute to resistance to therapy [20].

In conclusion, CML is the model disease for the development of modern targeted therapies. Moreover, it is also a good example of molecular diagnostics analyzing nucleic acids from peripheral blood being incorporated in patient treatment and guiding clinical decisions.

8.6 Acute Myeloid Leukemia

AML is characterized by a rapid growth of leukocytes of the myeloid lineage, thus leading to accumulation of transformed cells in the bone marrow and subsequent suppression of normal hematopoiesis. Patients generally present with an elevated leukocyte count and further alterations of the peripheral blood, such as thrombocytopenia or anemia. Clinical symptoms may include fatigue, shortness of breath, bleeding, weight loss, fever, and increased risk of infection.

Diagnosis is usually obtained from cytomorphology of the bone marrow. Historically, AML has been further classified based on the French–American–British (FAB) cooperative group [21]. While this traditional classification is still in use, it has rapidly lost its prominent role to cytogenetic and molecular markers.

AML exhibits numerous different chromosomal aberrations, which define characteristic individual entities with typical clinical attributes. The importance of those chromosomal aberrations is also clarified by their inclusion in the current World Health Organization (WHO) classification system [22]. Inversion inv16(p13q22) with *CBFB–MYH11* gene rearrangement or the translocation t(8;21)(q22,q22) of fusion gene *RUNX1–RUNX1T1* (*AML1–ETO*) are common aberrations in AML, which are associated with the former FAB classifications M4eo and M1/M2, respectively [23]. Both anomalies are associated with a relatively good clinical prognosis. 11q23 abnormalities involving the *MLL* gene are, however, related to a worse clinical outcome. A special subtype of AML is so-called promyelocytic leukemia FAB M3 with the typical translocation t(15;17)(q22;q12). Due to targeted therapy with all-*trans* retinoic acid (ATRA), which will be discussed in detail below, the rate of complete remission has been improved in this leukemia subtype. In addition to these mostly structural chromosomal changes, the most common numeric aberration in AML is trisomy 8 with a so far unknown prognostic impact. In contrast to the single genetic alterations, karyotypes, that possess three or more structural or numeric chromosomal anomalies are called complex-aberrant and are associated with a worse clinical outcome.

Most chromosomal abnormalities influence patient prognosis and survival. For instance, a normal karyotype in AML patients is associated with a merely

intermediate prognosis, while translocation t(8;21)(q22,q22) or inversion inv(16)(p13q22) with expression of the fusion proteins *CBFB–MYH11* and *RUNX1–RUNX1T1* (*AML1–ETO*), respectively, contribute to a favorable clinical outcome [24]. In contrast, 11q23 abnormalities involving *MLL* gene rearrangement worsen survival of AML patients, especially when diagnosed in therapy-related AML [23]. As mentioned above, a complex-aberrant karyotype with three or more chromosomal alterations defines a high-risk profile with poor prognosis when allogeneic stem cell transplantation is not feasible.

Molecular diagnostics has became increasingly important for AML treatment due to minor genetic alterations such as point mutations, deletions, or duplications, which are not covered via classical chromosomal or FISH analysis, but which are relevant for prognosis and therapeutic decisions in AML patients. The most frequent genetic modification in approximately 50% of AML patients with a normal karyotype is mutation in the nucleophosmin gene (*NPM1*) on chromosome 5q31 [25]. Female patients seem to be more affected than male patients. Also associated with a normal karyotype, but less common, is mutation of the transcription factor CEBPα. Mutations in the FMS-like tyrosine kinase 3 gene (*FLT3*) at chromosomal region 13q12 are predominantly found in patients with AML or myelodysplastic syndrome [26]. Adults in particular are affected (around 30%), while it is rare in younger patients aged 10 years or below. Apart from that, the previously mentioned AML-specific translocations t(15;17), t(8;21), and inv(16)/t(16;16) with expression of the fusion genes *PML-RARa*, *RUNX1–RUNX1T1* (*AML1–ETO*) and *CBFB–MYH11*, respectively, are also analyzed and quantified according to routine diagnostic procedures asthey contribute to the evaluation of prognosis and offer possible targets for therapy [24]. Another specific aberration worth mentioning is the so-called partial tandem duplication (PTD) of the *MLL* gene. It is mainly associated with a normal karyotype and only measurable by PCR or Southern Blot. In addition to several other molecular anomalies described in AML, increased expression of the transcription factor WT1 (Wilms Tumor protein 1) may be seen in AML and is analyzed in routine diagnostics by RT-PCR [27].

In AML, MRD monitoring is also performed, although half of the patients do not posses any molecular target that can be quantified [28]. Quantification via RT-PCR has been established for several fusion transcripts, including *RUNX1–RUNX1T1*, *CBFB–MYH11*, *PML–RARa*, and *MLL–PTD*, as well as for gene mutations of *NPM1*, *FLT3*, or expression of WT1 transcription factor [28,29].

Acute promyelocyte leukemia, also known as AML FAB M3, is characterized by translocation of chromosome 17 with other genetic partners. Translocation 15;17 occurs in about 95% of AML M3 cases leading to fusion of the promyelocytic gene (*PML*) at chromosome 15 with the retinoid acid receptor-α (*RARa*) gene on chromosome 17. ATRA, as a non-cytostatic drug, is able to induce complete remission by inducing terminal differentiation of malignant promyelocytes into mature granulocytes [30]. Beyond that, arsenic trioxide (ATO) is becoming more and more established in the treatment of promyelocyte leukemia.

Resistance to ATRA has recently been described in acute promyelocyte leukemia. Mutations in the ligand-binding domain of RARa basically contribute to treatment failure of ATRA in AML M3 patients. To overcome drug resistance, patients receive ATO, resulting in a subsequent good clinical response [31]. However, several reports indicate resistance to ATO as well due to mutations in the PML-B2 domain in PML-RARa. At present, no approach for solving this clinical problem has emerged [31].

8.7 Acute Lymphocytic Leukemia

In addition to AML, ALL as the second form of acute blood cancer can be further divided into two subgroups deriving from either lymphatic B or T cell lineages. It occurs mainly in childhood, although it shows a more aggressive clinical behavior in adults [32]. ALL is characterized by a rapid growth of leukocytes of lymphoid lineage with a blastoid shape, thus leading to accumulation of transformed cells in the bone marrow and subsequent suppression of normal hematopoiesis. Patients generally present with an elevated leukocyte count and further alterations of the peripheral blood, such as thrombocytopenia or anemia. Clinical symptoms may include fatigue, shortness of breath, bleeding, weight loss, fever, and increased risk of infection.

The frequency of cytogenetic abnormalities depends on the affected cell population (B or T cells) and on the age of first diagnosis (children, adolescents, or adults). Numerous defined translocations rendering oncogenic fusion products also offer very specific targets for molecular diagnostics based on PCR amplification. As outlined above, the prominent chromosomal aberration in CML, translocation t(9;22)(q34;q11) producing the BCR–ABL fusion protein, is also found in approximately 25% of adult patients with ALL, while only 3–5% of children are affected [4,33]. Prevalence increases constantly with age. The common or pre-ALL subtypes generally possess this kind of genetic alteration, although it never occurs in T-ALL or mature B-ALL.

Accordingly, one of the most common gene fusions monitored in ALL patients is *BCR–ABL1*. Quantification of fusion genes is mostly carried out by RT-PCR on the mRNA level from peripheral blood or bone marrow. However, patient-specific clonal rearrangement of immunoglobulin or T cell receptor genes can be identified on the DNA level via classical PCR with subsequent sequencing of the PCR product. Following sequence-specific design of PCR primers, the individual leukemic clone is exclusively amplified and quantified [34]. Alternatively, clonal rearrangements can also be identified by their size, although with a decreased sensitivity [34].

Another frequent chromosomal abnormality is translocation t(4;11)(q21;q23) with the concomitant expression of the *MLL–AF4* fusion protein. It occurs in about 6% of adults and 4% of children, and represents the most common *MLL* translocation. The second most prevalent *MLL* translocation is t(11;19)(q23;

p13.3) with the fusion gene *MLL–ENL*. In addition to the genetic anomalies already mentioned, there are several others in ALL diagnostics, such as t(1;19)(q23;p13), 9q34-aberration, t(8;14)(q24;q32), or t(10;14)(q24;q11) as the most frequent translocation in T-ALL. Only detectable by FISH analysis and associated mostly with the common ALL subtype, translocation t(12;21)(p13;q22) is the most frequent aberration in young patients with a prevalence of 4 years; only 1–2% of adults possess this translocation, whereas patients aged 35 or older are not affected.

Molecular diagnosis in patients with ALL depends on the immunophenotype of the malignant cells with the according ALL subtype. As well as testing *BCR–ABL1* expression in ALL diagnosis for high-risk stratification, the two most common *MLL* fusion genes (i.e., *MLL–AF4* and *MLL–ENL*) should be evaluated as *MLL–AF4* in particular is considered as a further high-risk factor. Considering mature B-ALL, investigation of *MYC–IGH* fusion proteins should be carried out as a marker for aggressive disease progression and the leukemic phase of Burkitt's lymphoma. Depending on the suspected ALL subtype, other genetic abnormalities are detectable by molecular diagnosis, such as *E2A–PBX1* in pre-B-ALL or *ETV6–RUNX1* rearrangement (*TEL–AML1*) as the most common alteration in common ALL in pediatric patients [35].

Genetic modification is also evident in T-ALL in about 50% of patients. In most cases, rearrangements of T cell receptor loci are affected with the involvement of several transcription factors. Although quite rare in T-ALL and not detectable by classical cytogenetic investigation, fusion protein NUP214–ABL1 due to 9q34 aberration is relevant with regard to disease monitoring and targeted therapy [36].

While the BCR–ABL fusion protein is basically found in CML, its existence in ALL patients opens new strategies to add tyrosine kinase inhibitors to chemotherapy in order to avoid allogeneic stem cell transplantation [37]. Similar to Philadelphia chromosome-positive ALL, 11q23 chromosomal changes involving the *MLL* gene, particularly expression of the fusion gene *MLL–AF4*, contribute to poor prognosis [38].

8.8 Chronic Lymphocytic Leukemia

The most frequent leukemia in the western hemisphere is chronic lymphocytic leukemia (CLL). In contrast to ALL, CLL is characterized by the comparably slower clonal proliferation and accumulation of mature, typically $CD5^+$ B cells within the blood, bone marrow, lymph nodes, and spleen.

The clinical variance, ranging from a very slowly evolving to a rapidly progressing clinical appearance, is reflected in a multitude of prognostic markers in CLL [39].

The first molecular prognostic marker in CLL was established using Sanger sequencing of the *IGVH* locus of assessment of somatic hypermutations in the

variable region of the heavy immunoglobulin chain (IgVH) [39]. A cutoff value of 2% or less mutated alleles compared with its germline counterpart is considered as the unmutated IgVH condition [40]. Unmutated IgVH is considered to reflect a more immature differentiation of B cells and patients display a more rapidly progressing disease with worse prognosis [39]. Additionally, the hypermutational status of IgVH is associated with typical chromosomal aberrations and expression of prognostic markers, and also influences the clinical outcome of CLL patients [41]. It could be shown in many previous publications that an unmutated IgVH contributes to the aggressive course of disease [42,43]. One exceptional case is detection of IgVH gene 3-21, which contributes to poor prognosis independently of mutated IgVH [44]. In addition, cyto- and molecular genetic changes in CLL partly correlate with each other. For instance, unmutated IgVH mutational status is associated with deletion of 17p and 11q [41].

Although cytogenetic analysis is not necessarily required for diagnosis, it is recommended before initiating anti-leukemic therapy. The preferential method is FISH analysis instead of classical chromosomal investigation. The most relevant cytogenetic abnormality considering prognosis in CLL is deletion of the short arm of chromosome 17 (del17p13). It is further described in previous publications that this aberration is connected to a *TP53* mutation in the remaining *TP53* allele in the 17p13 band involved. Nevertheless, there is evidence that *TP53* mutation occurs independently of 17p deletion [45]. A similar chromosomal alteration in CLL is deletion of the long arm of chromosome 11 (del11q22.3), which targets the ataxia telangiectasia mutated (*ATM*) gene responsible for regulation of the cell cycle, including apoptosis and DNA repair by phosphorylation of certain proteins [46]. As a more common anomaly, trisomy 12 occurs in about 20–25% of CLL patients. The most common cytogenetic defect with more than 50% of patients being affected is deletion of the long arm of chromosome 13 (del13q14).

11q22.3 deletion affecting the *ATM* gene in a subset of patients [46] predicts advanced disease and shorter survival [47]. While 11q deletion has a median survival time of 79 months, loss of 17p results in a median survival time of only 32 months due to rapid disease progression [47], but also because of increased resistance to chemotherapy, especially purine analogs or alkylating agents [48,49]. Beyond that, there is strong evidence that the tumor suppressor gene *TP53* is affected by 17p deletion in the majority of patients [47]. However, *TP53* mutation also occurs in the absence of 17p deletion and presents as an independent factor for poor prognosis.

In addition to the cytogenetic alterations that directly affect the DNA damage pathway, other aberrations are associated with slower progressing disease. Deletion of chromosome 13q, for example, improves survival of CLL patients, whereas a normal karyotype or detection of trisomy 12 are considered as alterations with intermediate risk [47]. Deletion of the 13q14 locus affects the miR-15 and miR-16 loci, and has been linked to the pathogenesis of CLL [50]. Recently, improved *in vitro* propagation of CLL cells to obtain metaphase cytogenetics in CLL cells revealed translocations as a prognostic factor in CLL [51]; however,

the widespread occurrence of breakpoints does not offer a specific marker for PCR-based diagnostics of translocations in CLL.

Cytogenetic aberrations in standard clinical diagnostics should be identified prior to therapy onset of first-line treatment. In particular, patients harboring aberrations affecting *TP53* via mutation or del17p exhibit resistance towards immunochemotherapeutic approaches like the BR or FCR regimen [52,53]. Here, interphase FISH and Sanger sequencing of the *TP53* gene will be carried out from the peripheral blood of CLL patients once an indication for treatment onset has been reached. This genetic diagnostics directly affects therapeutic concepts identifying *TP53*-deficient patients for alternative therapeutic concepts such as allogeneic bone marrow transplantation or clinical trials. With regard to physically fit CLL patients, the standard immunochemotherapy regime is rituximab as an anti-CD20 antibody combined with fludarabine and cyclophosphamide [53]. In the case of 17p deletion or mutation of *TP53*, resistance against chemotherapy, particularly fludarabine, has been reported. Alemtuzumab, an anti-CD52-antibody, shows good response rates in high-risk CLL patients [54]; however, patients without serious comorbidities should be further considered for allogeneic stem cell transplantation to maintain progression-free survival [55].

In addition to cytogenetic aberrations, further research currently focuses on gene expression profiling or single nucleotide polymorphism (SNP) array analysis to identify variations in the human genome that might have a relevant impact on the diagnosis and prognosis of hematologic malignancies. In this context, whole-genome SNP mapping arrays could already demonstrate a clear prediction of time-to-treatment and overall survival in CLL patients independent of the well-established prognosis factors [56]. Nevertheless, additional studies need to be performed to further assess the clinical outcome of leukemic patients.

Recently published whole-genome sequencing projects in CLL have shown a number of recurrent somatic mutations, including *NOTCH1*, *MYD88*, *TP53*, *ATM*, *SF3B1*, *FBXW7*, *POT1*, *CHD2*, and others [57,58].

In CLL research, for instance, the NOTCH pathway comes into focus as a new additional target for CLL therapy. For T-ALL, activating mutations of the *NOTCH1* gene with oncogenic potential have been repeatedly described in about 50% of cases [59]. As a therapeutic approach for T-ALL, γ-secretase inhibitors (GSIs) in combination with glucocorticoids were reported to inhibit the *NOTCH1* pathway, to restore steroid sensitivity and to reduce gastrointestinal side-effects of GSIs by the use of glucocorticoid [60].

For B cell malignancies, the role of NOTCH is not fully understood. Both oncogenic and tumor suppressive functions have been considered [61]. For CLL, controversial studies have been published postulating the importance of increased *NOTCH2* expression as a pro-survival stimulus, while others found no relevant activation of *NOTCH1* or *NOTCH2* in CLL cells compared with healthy controls [61]. Nevertheless, there are hints of a constitutive NOTCH activation contributing to apoptosis resistance in malignant cells [61], in particular in CLL, where mutations of *NOTCH1* constitute an independent predictor of poor

survival [62]. Further investigation has focused on selective inhibition of each NOTCH receptor with an emphasis on *NOTCH1* receptors by anti-*NOTCH1* inhibitory antibodies or small-peptide inhibitors of NOTCH signaling [63].

Although *NOTCH1* mutations generally occur independently of *TP53* disruption [62], there are hints that *NOTCH1* signaling and *TP53* mutation are in part connected [59].

As outlined above, patients possessing a *TP53* mutation normally show a decreased response to chemotherapeutics, especially to fludarabine. Therefore, efforts are being made to develop therapeutic strategies that avoid the *TP53* pathway in order to unfold full cytotoxic activity. Parthenolide, an NF-κB inhibitor, flavopiridol, ABT-737, 2-phenylacetlyensulfonamide, or geldanamycin as a heat-shock 70 and 90 inhibitor are some of the substances that at least partially bypass *TP53* signaling [59]. Another possibility for targeting *TP53* is the compound Nutlin, a small molecule upregulating *TP53* through inhibition of the protein–protein interaction between p53 and MDM2 in wild-type *TP53* samples. It has already been tested in hematopoietic malignancies in first clinical trials with good preliminary results [64].

In addition to the classical chromosomal or molecular genetic alterations in clinical diagnostics of acute and chronic leukemia, other proteins were identified for prognostic assessment. High expression of zeta chain-associated protein (ZAP-70) goes along with poor prognosis in CLL patients. Measurement of ZAP-70 expression was difficult and often not comparable as different laboratories used distinctive cutoff levels for ZAP-70 positivity. However, epigenetic investigations were recently shown to overcome this inaccuracy, making ZAP-70 testing feasible for clinical practice in CLL [65]. Specifically, loss of methylation at a specific single CpG dinucleotide in the ZAP-70 5′ regulatory sequence correlated with unfavorable prognosis in this study, and, hence, represents a reliable and especially reproducible biomarker for CLL risk stratification [65].

The currently most promising concept in the therapy of CLL is the introduction of Bruton's tyrosine kinase inhibitors such as ibrutinib [66]. Here, apoptosis of CLL cells is achieved through specific inhibition of Bruton's tyrosine kinase in the essential B cell receptor signaling pathway. Ibrutinib typically causes an initial lymphocytosis accompanied by a rapid reduction of lymphadenopathy. Due to its oral availability with no relevant myelosuppressive side-effects, ibrutinib is easy to manage for the patient. This drug represents a novel, promising therapeutic alternative for relapsed CLL patients independently of chromosomal alterations normally associated with a poor prognosis [67]. In addition to CLL, ibrutinib also shows excellent response rates in other B cell malignancies, such as relapsed mantle cell lymphoma [66,68]. Despite the already published studies concerning the efficacy of ibrutinib, further trials are currently ongoing to evaluate the optimal use of this new drug to maintain long-lasting remissions. Furthermore, the first reports indicate the occurrence of resistance-mediating mutations in ibrutinib-treated leukemia patients. Thus, future diagnostics in CLL will include sequencing approaches in order to detect specific mutations in order to tailor therapies. As a multitude of novel compounds will be available

in the near future, molecular diagnostics will gain a prominent role in treating CLL patients.

8.9 Outlook and Perspectives

Over recent decades, genetic alterations defining different kinds of leukemia have become more and more clarified due to improvements in cytogenetic and molecular biology methods. In the field of diagnosis and treatment of acute and chronic leukemia, numerous innovations have been achieved in the last decade due to improved diagnostic procedures. By detection of specific chromosomal and molecular alterations in leukemic patients, our understanding of the disease has deepened, and treatment strategies adjusted and stratified according to molecular diagnostics. Subsequent development of targeted drugs that selectively destroy the malignant cell with concomitantly fewer side-effects for the patient has considerably improved anti-leukemic therapy beyond the standard chemotherapeutic regimens. This development has already introduced an increasing application of targeted therapeutics in cancer treatment. However, further investigation is needed to improve personalized medicine. The advent of NGS technology will facilitate broad coverage of molecular markers and specific mutations simultaneously. Quantification of tumor heterogeneity might become more feasible in clinical practice and could also guide clinical concepts.

In conclusion, nucleic acid analysis in molecular diagnostics has already revolutionized clinical practice in hematology and will further evolve as a central pillar in patient treatment.

References

1 Emanuel, B.S., Selden, J.R., Wang, E., Nowell, P.C., and Croce, C.M. (1984) *In situ* hybridization and translocation breakpoint mapping. I. Nonidentical 22q11 breakpoints for the t(9;22) of CML and the t(8;22) of Burkitt lymphoma. *Cytogenetics and Cell Genetics*, **38**, 127–131.

2 Kurzrock, R., Blick, M.B., Talpaz, M., Velasquez, W.S., Trujillo, J.M., Kouttab, N.M. *et al.* (1986) Rearrangement in the breakpoint cluster region and the clinical course in Philadelphia-negative chronic myelogenous leukemia. *Annals of Internal Medicine*, **105**, 673–679.

3 Kurzrock, R., Gutterman, J.U., and Talpaz, M. (1988) The molecular genetics of Philadelphia chromosome-positive leukemias. *New England Journal of Medicine*, **319**, 990–998.

4 Kurzrock, R., Kantarjian, H.M., Druker, B.J., and Talpaz, M. (2003) Philadelphia chromosome-positive leukemias: from basic mechanisms to molecular therapeutics. *Annals of Internal Medicine*, **138**, 819–830.

5 Baccarani, M., Deininger, M.W., Rosti, G., Hochhaus, A., Soverini, S., Apperley, J.F. *et al.* (2013) European LeukemiaNet recommendations for the management of chronic myeloid leukemia: 2013. *Blood*, **122**, 872–884.

6 Bhamidipati, P.K., Kantarjian, H., Cortes, J., Cornelison, A.M., and Jabbour, E.

(2013) Management of imatinib-resistant patients with chronic myeloid leukemia. *Therapeutic Advances in Hematology*, **4**, 103–117.

7 Druker, B.J. and Lydon, N.B. (2000) Lessons learned from the development of an Abl tyrosine kinase inhibitor for chronic myelogenous leukemia. *Journal of Clinical Investigation*, **105**, 3–7.

8 Kantarjian, H.M., Shah, N.P., Cortes, J.E., Baccarani, M., Agarwal, M.B., Undurraga, M.S. *et al.* (2012) Dasatinib or imatinib in newly diagnosed chronic-phase chronic myeloid leukemia: 2-year follow-up from a randomized phase 3 trial (DASISION). *Blood*, **119**, 1123–1129.

9 Larson, R.A., Hochhaus, A., Hughes, T.P., Clark, R.E., Etienne, G., Kim, D.W. *et al.* (2012) Nilotinib vs imatinib in patients with newly diagnosed Philadelphia chromosome-positive chronic myeloid leukemia in chronic phase: ENESTnd 3-year follow-up. *Leukemia*, **26**, 2197–2203.

10 Weisberg, E., Manley, P.W., Cowan-Jacob, S.W., Hochhaus, A., and Griffin, J.D. (2007) Second generation inhibitors of BCR–ABL for the treatment of imatinib-resistant chronic myeloid leukaemia. *Nature Reviews Cancer*, **7**, 345–356.

11 Branford, S., Rudzki, Z., Walsh, S., Grigg, A., Arthur, C., Taylor, K. *et al.* (2002) High frequency of point mutations clustered within the adenosine triphosphate-binding region of BCR/ABL in patients with chronic myeloid leukemia or Ph-positive acute lymphoblastic leukemia who develop imatinib (STI571) resistance. *Blood*, **99**, 3472–3475.

12 Jabbour, E., le Coutre, P.D., Cortes, J., Giles, F., Bhalla, K.N., Pinilla-Ibarz, J. *et al.* (2013) Prediction of outcomes in patients with Ph+ chronic myeloid leukemia in chronic phase treated with nilotinib after imatinib resistance/intolerance. *Leukemia*, **27**, 907–913.

13 Jabbour, E., Morris, V., Kantarjian, H., Yin, C.C., Burton, E., and Cortes, J. (2012) Characteristics and outcomes of patients with V299L BCR–ABL kinase domain mutation after therapy with tyrosine kinase inhibitors. *Blood*, **120**, 3382–3383.

14 Apperley, J.F. (2007) Part I: mechanisms of resistance to imatinib in chronic myeloid leukaemia. *Lancet Oncology*, **8**, 1018–1029.

15 Soverini, S., De Benedittis, C., Machova Polakova, K., Brouckova, A., Horner, D., Iacono, M. *et al.* (2013) Unraveling the complexity of tyrosine kinase inhibitor-resistant populations by ultra-deep sequencing of the BCR–ABL kinase domain. *Blood*, **122**, 1634–1648.

16 Soverini, S., Gnani, A., Colarossi, S., Castagnetti, F., Abruzzese, E., Paolini, S. *et al.* (2009) Philadelphia-positive patients who already harbor imatinib-resistant Bcr–Abl kinase domain mutations have a higher likelihood of developing additional mutations associated with resistance to second- or third-line tyrosine kinase inhibitors. *Blood*, **114**, 2168–2171.

17 Soverini, S., Hochhaus, A., Nicolini, F.E., Gruber, F., Lange, T., Saglio, G. *et al.* (2011) BCR–ABL kinase domain mutation analysis in chronic myeloid leukemia patients treated with tyrosine kinase inhibitors: recommendations from an expert panel on behalf of European LeukemiaNet. *Blood*, **118**, 1208–1215.

18 Hochhaus, A., Kreil, S., Corbin, A.S., La Rosee, P., Muller, M.C., Lahaye, T. *et al.* (2002) Molecular and chromosomal mechanisms of resistance to imatinib (STI571) therapy. *Leukemia*, **16**, 2190–2196.

19 Hochhaus, A. and La Rosee, P. (2004) Imatinib therapy in chronic myelogenous leukemia: strategies to avoid and overcome resistance. *Leukemia*, **18**, 1321–1331.

20 Che, X.F., Nakajima, Y., Sumizawa, T., Ikeda, R., Ren, X.Q., Zheng, C.L. *et al.* (2002) Reversal of P-glycoprotein mediated multidrug resistance by a newly synthesized 1,4-benzothiazipine derivative, JTV-519. *Cancer Letters*, **187**, 111–119.

21 Bennett, J.M., Catovsky, D., Daniel, M.T., Flandrin, G., Galton, D.A., Gralnick, H.R. *et al.* (1976) Proposals for the classification of the acute leukaemias. French–American–British (FAB) co-operative group. *British Journal of Haematology*, **33**, 451–458.

22 Vardiman, J.W., Harris, N.L., and Brunning, R.D. (2002) The World Health

Organization (WHO) classification of the myeloid neoplasms. *Blood*, **100**, 2292–2302.

23 Schoch, C. and Haferlach, T. (2002) Cytogenetics in acute myeloid leukemia. *Current Oncology Reports*, **4**, 390–397.

24 Kihara, R., Nagata, Y., Kiyoi, H., Kato, T., Yamamoto, E., Suzuki, K. *et al.* (2014) Comprehensive analysis of genetic alterations and their prognostic impacts in adult acute myeloid leukemia patients. *Leukemia*, DOI: 10.1038/leu.2014.55 Epub ahead of print].

25 Falini, B., Nicoletti, I., Bolli, N., Martelli, M.P., Liso, A., Gorello, P. *et al.* (2007) Translocations and mutations involving the nucleophosmin (NPM1) gene in lymphomas and leukemias. *Haematologica*, **92**, 519–532.

26 Yamamoto, Y., Kiyoi, H., Nakano, Y., Suzuki, R., Kodera, Y., Miyawaki, S. *et al.* (2001) Activating mutation of D835 within the activation loop of FLT3 in human hematologic malignancies. *Blood*, **97**, 2434–2439.

27 Hou, H.A., Huang, T.C., Lin, L.I., Liu, C.Y., Chen, C.Y., Chou, W.C. *et al.* (2010) WT1 mutation in 470 adult patients with acute myeloid leukemia: stability during disease evolution and implication of its incorporation into a survival scoring system. *Blood*, **115**, 5222–5231.

28 Paietta, E. (2012) Minimal residual disease in acute myeloid leukemia: coming of age. *Hematology*, **2012**, 35–42.

29 Doubek, M., Palasek, I., Pospisil, Z., Borsky, M., Klabusay, M., Brychtova, Y. *et al.* (2009) Detection and treatment of molecular relapse in acute myeloid leukemia with RUNX1 (AML1), CBFB, or MLL gene translocations: frequent quantitative monitoring of molecular markers in different compartments and correlation with WT1 gene expression. *Experimental Hematology*, **37**, 659–672.

30 Tallman, M.S., Andersen, J.W., Schiffer, C. A., Appelbaum, F.R., Feusner, J.H., Ogden, A. *et al.* (1997) All-*trans*-retinoic acid in acute promyelocytic leukemia. *New England Journal of Medicine*, **337**, 1021–1028.

31 Tomita, A., Kiyoi, H., and Naoe, T. (2013) Mechanisms of action and resistance to all-trans retinoic acid (ATRA) and arsenic trioxide (As_2O_3) in acute promyelocytic leukemia. *International Journal of Hematology*, **97**, 717–725.

32 Larson, S. and Stock, W. (2008) Progress in the treatment of adults with acute lymphoblastic leukemia. *Current Opinion in Hematology*, **15**, 400–407.

33 Croce, C.M., Huebner, K., Isobe, M., Fainstain, E., Lifshitz, B., Shtivelman, E. *et al.* (1987) Mapping of four distinct BCR-related loci to chromosome region 22q11: order of BCR loci relative to chronic myelogenous leukemia and acute lymphoblastic leukemia breakpoints. *Proceedings of the National Academy of Sciences of the United States of America*, **84**, 7174–7178.

34 Campana, D. (2010) Minimal residual disease in acute lymphoblastic leukemia. *Hematology*, **2010**, 7–12.

35 Moorman, A.V., Ensor, H.M., Richards, S. M., Chilton, L., Schwab, C., Kinsey, S.E. *et al.* (2010) Prognostic effect of chromosomal abnormalities in childhood B-cell precursor acute lymphoblastic leukaemia: results from the UK Medical Research Council ALL97/99 randomised trial. *Lancet Oncology*, **11**, 429–438.

36 Clarke, S., O'Reilly, J., Romeo, G., and Cooney, J. (2011) NUP214-ABL1 positive T-cell acute lymphoblastic leukemia patient shows an initial favorable response to imatinib therapy post relapse. *Leukemia Research*, **35**, e131–e133.

37 Harrison, C.J. (2013) Targeting signaling pathways in acute lymphoblastic leukemia: new insights. *Hematology*, **2013**, 118–125.

38 Uckun, F.M., Herman-Hatten, K., Crotty, M.L., Sensel, M.G., Sather, H.N., Tuel-Ahlgren, L. *et al.* (1998) Clinical significance of MLL–AF4 fusion transcript expression in the absence of a cytogenetically detectable t(4;11)(q21;q23) chromosomal translocation. *Blood*, **92**, 810–821.

39 Cramer, P. and Hallek, M. (2011) Prognostic factors in chronic lymphocytic leukemia – what do we need to know? *Nature Reviews Clinical Oncology*, **8**, 38–47.

40 Matthews, C., Catherwood, M., Morris, T. C., and Alexander, H.D. (2004) Routine

analysis of IgVH mutational status in CLL patients using BIOMED-2 standardized primers and protocols. *Leukemia & Lymphoma*, **45**, 1899–1904.

41 Kharfan-Dabaja, M.A., Chavez, J.C., Khorfan, K.A., and Pinilla-Ibarz, J. (2008) Clinical and therapeutic implications of the mutational status of IgVH in patients with chronic lymphocytic leukemia. *Cancer*, **113**, 897–906.

42 Damle, R.N., Wasil, T., Fais, F., Ghiotto, F., Valetto, A., Allen, S.L. *et al.* (1999) Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood*, **94**, 1840–1847.

43 Hamblin, T.J., Davis, Z., Gardiner, A., Oscier, D.G., and Stevenson, F.K. (1999) Unmutated Ig V_H genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood*, **94**, 1848–1854.

44 Tobin, G., Thunberg, U., Johnson, A., Thorn, I., Soderberg, O., Hultdin, M. *et al.* (2002) Somatically mutated Ig V_H3-21 genes characterize a new subset of chronic lymphocytic leukemia. *Blood*, **99**, 2262–2264.

45 Mohr, J., Helfrich, H., Fuge, M., Eldering, E., Buhler, A., Winkler, D. *et al.* (2011) DNA damage-induced transcriptional program in CLL: biological and diagnostic implications for functional p53 testing. *Blood*, **117**, 1622–1632.

46 Bullrich, F., Rasio, D., Kitada, S., Starostik, P., Kipps, T., Keating, M. *et al.* (1999) ATM mutations in B-cell chronic lymphocytic leukemia. *Cancer Research*, **59**, 24–27.

47 Dohner, H., Stilgenbauer, S., Benner, A., Leupolt, E., Krober, A., Bullinger, L. *et al.* (2000) Genomic aberrations and survival in chronic lymphocytic leukemia. *New England Journal of Medicine*, **343**, 1910–1916.

48 Dohner, H., Fischer, K., Bentz, M., Hansen, K., Benner, A., Cabot, G. *et al.* (1995) p53 gene deletion predicts for poor survival and non-response to therapy with purine analogs in chronic B-cell leukemias. *Blood*, **85**, 1580–1589.

49 Stilgenbauer, S., Bullinger, L., Lichter, P., Dohner, H., and German CLL Study Group (2002) Genetics of chronic lymphocytic leukemia: genomic aberrations and V_H gene mutation status in pathogenesis and clinical course. *Leukemia*, **16**, 993–1007.

50 Klein, U., Lia, M., Crespo, M., Siegel, R., Shen, Q., Mo, T. *et al.* (2010) The DLEU2/miR-15a/16-1 cluster controls B cell proliferation and its deletion leads to chronic lymphocytic leukemia. *Cancer Cell*, **17**, 28–40.

51 Mayr, C., Speicher, M.R., Kofler, D.M., Buhmann, R., Strehl, J., Busch, R. *et al.* (2006) Chromosomal translocations are associated with poor prognosis in chronic lymphocytic leukemia. *Blood*, **107**, 742–751.

52 Goede, V., Fischer, K., Busch, R., Engelke, A., Eichhorst, B., Wendtner, C.M. *et al.* (2014) Obinutuzumab plus chlorambucil in patients with CLL and coexisting conditions. *New England Journal of Medicine*, **370**, 1101–1110.

53 Hallek, M., Fischer, K., Fingerle-Rowson, G., Fink, A.M., Busch, R., Mayer, J. *et al.* (2010) Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomised, open-label, phase 3 trial. *Lancet*, **376**, 1164–1174.

54 Stilgenbauer, S., Zenz, T., Winkler, D., Buhler, A., Schlenk, R.F., Groner, S. *et al.* (2009) Subcutaneous alemtuzumab in fludarabine-refractory chronic lymphocytic leukemia: clinical results and prognostic marker analyses from the CLL2H study of the German Chronic Lymphocytic Leukemia Study Group. *Journal of Clinical Oncology*, **27**, 3994–4001.

55 Schetelig, J., van Biezen, A., Brand, R., Caballero, D., Martino, R., Itala, M. *et al.* (2008) Allogeneic hematopoietic stem-cell transplantation for chronic lymphocytic leukemia with 17p deletion: a retrospective European Group for Blood and Marrow Transplantation analysis. *Journal of Clinical Oncology*, **26**, 5094–5100.

56 Schweighofer, C.D., Coombes, K.R., Majewski, T., Barron, L.L., Lerner, S., Sargent, R.L. *et al.* (2013) Genomic variation by whole-genome SNP mapping arrays predicts time-to-event outcome in

patients with chronic lymphocytic leukemia: a comparison of CLL and HapMap genotypes. *Journal of Molecular Diagnostics*, **15**, 196–209.

57 Quesada, V., Conde, L., Villamor, N., Ordonez, G.R., Jares, P., Bassaganyas, L. *et al.* (2012) Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nature Genetics*, **44**, 47–52.

58 Wang, L., Lawrence, M.S., Wan, Y., Stojanov, P., Sougnez, C., Stevenson, K. *et al.* (2011) SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *New England Journal of Medicine*, **365**, 2497–2506.

59 Wickremasinghe, R.G., Prentice, A.G., and Steele, A.J. (2011) p53 and Notch signaling in chronic lymphocytic leukemia: clues to identifying novel therapeutic strategies. *Leukemia*, **25**, 1400–1407.

60 Real, P.J. and Ferrando, A.A. (2009) NOTCH inhibition and glucocorticoid therapy in T-cell acute lymphoblastic leukemia. *Leukemia*, **23**, 1374–1377.

61 Rosati, E., Sabatini, R., Rampino, G., Tabilio, A., Di Ianni, M., Fettucciari, K. *et al.* (2009) Constitutively activated Notch signaling is involved in survival and apoptosis resistance of B-CLL cells. *Blood*, **113**, 856–865.

62 Rossi, D., Rasi, S., Fabbri, G., Spina, V., Fangazio, M., Forconi, F. *et al.* (2012) Mutations of NOTCH1 are an independent predictor of survival in chronic lymphocytic leukemia. *Blood*, **119**, 521–529.

63 Tzoneva, G. and Ferrando, A.A. (2012) Recent advances on NOTCH signaling in T-ALL. *Current Topics in Microbiology and Immunology*, **360**, 163–182.

64 Saha, M.N., Micallef, J., Qiu, L., and Chang, H. (2010) Pharmacological activation of the p53 pathway in haematological malignancies. *Journal of Clinical Pathology*, **63**, 204–209.

65 Claus, R., Lucas, D.M., Stilgenbauer, S., Ruppert, A.S., Yu, L., Zucknick, M. *et al.* (2012) Quantitative DNA methylation analysis identifies a single CpG dinucleotide important for ZAP-70 expression and predictive of prognosis in chronic lymphocytic leukemia. *Journal of Clinical Oncology*, **30**, 2483–2491.

66 de Rooij, M.F., Kuil, A., Geest, C.R., Eldering, E., Chang, B.Y., Buggy, J.J. *et al.* (2012) The clinically active BTK inhibitor PCI-32765 targets B-cell receptor- and chemokine-controlled adhesion and migration in chronic lymphocytic leukemia. *Blood*, **119**, 2590–2594.

67 Byrd, J.C., Furman, R.R., Coutre, S.E., Flinn, I.W., Burger, J.A., Blum, K.A. *et al.* (2013) Targeting BTK with ibrutinib in relapsed chronic lymphocytic leukemia. *New England Journal of Medicine*, **369**, 32–42.

68 Wang, M.L., Rule, S., Martin, P., Goy, A., Auer, R., Kahl, B.S. *et al.* (2013) Targeting BTK with ibrutinib in relapsed or refractory mantle-cell lymphoma. *New England Journal of Medicine*, **369**, 507–516.

9
Techniques of Nucleic Acid-Based Diagnosis in the Management of Bacterial and Viral Infectious Diseases

Irene Latorre, Verónica Saludes, Juana Díez, and Andreas Meyerhans

9.1
Importance of Nucleic Acid-Based Molecular Assays in Clinical Microbiology

Defining the cause of an infectious disease is a major challenge for microbiological diagnosis. A solution for this difficult task goes back to the German physician Robert Koch. He provided experimental support for the importance of microorganisms in causing human diseases and formulated four rigorous criteria, called Koch's postulates. These postulates provided the necessary tools for describing the cause of an infectious disease and highlighted the significance of culturing the causative agent. Koch's postulates were developed in the nineteenth century, and represented an enormous step forward in the study of infectious diseases and clinical microbiology. Many of the significant infectious diseases in humans were discovered with these postulates. In addition to their historical importance, Koch's postulates continue to inform us of how to proceed in microbiologic diagnosis, despite the fact that fulfillment of all four postulates demonstrating which infectious agent is causing the disease is technically not always possible.

Early diagnosis and treatment administration play an important role in preventing the development of infectious disease-related complications within a host and help, in addition to vaccination, to avoid the spread of an infectious agent within a host population. The classical infectious disease diagnostics consists of the detection, identification, and characterization of the causative microorganism from a sample collected of the infected host. As well as direct observation of the pathogen by microscopy, antigen detection, or genome detection, there are indirect assays that detect humoral or cellular immune responses generated after an infection. These indirect methods are very common in virus diagnostics. However, as an immune response needs time to develop, indirect assays usually have a longer diagnostic window in which the infection cannot be detected. Independent on the type of assay used for microbial diagnosis, assay quality is an important criteria. This is defined by sensitivity and specificity. The sensitivity measures the proportion of correctly identified positive results. A high

Nucleic Acids as Molecular Diagnostics, First Edition. Edited by Andreas Keller and Eckart Meese.

sensitivity is associated with a low frequency of false-negative results. In contrast, the specificity measures the proportion of correctly identified negative results. A high specificity is associated with a low frequency of false-positive results.

Conventional microbiological diagnosis was revolutionized when molecular-based technologies entered the clinic. This was initiated at the end of the 1980s with the polymerase chain reaction (PCR) – one of the most important molecular techniques ever developed. In a PCR, a target DNA region is amplified billions of times by means of several cycles of DNA denaturation, oligonucleotide primer hybridization, and extension by a DNA polymerase [1]. Thus, a PCR can directly amplify minute quantities of microbial genomes and make them assessable for further analysis, such as detection of resistance mutations against antiviral drugs or antibiotics. In some cases an important limitation may be the inability of not being able to discriminate a pathogenic infection from natural colonization. Likewise, an amplified genome may be derived from a live or a dead pathogen, as in antibiotic-treated active tuberculosis patients, complicating the interpretation of the assay results. In other cases, false-negative results may occur due to pathogen loads below the limit of detection or the presence of inhibitors [2]. Despite these limitations, the use of PCR and other nucleic acid-based assays has become well accepted, although their specificity and sensitivity require optimization in some cases [3,4].

Novel high-throughput molecular technologies such as next-generation sequencing (NGS) and mass spectrometry may be considered as the new ongoing revolution in clinical microbiology. Among other applications, they have made a decisive impact on strain typing and detecting antibiotic susceptibility. A very promising new area for NGS is metagenomics. It has emerged as a powerful tool to describe the microbiome of different human specimens, and enables us to analyze and relate the microbiota to human diseases [5]. Although we are still far from the routine use of these novel technologies in clinical microbiology laboratories, numerous efforts will be focused in this direction in the coming years.

Here, we provide an up-to-date review of the main nucleic acid-based molecular assays applied in clinical microbiology. Figure 9.1 shows a summary of the main molecular techniques used for the diagnosis of bacteria and viruses causing infectious diseases. Novel applications and technologies with potential interest for the diagnosis and management of infectious diseases will also be discussed.

9.2 Nucleic Acid Amplification Techniques

There are two commonly used strategies to visualize minute quantities of nucleic acids from pathogens: target amplification and signal amplification. Kits based on both principles are available and in use for the microbiological diagnosis of infectious diseases.

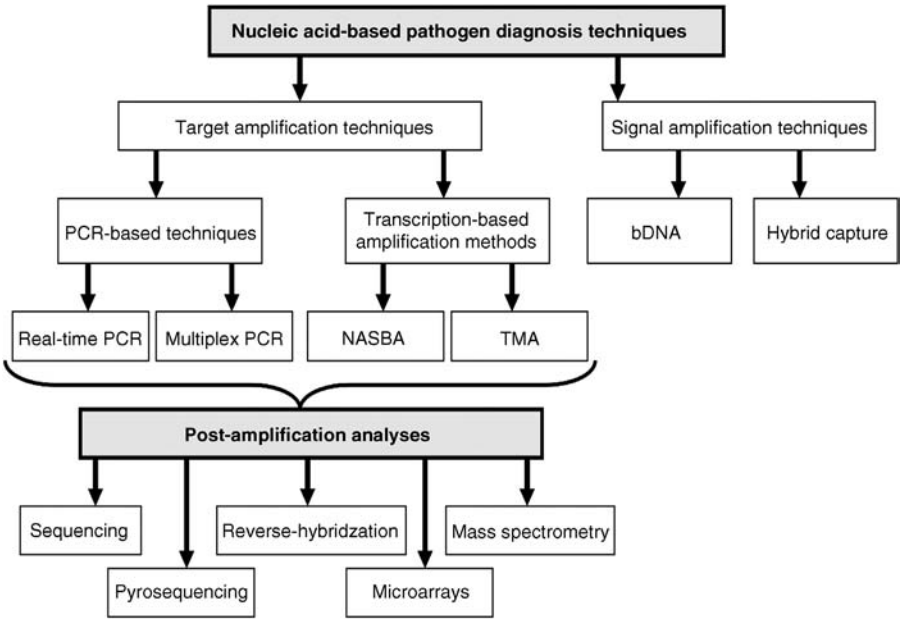


Figure 9.1 Main nucleic acid-based techniques used in the clinical setting for the diagnosis of pathogens causing infectious diseases. bDNA, branched DNA; NASBA, nucleic acid sequence-based amplification; PCR, polymerase chain reaction; TMA, transcription-mediated amplification.

9.2.1 Target Amplification Techniques

The two commonly applied target amplification methods in clinical microbiology are the PCR-based and the transcription-based amplification methods. Both techniques use enzyme-mediated processes to massively copy a targeted nucleic acid sequence. Due to their extreme sensitivity, it is important to avoid sample contamination with DNA from previously amplified templates.

9.2.1.1 PCR-Based Techniques

The PCR is by far the most universal methodology for genome amplification. In its current form, it uses thermostable DNA polymerases, specific primers, and nucleotide triphosphates to amplify a predefined DNA target sequence. Amplification is achieved through cycling temperatures that allow DNA denaturation, primer hybridization, and primer extension [6]. A special form of the PCR technique that is most commonly used in clinical microbiology today is real-time PCR. In contrast to conventional PCR, real-time PCR simultaneously amplifies a target region and quantifies the reaction product after each cycle. When used in combination with a standard DNA of known concentration, it provides quantitative information on the available target DNA. As the reaction in the tube/plate is

sealed and never opened after the amplification, the risk of amplicon carryover is significantly reduced. Monitoring of the amplification product is achieved by fluorescent dyes. They include fluorescent DNA-intercalating dyes such as SYBR Green, which preferentially binds to double-stranded DNA [7], and fluorescent-labeled internal DNA probes that are specifically designed to hybridize to the target region. Among the latter, TaqMan probes (Applied Biosystems), fluorescence resonance energy transfer (FRET), and molecular beacon probes are used [8–10]. The major drawback of real-time PCR in a clinical setting is its low multiplexing capacity for the detection of several targets simultaneously. Nonetheless, some multiplex PCR assays have been developed and commercialized for several applications, such as the detection of viral respiratory pathogens and viral infections of the central nervous system [11,12].

Commercially available PCR assays initially focused on the detection of sexually transmitted infections and blood-borne viruses. However, a rapid expansion in the development of new commercial fully automated and in-house PCR-based assays in human infection diagnostics has taken place (www.fda.org) [13]. Point-of-care tests have been developed for difficult-to-detect microorganisms such as *Mycobacterium tuberculosis* [14,15] or difficult-to-diagnose syndromes such as bacterial sepsis and pneumonia [16,17]. In addition, tests for the detection of virulence factors and microbial resistances like methicillin-resistant *Staphylococcus aureus* are available [3,18].

9.2.1.2 Transcription-Based Amplification Methods

Transcription-based amplification methods include nucleic acid sequence-based amplification (NASBA) and transcription-mediated amplification (TMA). In both methods, RNA sequences are reverse-transcribed into DNA. From DNA, RNA copies are synthesized with a RNA polymerase through isothermal RNA amplification. The main difference between these two techniques is that NASBA uses a reverse transcriptase (RT) from avian myeloblastosis virus and TMA uses a Moloney murine leukemia virus RT. NASBA also requires supplementary addition of RNase H [19–22]. These assays have several advantages, such as no requirement for a thermal cycler, rapid kinetics, and a single-stranded RNA product that does not require denaturation prior to detection. Commercial NASBA-based kits (bioMérieux, Marcy l'Etoile, France) are available for the detection and quantitation of human immunodeficiency virus type 1 (HIV-1) and cytomegalovirus (CMV), and the detection of enteroviruses and respiratory syncytial virus [23–26]. Gen-Probe (San Diego, CA, USA) has developed TMA-based assays to detect *M. tuberculosis*, *Chlamydia trachomatis*, *Neisseria gonorrhoeae*, hepatitis C virus (HCV) and HIV-1 [21,27–32].

9.2.2 Signal Amplification Techniques

Signal amplification assays are used to detect nucleic acid molecules in a specific clinical specimen by generating multiple signals that are directly proportional

to the amount of the targeted nucleic acid sequences. These assays do not require the amplification of a target sequence. Branched DNA (bDNA) and hybrid capture are the two main methods routinely used.

bDNA is based on a signal amplification after direct binding of a large hybridization complex to a target sequence creating a branched structure. Signal amplification takes place when complementary alkaline phosphatase-labeled nucleotides bind to the bDNA structure, generating a chemiluminiscent signal [33]. Commercial available assays based on this technology have been used for the quantitation of HCV, hepatitis B virus (HBV) and HIV-1 (Bayer Health Care, Diagnostics Division, Tarrytown, NY, USA) [34–39]. These methods have improved from first-generation to more accurate, automated and highly sensitive third-generation bDNA assays [27,28,40].

Hybrid capture technology is based on a nucleic acid hybridization assay. The DNA of the sample is denatured and afterwards hybridized to a RNA probe. The resultant DNA–RNA hybrid is then captured and immobilized onto the surface of a microplate coated with specific antibodies. Subsequently, multiple alkaline phosphatase-conjugated specific antibodies bind to the immobilized hybrids, creating a signal amplification. Commercially available assays based on this technology are used for the detection of human papillomaviruses (HPVs), *N. gonorrhoeae*, *C. trachomatis*, and CMV (Digene, Gaithersburg, MD, USA) [41–44].

9.3 Post-Amplification Analyses

Post-amplification analyses for detecting amplified DNA include methods such as traditional Sanger sequencing, pyrosequencing, reverse hybridization, and high-throughput nucleic acid-based analyses via DNA microarrays, mass spectrometry, or NGS.

9.3.1 Sequencing and Pyrosequencing

The dideoxy DNA sequencing technology developed by Sanger [45] is a standard research tool that has been introduced into clinical diagnostics. Today it is still used for sequencing PCR-amplified DNA. The sequence obtained can be compared with several databases for identification of microorganisms or polymorphisms. This method helped to obtain the first complete reference genome sequences of thousands of species [46]. Currently, it is still the principle way of identifying 16S rRNA sequences of bacteria, providing species-specific signature sequences useful for bacterial identification. It is also frequently used for the identification of mutations associated with antiviral drug or antibiotic resistance, as well as for HCV genotyping and monitoring HIV treatment efficacy [47–49].

Pyrosequencing is a short-read sequencing method that has been shown to be cost-effective and improve diagnostic capability [13,28]. It is based on the

principle of "sequencing-by-synthesis." In each step of the sequence reaction, the incorporation of a deoxyribonucleotide triphosphate (dNTP) leads to the release of a pyrophosphate (PP_i). This is subsequently converted into ATP and used to activate luciferase that produces a light signal proportional to the amount of the incorporated nucleotide [28]. Pyrosequencing is quantitative, fast, and inexpensive. Additional advantages include high accuracy, flexibility, and ability to automate sample preparation [50]. In microbiology, pyrosequencing has been used to detect drug resistance mutations, and to identify and type bacteria and viruses [51–56]. It has also been applied for typing and detecting β-lactamase-resistant *Klebsiella pneumonia* and *Escherichia coli* [57,58], as well as for identifying penicillin-resistant *Neisseria meningitidis* [59].

9.3.2 Reverse Hybridization

In the so-called reverse hybridization technique, an amplified PCR product is biotin-labeled and applied to a test strip that includes several species-specific probes located at specific positions. The PCR product hybridizes only with sequence-matched regions, which is subsequently detected after the addition of a color-developing substrate. Commercial assays have been developed to identify *Mycobacteria* and drug-resistant gene mutants, to detect *Helicobacter pylori* virulence genotypes, for HCV/HBV genotyping, and to detect drug-resistant mutations in HIV-1 [60–66].

9.3.3 High-Throughput Nucleic Acid-Based Analyses

Infectious disease diagnostics and management is moving toward simple, fast, and accurate high-throughput technologies that will provide a massive amount of information for personalized medicine. The hope is that this may lead to a more accurate diagnosis and successful treatment. There are three main candidate approaches within high-throughput technologies for analyzing multiple amplified nucleic acid sequences: DNA microarrays, mass spectrometry, and NGS technologies. While the first two are already being implemented in many clinical laboratories, NGS techniques are so far restricted to research laboratories.

9.3.3.1 DNA Microarrays

A DNA microarray is a huge collection of single-stranded DNA oligonucleotide probes attached to a solid support [67,68]. It can be used for diagnosis following the steps: (i) nucleic acid extraction from the clinical sample, (ii) amplification by solid- or solution-phase PCR or reverse transcription (RT)-PCR, (iii) labeling of amplified products with fluorescent dyes, (iv) hybridization on the microarray, and (v) scanning of the fluorescence emission to measure the specific DNA/probe interactions. The results are obtained within 24–48 h. The main advantage of this technology is the huge amount of potential targets that can be

discriminated at the same time and, thus, multiple pathogens can be tested simultaneously.

The use of microarrays in clinical diagnostics was first reported in 1993 for the detection of a hantavirus infection outbreak [69]. Since then, numerous microarray-based assays have been developed for the identification and detection of different genus/species/serotypes and their resistance/virulence determinants in both bacterial and viral infectious diseases. They are valuable for the identification and resistance determination of mycobacterial species, identification of influenza viruses, diagnosis of bacteremia and detection of methicillin resistance in *S. aureus*, identification of bacterial species associated with digestive infections or detection, and genotyping of sexually transmitted infections like *Chlamydia* or HPV [70–78]. In general, related studies have demonstrated that the microarray technology gives a similar or better sensitivity and specificity than reference serological-based or multiplexing PCR approaches.

9.3.3.2 Mass Spectrometry

More recently, mass spectrometry has been incorporated into microbiology diagnosis laboratories as an excellent technology for rapid detection, identification, and classification of microorganisms by means of the detection of their proteins or nucleic acids [79]. Mass spectrometry works by ionizing chemical compounds to generate charged molecules and measuring their mass-to-charge (m/z) ratios. The most common mass spectrometry technologies are matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) coupled with a time-of-flight (TOF) analyzer (i.e., MALDI-TOF and ESI-TOF, respectively).

In MALDI-TOF technology, the sample to be analyzed (i.e., PCR-amplified nucleic acids) is inserted into a matrix crystal and then ionized by a pulsed laser light. In ESI-TOF technology, the sample is dissolved in an organic solvent, injected through a steel capillary, and then ionized by applying a high voltage. In both technologies, a mass ion detector finally measures the flight time from the ion source to the detector. Data are recorded as a mass spectrum that shows ion intensity versus mass-to-charge ratio values. TOF analyzers require short analysis times within the millisecond range.

Mass spectrometry technologies that detect microbial nucleic acids have been used for the efficient identification of bacteria, viruses, and virulence and/or resistance factors [80]. By using MALDI-TOF technology, Sequenom (San Diego, CA, USA) has provided an alternative to multilocus sequence typing (MLST) based on the detection of RNA fragments that are derived from bacterial targeted genomic loci subjected to PCR and transcription [81]. The generated mass patterns are matched with *in silico* MLST-derived patterns for species identification. Results can be obtained within 1 day. This technology has been used for typing *N. meningitidis* [81], and for genotyping HBV and HCV [82,83]. Regarding the use of the ESI-TOF technology, Ibis Biosciences (Carlsbad, CA, USA) has developed a technology called PLEX-ID. It is based on a multilocus

PCR amplification of informative genomic loci with broad-range primers followed by ESI-based detection of PCR products and base composition analysis [84]. This technology allows the identification and quantification of a broad range of microorganisms, including bacteria and viruses, and the identification of resistance and/or virulence factors, by targeting conserved sequences using universal primers. Among bacterial pathogens, different species of *M. tuberculosis* (including multiresistant *M. tuberculosis*), and respiratory pathogens such as *Haemophilus influenzae*, *N. meningitidis*, and *Streptococcus pyogenes* have been detected [84,85]. Interestingly, ESI-TOF is also a promising technology for the rapid diagnosis of sepsis [86]. In addition, this technology has been applied for the detection of antibiotic resistance in *Acinetobacter* spp. [87]. Regarding the identification of viruses, this methodology was developed for the identification of the H1N1 influenza A virus [88], among others.

9.3.3.3 NGS

The term "next-generation sequencing" refers to new high-throughput nucleic acid technologies [89] that generate nucleotide sequence information. The first high-throughput sequencing technology was published in 2005 [90] and several platforms have since been developed. The current platforms can be divided into (i) high-capacity sequencers, which are especially used to identify gene polymorphisms, and (ii) long-read sequencers, which generate sequences suitable for linkage studies or *de novo* sequence characterization.

NGS methods use a three-step protocol: (i) library preparation, (ii) DNA capture and enrichment, and (iii) sequencing/detection [91]. The detection step can be based on pyrophosphate (pyrosequencing), H^+ ions (pH variation), fluorescence, or ligation combined with fluorescence detection. After detection, generated signals are converted into readable sequences in a FASTA format. Unreliable sequences are eliminated and final sequences are aligned. However, there is a need for automated data interpretation, user-friendly bioinformatic tools, and for a reduction of the equipment costs in order to implement NGS technologies in the clinical setting.

NGS has already been demonstrated to be very fruitful in microbiology, such as in the characterization of transmission pathways of healthcare-associated infections, the identification of genetic determinants of antimicrobial resistance, the full-genome sequencing of several viruses (including HIV), and the characterization of intrahost genetic variability for highly variable viruses [92–95]. With this technology the pre-existence of amino acid substitutions that confer HBV, HCV, and HIV resistance to specific antiviral drugs has been demonstrated [96–98]. Microbial metagenomics is also among the potential applications. The term "metagenomics" refers to the analysis of all nucleic acids present in a given sample, allowing the profiling of complex microbial communities associated with human diseases (e.g., the gut bacterial metagenome and the human virome) [99,100]. In addition, metagenomic studies led to the discovery of novel viruses from clinical samples, like the Ebola virus Bundiubugyo [101].

9.4
General Overview and Concluding Remarks

For an efficient clinical management of infectious diseases, it is important to accelerate the identification of pathogenic microorganisms and their sensitivity towards antiviral drugs or antibiotics. In this regard, molecular assays for detecting nucleic acids have revolutionized microbiology and clinical laboratories. While some of these assays, especially the high-throughput molecular techniques such as NGS and mass spectrometry, are currently only used in reference settings or research laboratories, their implementation will increase over the next decade. This is easily envisaged due to their potential of being very powerful for personalized diagnosis and therapy. However, their cost-effectiveness and simplicity should increase significantly to replace conventional methods such as culture.

Acknowledgments

This work was supported by grants from the Spanish Ministerio de Ciencia e Innovación BFU2010-20803 and SAF2010-21336. A.M. is a researcher funded from the Institució Catalana de Recerca i Estudis Avançats (ICREA). I.L. is the recipient of a Joint ERS/SEPAR Long-Term Fellowship (LTRF 82-2012).

References

1 Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239** (4839), 487–491.

2 Maurer, J.J. (2011) Rapid detection and limitations of molecular techniques. *Annual Review of Food Science Technology*, **2**, 259–279.

3 van Belkum, A., Durand, G., Peyret, M., Chatellier, S., Zambardi, G., Schrenzel, J., Shortridge, D., Engelhardt, A., and Dunne, W.M.Jr. (2013) Rapid clinical bacteriology and its future impact. *Annals of Laboratory Medicine*, **33** (1), 14–27.

4 Hadgu, A., Dendukuri, N., and Hilden, J. (2005) Evaluation of nucleic acid amplification tests in the absence of a perfect gold-standard test: a review of the statistical and epidemiologic issues. *Epidemiology*, **16** (5), 604–612.

5 Nakamura, S., Nakaya, T., and Iida, T. (2011) Metagenomic analysis of bacterial infections by means of high-throughput DNA sequencing. *Experimental Biology and Medicine*, **236** (8), 968–971.

6 Yang, S. and Rothman, R.E. (2004) PCR-based diagnostics for infectious diseases: uses, limitations, and future applications in acute-care settings. *Lancet Infectious Diseases*, **4** (6), 337–348.

7 Higuchi, R., Dollinger, G., Walsh, P.S., and Griffith, R. (1992) Simultaneous amplification and detection of specific DNA sequences. *Biotechnology*, **10** (4), 413–417.

8 Heid, C.A., Stevens, J., Livak, K.J., and Williams, P.M. (1996) Real time quantitative PCR. *Genome Research*, **6** (10), 986–994.

9 Chen, X., Zehnbauer, B., Gnirke, A., and Kwok, P.Y. (1997) Fluorescence energy transfer detection as a homogeneous DNA diagnostic method. *Proceedings of the National Academy of Sciences of the United States of America*, **94** (20), 10756–10761.

10 Tyagi, S. and Kramer, F.R. (1996) Molecular beacons: probes that fluoresce upon hybridization. *Nature Biotechnology*, **14** (3), 303–308.

11 Reddington, K., Tuite, N., Barry, T., O'Grady, J., and Zumla, A. (2013) Advances in multiparametric molecular diagnostics technologies for respiratory tract infections. *Current Opinion in Pulmonary Medicine*, **19** (3), 298–304.

12 Leveque, N., Renois, F., and Andreoletti, L. (2013) The microarray technology: facts and controversies. *Clinical Microbiology and Infection*, **19** (1), 10–14.

13 Sibley, C.D., Peirano, G., and Church, D. L. (2012) Molecular methods for pathogen and microbial community detection and characterization: current and potential application in diagnostic microbiology. *Infection, Genetics and Evolution*, **12** (3), 505–521.

14 Boehme, C.C., Nabeta, P., Hillemann, D., Nicol, M.P., Shenai, S., Krapp, F., Allen, J., Tahirli, R., Blakemore, R., Rustomjee, R., Milovic, A., Jones, M., O'Brien, S.M., Persing, D.H., Ruesch-Gerdes, S., Gotuzzo, E., Rodrigues, C., Alland, D., and Perkins, M.D. (2010) Rapid molecular detection of tuberculosis and rifampin resistance. *New England Journal of Medicine*, **363** (11), 1005–1015.

15 Dheda, K., Ruhwald, M., Theron, G., Peter, J., and Yam, W.C. (2013) Point-of-care diagnosis of tuberculosis: past, present and future. *Respirology*, **18** (2), 217–232.

16 Ecker, D.J., Sampath, R., Li, H., Massire, C., Matthews, H.E., Toleno, D., Hall, T. A., Blyn, L.B., Eshoo, M.W., Ranken, R., Hofstadler, S.A., and Tang, Y.W. (2010) New technology for rapid molecular diagnosis of bloodstream infections. *Expert Review of Molecular Diagnostics*, **10** (4), 399–415.

17 Tenover, F.C. (2011) Developing molecular amplification methods for rapid diagnosis of respiratory tract infections caused by bacterial pathogens. *Clinical Infectious Diseases*, **52** (Suppl 4), S338–S345.

18 van Hal, S.J., Jennings, Z., Stark, D., Marriott, D., and Harkness, J. (2009) MRSA detection: comparison of two molecular methods (BD GeneOhm PCR assay and Easy-Plex) with two selective MRSA agars (MRSA-ID and Oxoid MRSA) for nasal swabs. *European Journal of Clinical Microbiology & Infectious Diseases*, **28** (1), 47–53.

19 Compton, J. (1991) Nucleic acid sequence-based amplification. *Nature*, **350** (6313), 91–92.

20 Kwoh, D.Y., Davis, G.R., Whitfield, K.M., Chappelle, H.L., DiMichele, L.J., and Gingeras, T.R. (1989) Transcription-based amplification system and detection of amplified human immunodeficiency virus type 1 with a bead-based sandwich hybridization format. *Proceedings of the National Academy of Sciences of the United States of America*, **86** (4), 1173–1177.

21 Nolte, F.S. and Caliendo, A.M. (2007) Molecular detection and identification of microorganisms, in *Manual of Clinical Microbiology*, 9th edn (eds P.R. Murray, E.J. Baron, J.H. Jorgensen, M.L. Landry, and M.A. Pfaller), ASM Press, Washington, DC, pp. 218–244.

22 Langabeer, S.E., Gale, R.E., Harvey, R.C., Cook, R.W., Mackinnon, S., and Linch, D. C. (2002) Transcription-mediated amplification and hybridisation protection assay to determine BCR–ABL transcript levels in patients with chronic myeloid leukaemia. *Leukemia*, **16** (3), 393–399.

23 de Mendoza, C., Koppelman, M., Montes, B., Ferre, V., Soriano, V., Cuypers, H., Segondy, M., and Oosterlaken, T. (2005) Multicenter evaluation of the NucliSens EasyQ HIV-1 v1.1 assay for the quantitative detection of HIV-1 RNA in plasma. *Journal of Virological Methods*, **127** (1), 54–59.

24 Keightley, M.C., Rinaldo, C., Bullotta, A., Dauber, J., and St George, K. (2006) Clinical utility of CMV early and late transcript detection with NASBA in

bronchoalveolar lavages. *Journal of Clinical Virology*, **37** (4), 258–264.

25 Landry, M.L., Garner, R., and Ferguson, D. (2003) Comparison of the NucliSens Basic kit (Nucleic Acid Sequence-Based Amplification) and the Argene Biosoft Enterovirus Consensus Reverse Transcription-PCR assays for rapid detection of enterovirus RNA in clinical specimens. *Journal of Clinical Microbiology*, **41** (11), 5006–5010.

26 Moore, C., Valappil, M., Corden, S., and Westmoreland, D. (2006) Enhanced clinical utility of the NucliSens EasyQ RSV A+B Assay for rapid detection of respiratory syncytial virus in clinical samples. *European Journal of Clinical Microbiology & Infectious Diseases*, **25** (3), 167–174.

27 Cobo, F. (2012) Application of molecular diagnostic techniques for viral testing. *Open Virology Journal*, **6**, 104–114.

28 Muldrew, K.L. (2009) Molecular diagnostics of infectious diseases. *Current Opinion in Pediatrics*, **21** (1), 102–111.

29 La Rocco, M.T., Wanger, A., Ocera, H., and Macias, E. (1994) Evaluation of a commercial rRNA amplification assay for direct detection of *Mycobacterium tuberculosis* in processed sputum. *European Journal of Clinical Microbiology & Infectious Diseases*, **13** (9), 726–731.

30 Wada, K., Uehara, S., Mitsuhata, R., Kariyama, R., Nose, H., Sako, S., Ishii, A., Watanabe, T., Matsumoto, A., Monden, K., Uno, S., Araki, T., and Kumon, H. (2012) Prevalence of pharyngeal *Chlamydia trachomatis* and *Neisseria gonorrhoeae* among heterosexual men in Japan. *Journal of Infection and Chemotherapy*, **18** (5), 729–733.

31 Moller, J.K., Pedersen, L.N., and Persson, K. (2008) Comparison of Gen-probe transcription-mediated amplification, Abbott PCR, and Roche PCR assays for detection of wild-type and mutant plasmid strains of *Chlamydia trachomatis* in Sweden. *Journal of Clinical Microbiology*, **46** (12), 3892–3895.

32 Linnen, J.M., Gilker, J.M., Menez, A., Vaughn, A., Broulik, A., Dockter, J., Gillotte-Taylor, K., Greenbaum, K., Kolk, D.P., Mimms, L.T., and Giachetti, C. (2002) Sensitive detection of genetic variants of HIV-1 and HCV with an HIV-1/HCV assay based on transcription-mediated amplification. *Journal of Virological Methods*, **102** (1–2), 139–155.

33 Nolte, F.S. (1998) Branched DNA signal amplification for direct quantitation of nucleic acid sequences in clinical specimens. *Advances in Clinical Chemistry*, **33**, 201–235.

34 Pittaluga, F., Allice, T., Abate, M.L., Ciancio, A., Cerutti, F., Varetto, S., Colucci, G., Smedile, A., and Ghisetti, V. (2008) Clinical evaluation of the COBAS Ampliprep/COBAS TaqMan for HCV RNA quantitation in comparison with the branched-DNA assay. *Journal of Medical Virology*, **80** (2), 254–260.

35 Giraldi, C., Noto, A., Tenuta, R., Greco, F., Perugini, D., Spadafora, M., Bianco, A. M., Savino, O., and Natale, A. (2006) A comparative evaluation between real time Roche COBas TAQMAN 48 HCV and bDNA Bayer Versant HCV 3.0. *New Microbiologica*, **29** (4), 243–250.

36 Pawlotsky, J.M., Bastie, A., Hezode, C., Lonjon, I., Darthuy, F., Remire, J., and Dhumeaux, D. (2000) Routine detection and quantification of hepatitis B virus DNA in clinical laboratories: performance of three commercial assays. *Journal of Virological Methods*, **85** (1–2), 11–21.

37 Gleaves, C.A., Welle, J., Campbell, M., Elbeik, T., Ng, V., Taylor, P.E., Kuramoto, K., Aceituno, S., Lewalski, E., Joppa, B., Sawyer, L., Schaper, C., McNairn, D., and Quinn, T. (2002) Multicenter evaluation of the Bayer VERSANT HIV-1 RNA 3.0 assay: analytical and clinical performance. *Journal of Clinical Virology*, **25** (2), 205–216.

38 Elbeik, T., Markowitz, N., Nassos, P., Kumar, U., Beringer, S., Haller, B., and Ng, V. (2004) Simultaneous runs of the Bayer VERSANT HIV-1 version 3.0 and HCV bDNA version 3. 0 quantitative assays on the system 340 platform provide reliable quantitation and improved work flow. *Journal of Clinical Microbiology*, **42** (7), 3120–3127.

39 Elbeik, T., Surtihadi, J., Destree, M., Gorlin, J., Holodniy, M., Jortani, S.A., Kuramoto, K., Ng, V., Valdes, R.Jr., Valsamakis, A., and Terrault, N.A. (2004) Multicenter evaluation of the performance characteristics of the Bayer VERSANT HCV RNA 3.0 assay (bDNA). *Journal of Clinical Microbiology*, **42** (2), 563–569.

40 Tsongalis, G.J. (2006) Branched DNA technology in molecular diagnostics. *American Journal of Clinical Pathology*, **126** (3), 448–453.

41 Zaravinos, A., Mammas, I.N., Sourvinos, G., and Spandidos, D.A. (2009) Molecular detection methods of human papillomavirus (HPV). *International Journal of Biological Markers*, **24** (4), 215–222.

42 Van Der Pol, B., Williams, J.A., Smith, N. J., Batteiger, B.E., Cullen, A.P., Erdman, H., Edens, T., Davis, K., Salim-Hammad, H., Chou, V.W., Scearce, L., Blutman, J., and Payne, W.J. (2002) Evaluation of the Digene Hybrid Capture II Assay with the Rapid Capture System for detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae*. *Journal of Clinical Microbiology*, **40** (10), 3558–3564.

43 Darwin, L.H., Cullen, A.P., Arthur, P.M., Long, C.D., Smith, K.R., Girdner, J.L., Hook, E.W.3rd, Quinn, T.C., and Lorincz, A.T. (2002) Comparison of Digene Hybrid Capture 2 and conventional culture for detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* in cervical specimens. *Journal of Clinical Microbiology*, **40** (2), 641–644.

44 Pancholi, P., Wu, F., and Della-Latta, P. (2004) Rapid detection of cytomegalovirus infection in transplant patients. *Expert Review of Molecular Diagnostics*, **4** (2), 231–242.

45 Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, **74** (12), 5463–5467.

46 Relman, D.A. (2011) Microbial genomics and infectious diseases. *New England Journal of Medicine*, **365** (4), 347–357.

47 Arens, M. (2001) Clinically relevant sequence-based genotyping of HBV, HCV, CMV, and HIV. *Journal of Clinical Virology*, **22** (1), 11–29.

48 Martro, E., Gonzalez, V., Buckton, A.J., Saludes, V., Fernandez, G., Matas, L., Planas, R., and Ausina, V. (2008) Evaluation of a new assay in comparison with reverse hybridization and sequencing methods for hepatitis C virus genotyping targeting both 5′ noncoding and nonstructural 5b genomic regions. *Journal of Clinical Microbiology*, **46** (1), 192–197.

49 Dunn, D.T., Coughlin, K., and Cane, P.A. (2011) Genotypic resistance testing in routine clinical care. *Current Opinion in HIV and AIDS*, **6** (4), 251–257.

50 Novais, R.C. and Thorstenson, Y.R. (2011) The evolution of Pyrosequencing® for microbiology: from genes to genomes. *Journal of Microbiological Methods*, **86** (1), 1–7.

51 Lacoma, A., Garcia-Sierra, N., Prat, C., Maldonado, J., Ruiz-Manzano, J., Haba, L., Gavin, P., Samper, S., Ausina, V., and Dominguez, J. (2012) GenoType MTBDRsl for molecular detection of second-line-drug and ethambutol resistance in *Mycobacterium tuberculosis* strains and clinical samples. *Journal of Clinical Microbiology*, **50** (1), 30–36.

52 Heller, L.C., Jones, M., and Widen, R.H. (2008) Comparison of DNA pyrosequencing with alternative methods for identification of mycobacteria. *Journal of Clinical Microbiology*, **46** (6), 2092–2094.

53 Luna, R.A., Fasciano, L.R., Jones, S.C., Boyanton, B.L.Jr., Ton, T.T., and Versalovic, J. (2007) DNA pyrosequencing-based bacterial pathogen identification in a pediatric hospital setting. *Journal of Clinical Microbiology*, **45** (9), 2985–2992.

54 Boyanton, B.L.Jr., Luna, R.A., Fasciano, L. R., Menne, K.G., and Versalovic, J. (2008) DNA pyrosequencing-based identification of pathogenic *Candida* species by using the internal transcribed spacer 2 region. *Archives of Pathology & Laboratory Medicine*, **132** (4), 667–674.

55 Tenover, F.C. (2007) Rapid detection and identification of bacterial pathogens using novel molecular technologies: infection control and beyond. *Clinical Infectious Diseases*, **44** (3), 418–423.

56 Deyde, V.M. and Gubareva, L.V. (2009) Influenza genome analysis using pyrosequencing method: current applications for a moving target. *Expert Review of Molecular Diagnostics*, **9** (5), 493–509.

57 Haanpera, M., Forssten, S.D., Huovinen, P., and Jalava, J. (2008) Typing of SHV extended-spectrum beta-lactamases by pyrosequencing in *Klebsiella pneumoniae* strains with chromosomal SHV beta-lactamase. *Antimicrobial Agents and Chemotherapy*, **52** (7), 2632–2635.

58 Nyberg, S.D., Osterblad, M., Hakanen, A. J., Huovinen, P., Jalava, J., and Resistance, T.F. (2007) Detection and molecular genetics of extended-spectrum beta-lactamases among cefuroxime-resistant *Escherichia coli* and *Klebsiella* spp. isolates from Finland, 2002–2004. *Scandinavian Journal of Infectious Diseases*, **39** (5), 417–424.

59 Thulin, S., Olcen, P., Fredlund, H., and Unemo, M. (2008) Combined real-time PCR and pyrosequencing strategy for objective, sensitive, specific, and high-throughput identification of reduced susceptibility to penicillins in *Neisseria meningitidis*. *Antimicrobial Agents and Chemotherapy*, **52** (2), 753–756.

60 Bwanga, F., Hoffner, S., Haile, M., and Joloba, M.L. (2009) Direct susceptibility testing for multi drug resistant tuberculosis: a meta-analysis. *BMC Infectious Diseases*, **9**, 67.

61 Acuna-Villaorduna, C., Vassall, A., Henostroza, G., Seas, C., Guerra, H., Vasquez, L., Morcillo, N., Saravia, J., O'Brien, R., Perkins, M.D., Cunningham, J., Llanos-Zavalaga, L., and Gotuzzo, E. (2008) Cost-effectiveness analysis of introduction of rapid, alternative methods to identify multidrug-resistant tuberculosis in middle-income countries. *Clinical Infectious Diseases*, **47** (4), 487–495.

62 Stuyver, L., Rossau, R., Wyseur, A., Duhamel, M., Vanderborght, B., Van Heuverswyn, H., and Maertens, G. (1993) Typing of hepatitis C virus isolates and characterization of new subtypes using a line probe assay. *Journal of General Virology*, **74** (Pt 6), 1093–1102.

63 Stuyver, L., Wyseur, A., Rombout, A., Louwagie, J., Scarcez, T., Verhofstede, C., Rimland, D., Schinazi, R.F., and Rossau, R. (1997) Line probe assay for rapid detection of drug-selected mutations in the human immunodeficiency virus type 1 reverse transcriptase gene. *Antimicrobial Agents and Chemotherapy*, **41** (2), 284–291.

64 Thomas, L.B., Foulis, P.R., Mastorides, S. M., Djilan, Y.A., Skinner, O., and Borkowski, A.A. (2012) Hepatitis C genotype analysis: results in a large veteran population with review of the implications for clinical practice. *Annals of Clinical and Laboratory Science*, **42** (4), 355–362.

65 Guirgis, B.S., Abbas, R.O., and Azzazy, H. M. (2010) Hepatitis B virus genotyping: current methods and clinical implications. *International Journal of Infectious Diseases*, **14** (11), e941–953.

66 Soltermann, A., Koetzer, S., Eigenmann, F., and Komminoth, P. (2007) Correlation of *Helicobacter pylori* virulence genotypes *vacA* and *cagA* with histological parameters of gastritis and patient's age. *Modern Pathology*, **20** (8), 878–883.

67 Donatin, E. and Drancourt, M. (2012) DNA microarrays for the diagnosis of infectious diseases. *Médecine et Maladies Infectieuses*, **42** (10), 453–459.

68 Chandler, D.P., Bryant, L., Griesemer, S. B., Gu, R., Knickerbocker, C., Kukhtin, A., Parker, J., Zimmerman, C., St. George, K., and Cooney, C.G. (2012) Integrated amplification microarrays for infectious disease diagnostics. *Microarrays*, **1**, 107–124.

69 Nichol, S.T., Spiropoulou, C.F., Morzunov, S., Rollin, P.E., Ksiazek, T.G., Feldmann, H., Sanchez, A., Childs, J., Zaki, S., and Peters, C.J. (1993) Genetic identification of a hantavirus associated with an outbreak of acute respiratory illness. *Science*, **262** (5135), 914–917.

70 Tissari, P., Zumla, A., Tarkka, E., Mero, S., Savolainen, L., Vaara, M., Aittakorpi,

A., Laakso, S., Lindfors, M., Piiparinen, H., Maki, M., Carder, C., Huggett, J., and Gant, V. (2010) Accurate and rapid identification of bacterial species from positive blood cultures with a DNA-based microarray platform: an observational study. *Lancet*, **375** (9710), 224–230.

71 Teo, J., Di Pietro, P., San Biagio, F., Capozzoli, M., Deng, Y.M., Barr, I., Caldwell, N., Ong, K.L., Sato, M., Tan, R., and Lin, R. (2011) VereFlu: an integrated multiplex RT-PCR and microarray assay for rapid detection and identification of human influenza A and B viruses using lab-on-chip technology. *Archives of Virology*, **156** (8), 1371–1378.

72 Zhang, Z., Li, L., Luo, F., Cheng, P., Wu, F., Wu, Z., Hou, T., Zhong, M., and Xu, J. (2012) Rapid and accurate detection of RMP- and INH-resistant *Mycobacterium tuberculosis* in spinal tuberculosis specimens by CapitalBio DNA microarray: a prospective validation study. *BMC Infectious Diseases*, **12**, 303.

73 Zhu, L., Jiang, G., Wang, S., Wang, C., Li, Q., Yu, H., Zhou, Y., Zhao, B., Huang, H., Xing, W., Mitchelson, K., Cheng, J., Zhao, Y., and Guo, Y. (2010) Biochip system for rapid and accurate identification of mycobacterial species from isolates and sputum. *Journal of Clinical Microbiology*, **48** (10), 3654–3660.

74 You, Y., Fu, C., Zeng, X., Fang, D., Yan, X., Sun, B., Xiao, D., and Zhang, J. (2008) A novel DNA microarray for rapid diagnosis of enteropathogenic bacteria in stool specimens of patients with diarrhea. *Journal of Microbiological Methods*, **75** (3), 566–571.

75 Kim, D.H., Lee, B.K., Kim, Y.D., Rhee, S. K., and Kim, Y.C. (2010) Detection of representative enteropathogenic bacteria, *Vibrio* spp., pathogenic *Escherichia coli*, *Salmonella* spp., *Shigella* spp., and *Yersinia enterocolitica*, using a virulence factor gene-based oligonucleotide microarray. *Journal of Microbiology (Seoul, Korea)*, **48** (5), 682–688.

76 Wang, Q., Wang, S., Beutin, L., Cao, B., Feng, L., and Wang, L. (2010) Development of a DNA microarray for detection and serotyping of enterotoxigenic *Escherichia coli*. *Journal of Clinical Microbiology*, **48** (6), 2066–2074.

77 Borel, N., Kempf, E., Hotzel, H., Schubert, E., Torgerson, P., Slickers, P., Ehricht, R., Tasara, T., Pospischil, A., and Sachse, K. (2008) Direct identification of chlamydiae from clinical samples using a DNA microarray assay: a validation study. *Molecular and Cellular Probes*, **22** (1), 55–64.

78 Garcia-Sierra, N., Martro, E., Castella, E., Llatjos, M., Tarrats, A., Bascunana, E., Diaz, R., Carrasco, M., Sirera, G., Matas, L., and Ausina, V. (2009) Evaluation of an array-based method for human papillomavirus detection and genotyping in comparison with conventional methods used in cervical cancer screening. *Journal of Clinical Microbiology*, **47** (7), 2165–2169.

79 Sauer, S. and Kliem, M. (2010) Mass spectrometry tools for the classification and identification of bacteria. *Nature Reviews Microbiology*, **8** (1), 74–82.

80 Jordana-Lluch, E., Martro Catala, E., and Ausina Ruiz, V. (2012) Mass spectrometry in the clinical microbiology laboratory. *Enfermedades Infecciosas y Microbiología Clínica*, **30** (10), 635–644.

81 Honisch, C., Chen, Y., Mortimer, C., Arnold, C., Schmidt, O., van den Boom, D., Cantor, C.R., Shah, H.N., and Gharbia, S.E. (2007) Automated comparative sequence analysis by base-specific cleavage and mass spectrometry for nucleic acid-based microbial typing. *Proceedings of the National Academy of Sciences of the United States of America*, **104** (25), 10649–10654.

82 Malakhova, M.V., Ilina, E.N., Govorun, V. M., Shutko, S.A., Dudina, K.R., Znoyko, O.O., Klimova, E.A., and Iushchuk, N.D. (2009) Hepatitis B virus genetic typing using mass-spectrometry. *Bulletin of Experimental Biology and Medicine*, **147** (2), 220–225.

83 Ilina, E.N., Malakhova, M.V., Generozov, E.V., Nikolaev, E.N., and Govorun, V.M. (2005) Matrix-assisted laser desorption ionization-time of flight (mass spectrometry) for hepatitis C virus

genotyping. *Journal of Clinical Microbiology*, **43** (6), 2810–2815.
84 Ecker, D.J., Sampath, R., Massire, C., Blyn, L.B., Hall, T.A., Eshoo, M.W., and Hofstadler, S.A. (2008) Ibis T5000: a universal biosensor approach for microbiology. *Nature Reviews Microbiology*, **6** (7), 553–558.
85 Massire, C., Ivy, C.A., Lovari, R., Kurepina, N., Li, H., Blyn, L.B., Hofstadler, S.A., Khechinashvili, G., Stratton, C.W., Sampath, R., Tang, Y.W., Ecker, D.J., and Kreiswirth, B.N. (2011) Simultaneous identification of mycobacterial isolates to the species level and determination of tuberculosis drug resistance by PCR followed by electrospray ionization mass spectrometry. *Journal of Clinical Microbiology*, **49** (3), 908–917.
86 Jordana-Lluch, E., Carolan, H.E., Gimenez, M., Sampath, R., Ecker, D.J., Quesada, M.D., Modol, J.M., Armestar, F., Blyn, L.B., Cummins, L.L., Ausina, V., and Martro, E. (2013) Rapid diagnosis of bloodstream infections with PCR followed by mass spectrometry. *PLoS One*, **8** (4), e62108.
87 Hujer, K.M., Hujer, A.M., Endimiani, A., Thomson, J.M., Adams, M.D., Goglin, K., Rather, P.N., Pennella, T.T., Massire, C., Eshoo, M.W., Sampath, R., Blyn, L.B., Ecker, D.J., and Bonomo, R.A. (2009) Rapid determination of quinolone resistance in *Acinetobacter* spp. *Journal of Clinical Microbiology*, **47** (5), 1436–1442.
88 Faix, D.J., Sherman, S.S., and Waterman, S.H. (2009) Rapid-test sensitivity for novel swine-origin influenza A (H1N1) virus in humans. *New England Journal of Medicine*, **361** (7), 728–729.
89 Capobianchi, M.R., Giombini, E., and Rozera, G. (2013) Next-generation sequencing technology in clinical virology. *Clinical Microbiology and Infection*, **19** (1), 15–22.
90 Metzker, M.L. (2005) Emerging technologies in DNA sequencing. *Genome Research*, **15** (12), 1767–1776.
91 Metzker, M.L. (2010) Sequencing technologies – the next generation. *Nature Reviews Genetics*, **11** (1), 31–46.
92 Eyre, D.W., Golubchik, T., Gordon, N.C., Bowden, R., Piazza, P., Batty, E.M., Ip, C. L., Wilson, D.J., Didelot, X., O'Connor, L., Lay, R., Buck, D., Kearns, A.M., Shaw, A., Paul, J., Wilcox, M.H., Donnelly, P.J., Peto, T.E., Walker, A.S., and Crook, D.W. (2012) A pilot study of rapid benchtop sequencing of *Staphylococcus aureus* and *Clostridium difficile* for outbreak detection and surveillance. *BMJ Open*, **2**, e001124.
93 Koser, C.U., Holden, M.T., Ellington, M. J., Cartwright, E.J., Brown, N.M., Ogilvy-Stuart, A.L., Hsu, L.Y., Chewapreecha, C., Croucher, N.J., Harris, S.R., Sanders, M., Enright, M.C., Dougan, G., Bentley, S.D., Parkhill, J., Fraser, L.J., Betley, J.R., Schulz-Trieglaff, O.B., Smith, G.P., and Peacock, S.J. (2012) Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *New England Journal of Medicine*, **366** (24), 2267–2275.
94 McAdam, P.R., Holmes, A., Templeton, K.E., and Fitzgerald, J.R. (2011) Adaptive evolution of Staphylococcus aureus during chronic endobronchial infection of a cystic fibrosis patient. *PLoS One*, **6** (9), e24301.
95 Willerth, S.M., Pedro, H.A., Pachter, L., Humeau, L.M., Arkin, A.P., and Schaffer, D.V. (2010) Development of a low bias method for characterizing viral populations using next generation sequencing technology. *PLoS One*, **5** (10), e13564.
96 Rodriguez, C., Chevaliez, S., Bensadoun, P., and Pawlotsky, J.M. (2013) Characterization of the dynamics of hepatitis B virus resistance to adefovir by ultra-deep pyrosequencing. *Hepatology*, **58** (3), 890–901.
97 Chevaliez, S., Rodriguez, C., and Pawlotsky, J.M. (2012) New virologic tools for management of chronic hepatitis B and C. *Gastroenterology*, **142** (6), 1303–1313.
98 Alteri, C., Santoro, M.M., Abbate, I., Rozera, G., Bruselles, A., Bartolini, B., Gori, C., Forbici, F., Orchi, N., Tozzi, V., Palamara, G., Antinori, A., Narciso, P., Girardi, E., Svicher, V., Ceccherini-Silberstein, F., Capobianchi, M.R., and Perno, C.F. (2011) 'Sentinel' mutations in

standard population sequencing can predict the presence of HIV-1 reverse transcriptase major mutations detectable only by ultra-deep pyrosequencing. *Journal of Antimicrobial Chemotherapy*, **66** (11), 2615–2623.

99 Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D.R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J.M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H.B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Jian, M., Zhou, Y., Li, Y., Zhang, X., Qin, N., Yang, H., Wang, J., Brunak, S., Dore, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Bork, P., and Ehrlich, S. D. (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, **464** (7285), 59–65.

100 de Vries, M., Deijs, M., Canuti, M., van Schaik, B.D., Faria, N.R., van de Garde, M.D., Jachimowski, L.C., Jebbink, M.F., Jakobs, M., Luyf, A.C., Coenjaerts, F.E., Claas, E.C., Molenkamp, R., Koekkoek, S. M., Lammens, C., Leus, F., Goossens, H., Ieven, M., Baas, F., and van der Hoek, L. (2011) A sensitive assay for virus discovery in respiratory clinical samples. *PLoS One*, **6** (1), e16118.

101 Towner, J.S., Sealy, T.K., Khristova, M.L., Albarino, C.G., Conlan, S., Reeder, S.A., Quan, P.L., Lipkin, W.I., Downing, R., Tappero, J.W., Okware, S., Lutwama, J., Bakamutumaho, B., Kayiwa, J., Comer, J. A., Rollin, P.E., Ksiazek, T.G., and Nichol, S.T. (2008) Newly discovered Ebola virus associated with hemorrhagic fever outbreak in Uganda. *PLoS Pathog*, **4** (11), e1000212.

10
MicroRNAs in Human Microbial Infections and Disease Outcomes

Verónica Saludes, Irene Latorre, Andreas Meyerhans, and Juana Díez

10.1
Introduction

Searching for "miRNA and infection" in PubMed (www.ncbi.nlm.nih.gov/pubmed) gives 1132 references (July 2013), showing the importance of microRNAs (miRNAs) in human microbial infections. This chapter reviews, after a brief introduction, general aspects of miRNAs involved in bacterial and viral infectious diseases. The major focus will be on *Mycobacterium tuberculosis* bacteria and hepatitis C virus (HCV) infections – two major pathogens that threaten global health.

Technological advances in miRNA pattern detection (the "miRNome") and its correlation with different human pathologies has boosted biomarker research in recent years. miRNAs are small non-coding single-stranded RNAs (17–24 nucleotides long) that regulate post-transcriptional gene expression through translational repression and messenger RNA (mRNA) degradation. They control different aspects of cellular activity, including metabolism, cell proliferation, apoptosis, cell differentiation, and inflammatory responses [1–3]. More than 2000 human miRNAs have been described (see miRBase v19, www.mirbase.org) and it is known that approximately 60% of mRNAs are controlled by miRNAs [4]. The deregulation of miRNA expression patterns has been associated with clinical manifestations of major public health concern, such as cancer, diabetes, cardiac and neurodegenerative diseases, viral infections, immune dysfunction, and fibroproliferative disorders [5–7]. Although most of the respective studies analyzing miRNA expression patterns have been carried out with tissue samples from biopsies, more recently it has been demonstrated that deregulated miRNA expression patterns in peripheral blood cells are correlated with several cancer types [8]. Moreover, miRNAs can be found in a stable form in serum, inside exosomes, or in association with proteins protected from endogenous RNase activity. These miRNAs are released to serum from injured tissues after cell death or from other cells that actively secrete them, likely to act as messengers biomolecules [3,9]. Together, this opens the very attractive possibility to develop simple miRNA-based blood tests for clinical diagnosis and prognosis.

Nucleic Acids as Molecular Diagnostics, First Edition. Edited by Andreas Keller and Eckart Meese.

miRNAs are important regulatory components of immune system development and function, and can modulate innate and adaptive immune responses against pathogens. The innate immunity is the first barrier of defense against microorganisms, and mainly involves immune cells such as monocyte/macrophages, granulocytes, and dendritic cells that respond to pathogens after recognizing pathogen-associated molecular patterns (PAMPs) through host pattern recognition receptors called Toll-like receptors (TLRs) or RIG-1-like receptors. The binding of bacterial or viral ligands to pattern recognition receptors triggers the initiation of the host innate immune responses, partly mediated by transcription factors such as the nuclear factor-κB (NF-κB). Specifically, the recognition of viral PAMPs triggers the interferon (IFN) signaling system, which is a main defense mechanism against viruses. Adaptive immunity is predominantly mediated by lymphocytes and involves the specific recognition of pathogen components. miRNAs have a role in TLR signaling, cytokine production, B and T cell lymphocyte differentiation, T cell receptor signaling, regulatory T cell function, and antigen presentation [10,11].

miRNAs are involved in controlling TLR signaling and innate immune pathways. Numerous miRNAs are induced in innate immune cells by TLR activation, targeting the 3′-untranslated regions of mRNAs involved in the innate immune system. The most consistent and ubiquitous miRNAs induced by innate immune activation following TLR activation are the trio miR-155, miR-146 and miR-21. In a pioneering work conducted by David Baltimore's group, 200 miRNAs were profiled in human monocytes after TLR4 stimulation by lipopolysaccharide (LPS). Several miRNAs, such as miR-146a/b, miR-132, and miR-155, had an induced expression by LPS. Furthermore, miR-146a/b and miR-155 were induced in a NF-κB-dependent manner [12]. miRNAs play also a crucial role in T cell biology – an essential player in adaptive immune response. Deletion of a key protein involved in miRNA processing (Dicer protein) in mice resulted in an impaired T cell development, aberrant T helper cell differentiation, and cytokine production [13]. Furthermore, miRNAs are also very important in B cell development. For example, the deletion of Dicer led to arrest of the development of early B cells at a pro-B cell stage [14].

10.2
General Aspects of miRNAs in Infectious Diseases

10.2.1
miRNAs in Bacterial Infections

Bacterial pathogens control a large range of host cellular functions, such as modulation of host miRNAs, to ensure their survival and replication. The study of miRNAs in bacterial infections is not as well explored as in viral infections. The majority of the existing studies on the modulation of host miRNAs by bacteria are described for important pathogens such as *Helicobacter pylori*, *Salmonella*

spp., *Listeria monocytogenes*, or *M. tuberculosis*. Regarding *H. pylori* infections, several miRNAs have been related to gastric inflammatory stages and cancer caused by these Gram-negative bacteria [15]. In this context, some studies have evidenced the role of miR-155 as a prominent miRNA that modulates the inflammatory response in *H. pylori* infection [16–19]. It has also been described that the expression of miR-146a is upregulated by this pathogen in gastric epithelial cells and mucosa tissues in a NF-κB-dependent manner. In addition, upregulation of these two important miRNAs has been observed likewise in response to the intracellular infection caused by *Salmonella* [11,20–22]. Schulte *et al.* [21] demonstrated that members of the let-7 miRNA family were downregulated by *Salmonella* in macrophages and epithelial cells. This set of miRNAs is involved in the repression of interleukin-6 and -10 cytokine release; therefore, their modulation resulted in the induction of both interleukins. Similarly to Gram-negative bacteria, the intracellular Gram-positive pathogen *L. monocytogenes* induces significant changes in miRNA expression patterns, in particular those implicated in the regulation of immune-related genes, such as members of the miR-155, miR-146a/b, and let-7 miRNA family [23–25]. A recent study has reported that mice lacking miR-155 had an impaired $CD8^+$ T cell response to *L. monocytogenes* infection, suggesting that this miRNA could be a good target for modulating the immune response [25]. Moreover, Ma *et al.* [26] have elegantly demonstrated that mice infected with *L. monocytogenes* or *Mycobacterium bovis* bacillus Calmette-Guérin (BCG) downregulated the expression of miR-29 in natural killer (NK) cells. In addition, this miRNA inhibited the expression of IFN-γ mRNA in NK cells. Furthermore, regulation of miRNAs in response to *M. tuberculosis* infections is currently being explored.

An uncontrolled bacterial infection commonly results in the onset of sepsis. An *in vivo* profile of human leukocyte miRNA response to endotoxemia triggered by *Escherichia coli* LPS infusion revealed that miR-146b, miR-150, miR-342, and let-7 g were downregulated, and miR-143 was upregulated [27]. By correlating with the transcriptome expression in leukocytes, the predicted inflammation-related target genes contributed to our knowledge on the innate immunity initiated by pathogens.

Taken together, the available studies show that miR-155 and miR-146 are the most commonly induced miRNAs during bacterial infections. They are dependent on the NF-κB pathway. The identification of host miRNAs induced after bacterial infections is very important for the development of novel therapies against these pathogens. However, at this point, most studies are solely descriptive and do not provide a mechanistic understanding of the regulation of miRNAs in bacterial infections.

10.2.2
miRNAs in Viral Infections

miRNAs play a significant regulatory role at different levels of viral infection. While in some infections cellular miRNAs are part of the host antiviral response,

in others viruses exploit cellular miRNAs for their own benefit to either avoid the host antiviral response or promote their life cycles. Numerous DNA viruses, especially those belonging to the herpesvirus family, encode miRNAs able to evade the host immune response or to promote viral pathogenesis (Table 10.1).

10.2.2.1 Cellular miRNAs Control Viral Infections

Effective virus recognition and subsequent activation of antiviral innate immunity are essential to control an infection. All the steps are highly regulated by multiple mechanisms, including cellular miRNAs. Type I IFN signaling, which plays critical roles in the antiviral innate immunity, regulates miRNA expression

Table 10.1 miRNAs encoded by human pathogenic viruses involved in the evasion of the innate immune system.

Virus family or subfamily	Virus-encoded miRNA	Targeted innate immunity factor	Function	Reference
HCMV (β-herpesvirinae)	miR-UL112	MICB	NK cell inhibition	[28]
	miR-US4-1	ERAP1	non-presentation of antigenic HCMV peptides to $CD8^+$ T cells	[29]
	miR-UL148D	RANTES	prevention of the recruitment of immune effector cells	[30]
HHV-8 (γ-herpesvirinae)	miR-K7	MICB	NK cell inhibition	[31]
	miR-K10	TWEAKR	apoptosis inhibition	[32]
	miR-K9	IRAK1	TLR signaling inhibition	[33]
	miR-K5	MyD88	TLR signaling inhibition	[33]
	miR-K10	TGBRII	apoptosis inhibition	[34]
	miR-K1, miR-K3, miR-K4	caspase-3	apoptosis inhibition	[35]
	miR-K1	p21	inhibition of cell cycle arrest	[36]
EBV (γ-herpesvirinae)	miR-BART5, miR-BART19	PUMA	apoptosis inhibition	[37,38]
	miR-BART4, miR-BART15	BIM	apoptosis inhibition	[39,40]
	miR-BART1, miR-BART3, miR-BART5, miR-BART9	MICB	NK cell inhibition	[39,41]
JC and BK virus (*Polyomaviridae*)	miR-J1-3p	ULBP3	NK cell inhibition	[42]

and miRNAs regulate type I IFN signaling, together in a regulatory loop. The first report showing a direct control of a viral infection by cellular miRNAs regulated by type I IFN signaling was on HCV infection [43]. In liver cells, IFN-β led to the upregulation of miRNAs with anti-HCV activity, such as miR-196, miR-296, miR-351, miR-431, and miR-448, and to the downregulation of miR-122, which is a promoter of HCV replication. These miRNAs directly target the HCV genome. Later, it was demonstrated that miR-155 expression induced in macrophages by vesicular stomatitis virus (VSV) infection enhances type I IFN signaling by targeting SOCS1 (suppressor of cytokine signaling 1) mRNA [44]. Additionally, type I IFN signaling regulates the expression of the enzyme Dicer [45].

Several cellular miRNAs have been associated with the control of viral infections by directly targeting the viral genome. Among them, miR-323, miR-491, and miR-654 control H1N1 influenza A virus replication by degrading the mRNA encoding for the viral PB1 polymerase subunit in infected cells [46]. In the case of retroviruses, it has been described that the highly abundant miR-29a suppresses HIV-1 mRNA expression in human T lymphocytes [47] and miR-32 inhibits primate foamy virus type 1 replication in human cells [48].

10.2.2.2 Viruses Use miRNAs for Their Own Benefit

Viral Evasion of Host Antiviral Response by miRNAs Many viruses have evolved different mechanisms to resist the antiviral innate immune response, including the use of miRNAs encoded by the host or their own genome. Regarding the use of cellular miRNAs for viral survival, VSV infection enhances miR-146a expression in macrophages and this miRNA downregulates type I IFN production [49]. In contrast, Epstein–Barr virus (EBV) infection downregulates the expression of miR-146a in order to produce a constant low level of IFN to make the virus-infected cells refractory to this cytokine [50]. Human herpesvirus (HHV)-8 infection upregulates miR-132 expression in lymphatic endothelial cells, which negatively regulates the expression of type I IFN-stimulated genes by suppressing a transcriptional co-activator [51]. It has also been described that miR-451 negatively regulates dendritic cell cytokine responses to influenza virus infection, enhancing its replication [52]. Several DNA viruses encode miRNAs that downregulate the expression of cellular factors involved in the innate immune system. Examples are EBV, human cytomegalovirus (HCMV), HHV-8, JC virus, and BK virus. The respective observations are summarized in the Table 10.1.

miRNAs Involved in Viral Pathogenesis Both host and virus-encoded miRNAs can regulate viral infections at different levels to enhance pathogenicity. Regarding the former, the most widely reported case is the use of the liver-specific miR-122 by HCV to promote its life cycle. miR-122 positively regulates HCV translation and replication by binding to the viral genome [53,54]. This finding has important clinical implications as a novel antagonist targeting the miR-122 has shown potent antiviral effects in chronic

HCV-infected patients in clinical trials [55]. HIV-1 mRNAs are targeted by a cluster of cellular miRNAs (miR-28, miR-125b, miR-150, miR-223, and miR-382) in resting $CD4^+$ T cells involved in proviral latency [56]. This discovery suggested that manipulation of cellular miRNAs could represent a novel approach for purging the HIV-1 reservoir. In the same way, EBV is able to upregulate miR-155 in B cells to promote cell immortalization and maintenance of EBV genomes during latent infection [57]. More recently, it was shown that enterovirus 71 (EV71) upregulates miR-141, which inhibits cap-dependent translation and consequently increases viral propagation [58]. Among virally encoded miRNAs, the herpes simplex virus (HSV)-1-encoded miR-LAT, which is generated from the exon 1 region of the HSV-1 *LAT* gene, promotes viral latency by targeting components of the transforming growth factor (TGF)-β pathway [59], while miR-H6 and miR-H2-3p promote viral latency by post-transcriptionally regulating viral gene expression [60]. Another virus-encoded miRNA involved in latency is the HCMV-encoded miR-UL112-1, which regulates the expression of viral genes involved in replication [61]. Finally, the HHV-8 genome encodes multiple miRNAs that downregulate the expression of a tumor suppressor factor, thus contributing to tumorigenesis [62].

10.3 miRNAs as Biomarkers and Therapeutic Agents in Tuberculosis and Hepatitis C Infections

10.3.1 Tuberculosis: A Major Bacterial Pathogen

10.3.1.1 Tuberculosis Diagnosis and the Need for Immunological Biomarkers

M. tuberculosis is a major microbial pathogen that threatens global health. About one-third of the world's population carries the bacteria, albeit only a minority will ever develop tuberculosis (TB) [63]. Nevertheless, WHO estimates gave a number of 8.7 million new TB cases for 2011 and around 1.4 million deaths due to TB per year [64], making it the most deadly bacterial pathogen of humans.

In *M. tuberculosis*-infected individuals, failure of clearance by the immune system results in different outcomes. Bacteria may exist in a quiescent state for prolonged periods of time (latent *M. tuberculosis* infection) or may ultimately start multiplying and escape immune control, resulting in clinically active TB in up to 10% of infected cases [63,65]. Live *M. tuberculosis* may also persist after initial successful treatment of active TB, and dynamic interactions between the host and pathogen determine whether relapse to active disease will occur or not [65].

One essential factor for controlling the spread of TB is the ability to diagnose it at an early stage and to supply adequate treatment. However, the commonly used test systems are still insufficient. For example, TB diagnoses by microscopy or tuberculin skin testing (TST) have a variable and low sensitivity ranging from

20 to 60% for microscopy [66,67] and from 67 to 80% for TST [68,69], respectively. Tests based on *M. tuberculosis* cultivation, the current gold standard of TB diagnosis, can take several weeks, requiring incubation periods of up to 2 months [70]. A rapid nucleic acid *M. tuberculosis* gene amplification test (GeneXpert MTB/RIF) has recently been developed, and holds promise for improved diagnosis of active TB and detection of rifampicin resistance [71,72]. However, the use of this system cannot detect latent *M. tuberculosis* infection and awaits further validation in a point-of-care setting. Potential host biomarkers are therefore needed to diagnose TB. The ideal TB biomarker should provide correlates of risk to TB, protection against active disease, and success of treatment. Today, owing to the limited knowledge of promising TB biomarkers, global "omics" approaches are attractive options to follow and the initial studies in this direction provide encouraging results [65].

10.3.1.2 miRNAs Regulation in Response to M. tuberculosis

Given the great success of recent miRNA-based diagnostic studies in various human pathologies, encouraging results from miRNA studies have appeared for bacterial infections like *M. tuberculosis*. A summary of our current knowledge on regulated host miRNAs in response to *M. tuberculosis* is given in Table 10.2. The first two studies on miRNAs in active pulmonary TB appeared in 2011 [73,74]. miR-223, miR-424, miR-451, and miR-144, which are involved in hematopoietic differentiation processes, were highly expressed in peripheral blood mononuclear cells (PBMCs) of pulmonary active TB patients [73]. Additionally, 92 differential miRNAs were significantly detected in TB serum patients compared with healthy control serum [74]. Additionally, miR-29a and miR-93* were present in the serum and sputum of active TB patients.

Other studies have described the mechanism that miRNAs play in the immune response to intracellular *Mycobacteria* infection. Ma *et al.* discovered that infection of mice with BCG downregulated miR-29 in NK cells, $CD4^+$ T cells, and $CD8^+$ T cells [26]. They also showed that this miRNA suppressed the immune response to intracellular pathogens by targeting IFN-γ mRNA. A recent study demonstrated that miR-29a shows lower expression in adults and children with active TB compared with latently infected individuals [81]. In contrast to the previous study, the authors found that miR-29a expression did not correlate with *M. tuberculosis*-specific IFN-γ-expressing T cells. It has also been demonstrated that macrophages incubated with lipomannan (an external lipoglycoconjugate cell wall molecule of *M. tuberculosis*) and live *M. tuberculosis* induced miR-125b and repressed miR-155 expression [77]. miR-125b directly destabilizes the 3′-untranslated region of tumor necrosis factor (TNF) mRNA, which is a key cytokine for the control of *M. tuberculosis* infection, leading to reduced TNF production. On the contrary, miR-155 targets and degrades a transcript that negatively regulates TNF production. Recently, it has been determined that miR-99b also targets the expression of TNF-α mRNA [82]. This miRNA was upregulated in infected dendritic cells and macrophages, suggesting that *M. tuberculosis* uses miR-99b as a host evasion mechanism.

Table 10.2 Host miRNAs expression changes in response to *M. tuberculosis* infection.

miRNA changes in active TB patients	Groups of individuals compared	Type of sample	miRNA analysis technology	Reference
Up: miR-93, miR-29a	pulmonary active TB; healthy controls	serum	microarrays	[74]
Up: miR-223, miR-451, miR-424, miR-144	pulmonary active TB; latent TB infection individuals; healthy controls	PBMCs	microarrays	[73]
Up: miR-144*	pulmonary active TB; healthy controls	PBMCs	microarrays	[75]
Up: miR-155, miR-193	pulmonary active TB; healthy controls	PBMCs	microarrays	[76]
Up: miR-125b; *down*: miR-155	—	human macrophages	qRT-PCR of miR-125b and miR-155	[77]
Up: miR-144	pulmonary active TB; sarcoidosis; healthy controls	blood	microarrays	[78]
Up: miR-3179, miR-147; *down*: miR-19b-2*	Pulmonary active TB; healthy controls	sputum	microarrays	[79]
Up: miR-155	—	bone marrow-derived murine macrophages	qRT-PCR of miR-155	[80]
Down: miR-26a, miR-29a, miR-142-3p	pulmonary active TB; latent TB infection individuals	$CD4^+$ T cells; blood	qRT-PCR of 29 candidate miRNAs	[81]
Up: miR-99b	—	murine dendritic cells	microarrays	[82]
Up: miR-182, miR-197	pulmonary active TB; lung cancer; pneumonia	serum	microarrays	[83]

qRT-PCR, quantitative real-time polymerase chain reaction.

The expression profile of purified protein derivative-induced miRNAs in PBMCs from active TB patients has been explored, showing that miR-155 and miR-155* were highly expressed [76]. Specific *M. tuberculosis* antigen ESAT-6 also induces miR-155 [80]. Indeed, this miRNA was the most upregulated in *M. tuberculosis*-infected bone marrow-derived macrophages. Furthermore, miR-155 could inhibit SHIP1 (an inhibitor of AKT kinase activation). This inhibition favors AKT activation, which is essential for *M. tuberculosis* survival in macrophages. The function of miR-144* has also been explored. This miRNA has been found to be overexpressed in PBMCs of active TB patients. Additionally, transfection of T cells with miR-144* precursor RNA confirmed that it could reduce the release of IFN-γ and TNF-α, as well as T cell proliferation [75].

It is important to determine whether identified miRNAs are exclusive signatures of TB or shared with other similar inflammatory pulmonary diseases. The miRNA expression profile between active TB and sarcoidosis patients has shown similar patterns. The most strongly upregulated miRNA in these two diseases was miR-144 [78]. Moreover, a recent study found that serum levels of miR-182 and miR-197 were significantly higher in patients with lung cancer and TB compared with healthy controls [83].

10.3.1.3 Future Perspectives

The basis for controlling the spread of TB consists of the ability to diagnose it in its early stages and to treat adequately patients with active TB. Additionally, the monitoring of anti-TB treatment based solely on clinical findings is difficult in patients with TB. There is a need for a specific marker to better evaluate and/or predict the success of therapy in order to assess an adequate treatment. An elegant study conducted by Anne O'Garra's group has identified significant changes in the transcriptional signatures detectable just 2 weeks after TB treatment initiation [84]. Further work is required to evaluate whether miRNA expression patterns can predict anti-TB therapy success.

Several miRNAs have been differentially identified as potential biomarkers between active TB, latent *M. tuberculosis*, and healthy individuals. However, the results obtained are not conclusive and have to be validated satisfactorily. Currently, next-generation sequencing (NGS) has not been used for miRNA expression pattern detection. Therefore, novel miRNAs involved in *M. tuberculosis* pathogenesis could not be discovered. Finally, if an accurate differentiation between infection and disease could be obtained, it would be possible to build up a simple point-of-care test. Given the ample evidence of miRNA-dependent modulation of TB–host interactions, therapeutic options that aim to rescue these modifications should be considered eventually.

10.3.2 Chronic Hepatitis C: A Major Viral Disease

10.3.2.1 Liver Fibrosis Progression and Treatment Outcome

With a seroprevalence of 2.8% (more than 185 million people) worldwide [85], HCV infection is a leading cause of chronic hepatitis, cirrhosis, and hepatocellular carcinoma, and the main cause of liver transplantation in developed countries [86]. There is no vaccine available to prevent this infection and the antiviral treatment options are complex, having various limitations [87]. Thus, HCV infection is a serious viral threat to public health.

Fibrogenesis is an important pathogenic consequence of chronic HCV infection that leads to cirrhosis in 10–40% of the cases. Death associated with complications of cirrhosis can take place with an incidence of 4% per year. Progression to hepatocellular carcinoma has an incidence of 1–5% per year and these patients have a probability of death of 33% during the first year. Antiviral treatment against chronic hepatitis C is able to reverse fibrosis progression in

patients that respond successfully to treatment [86]. However, there is no specific and effective therapy against liver fibrosis [88]. This is especially needed for HCV patients that respond poorly to HCV therapy and for liver transplant recipients with HCV recurrence as they have an accelerated liver disease progression. Unfortunately, the molecular mechanisms involved in liver fibrogenesis are not completely understood. Therefore, a deeper understanding of the molecular mechanisms involved in liver fibrogenesis is key to finding a more effective and specific antifibrotic therapy.

For almost a decade, the standard of care against chronic hepatitis C consisted of the combined administration of pegylated IFN-α and ribavirin (PEG-IFN/RBV) [86]. This therapy is long, costly, and involves multiple severe side-effects, such as depression and hemolytic anemia, that may result in therapy discontinuation. A treatment response is considered successful when a patient achieves a sustained virological response (SVR), previously defined as having a negative viral load 24 weeks after treatment cessation. Among patients infected with HCV genotype 1 (HCV-1), the most prevalent genotype worldwide and in Europe [89], only about 40–54% of the cases achieve such a SVR in comparison with patients infected with the HCV genotypes 2 or 3 (65–82%) [86]. To improve the SVR rates in chronic HCV-1-infected patients, different compounds with direct antiviral activity have been developed or are still under development. Triple therapies with the HCV protease inhibitors boceprevir or telaprevir together with PEG-IFN/RBV improved the SVR rates up to around 70% [90–92] and thus it recently became the new standard of care for the treatment of chronic HCV-1 infections [93]. However, triple therapy involves more frequent side-effects (anemia, rash) and considerably increases the costs associated with PEG-IFN/RBV dual therapy. In addition, the use of drugs directly targeting viral proteins may lead to the selection of resistant virus variants that are associated with virological failure and relapse. Furthermore, they can cause pharmacological interactions with drugs metabolized in the liver through cytochrome P450 3A4. Thus, a reliable prediction of treatment outcomes would avoid the unnecessary use of triple therapy with its associated costs and secondary effects, and would be a step forward towards personalized treatment regimens.

Altogether, two important priorities in the HCV field are: (i) to better understand the molecular mechanisms involved in liver fibrogenesis in order to develop antifibrotic therapies and (ii) to define reliable predictors of treatment outcome in chronic HCV-1-infected patients, since although novel triple therapies have increased the success rate, even the rich countries cannot afford the associated costs. Similar to TB infection, the use of global "omics" approaches to study liver fibrogenesis and to predict treatment outcome in chronic HCV-infected patients is a promising direction to follow, and several studies have already shown encouraging results.

10.3.2.2 miRNAs Involved in Liver Fibrogenesis

Deregulation of miRNAs play a key role in many chronic liver diseases, including viral hepatitis C and B, alcoholic and non-alcoholic fatty liver disease, and in associated liver necroinflammatory activity, progression to cirrhosis and

Table 10.3 Deregulated expression of miRNAs associated with the stage of liver fibrosis in chronic HCV-infected patients.

miRNA changes	Associated stage of liver fibrosis	Groups of individuals compared	Type of sample	miRNA analysis technology	Reference
Up: miR-20a, miR-92a	cirrhosis and advanced fibrosis	cirrhosis and advanced fibrosis versus early fibrosis stage	serum	microarrays	[109]
Up: miR-199a-5p, miR-199a-3p, miR-221, miR-222	advanced fibrosis	advanced versus early fibrosis stage	liver	microarrays	[110]
Up: miR-483-5p, miR-671-5p; *down*: Let-7a, miR-106b, miR-1274a, miR-130b, miR-140-3p, miR-151-3p, miR-181a, miR-19b, miR-21, miR-24, miR-375, miR-5481, miR-93, miR-941	advanced fibrosis	fibrosis versus non-fibrosis	serum	microarrays	[111]
Up: miR-513-3p, miR-571; *down*: miR-652	cirrhosis	fibrosis versus cirrhosis	serum	microarrays	[98]
Up: miR-199a, miR-199a*, miR-200a, miR-200b	advanced fibrosis	advanced versus early fibrosis stage	liver	microarrays	[96]
Up: miR-34a, miR-122	cirrhosis and advanced fibrosis	cirrhosis and advanced fibrosis versus early fibrosis stage	serum	qRT-PCR of four miRNAs	[112]
Down: miR-122; *up*: miR-21	advanced fibrosis	advanced versus early fibrosis stage	liver	qRT-PCR of miR-122 and miR-21	[107]

qRT-PCR, quantitative real-time polymerase chain reaction.

hepatocellular carcinoma [94]. Several studies have analyzed miRNA expression patterns in relation to liver fibrogenesis in chronic HCV infection, supporting the hypothesis that miRNA-mediated regulation is mechanistically involved in this process. Table 10.3 provides a summary of the data from the respective *in vivo* studies in liver and serum from chronic HCV-infected patients. In addition, several *in vitro* studies with hepatic stellate cells (HSCs) have correlated mechanistically miRNA alterations with liver fibrogenesis. These cells are activated in the setting of chronic liver injury mainly through stimulation by TGF-β

and are the principal cell type contributing to extracellular matrix deposition during liver fibrogenesis [95]. Several profibrotic and antifibrotic miRNAs have been described. Regarding the former, overexpression of miR-199 and miR-200 families in HSCs leads to an increased expression of fibrosis related-genes associated with the TGF-β signaling pathway [96]. Overexpression of miR-15b and miR-16 inhibits HSC apoptosis by targeting Bcl-2 and the caspase signaling pathway [97]. Another profibrotic miRNA is the miR-571, whose expression is upregulated in activated HSCs and likely targets the gene transcription modulator CREB-binding protein [98]. Regarding antifibrotic miRNAs, overexpression of miR-122 in HSC leads to decreased extracellular matrix production by suppressing P4HA1 expression – a key enzyme in collagen maturation [99]. Overexpression of miR-27a and miR-27b causes the reversion of activated HSCs to a quiescent phenotype by regulating the retinoid X receptor [100], and both miR-29 [101] and miR-133 [102] cause a reduction in collagen expression. Furthermore, overexpression of miR-150 and miR-194 targets c-*MYB* and *RAC1* genes [103], miR-146a negatively regulates TGF-β signaling by decreasing the expression of SMAD4 [104], and miR-19b negatively regulates TGF-β signaling by decreasing TGF-β receptor II, SMAD3, and procollagen expression [105], in all cases leading to inhibition of HSC activation. A mechanistic link between miRNA expression and liver fibrogenesis is further substantiated by studies of liver fibrosis in mice. Fibrogenesis has been associated with downregulation of miR-122 [99] and miR-19b [105], downregulation of the miR-29 family [106] and miR-146a [104] through the TGF-β signaling pathway, upregulation of miR-21 that targets a negative regulator of TGF-β signaling pathway called the *SMAD7* gene [107], and downregulation of miR-132 that leads to the activation of an epigenetic pathway involved in HSC activation [108].

10.3.2.3 Prediction of Treatment Outcome in Chronic HCV-1 Infected Patients

Prediction of Treatment Outcome Using Classical Predictors Currently, there are no reliable baseline predictors of the SVR at an individual level. However, several baseline viral and host factors help to predict the SVR to PEG-IFN/RBV [113,114]. Of special relevance are several single nucleotide polymorphisms (SNPs) located near the human *IL28B* gene, which codes for IFN-λ3 and has been associated with treatment outcome. Specifically, the CC genotype for the SNP rs12979860 was the strongest baseline predictor of SVR when compared with other host and viral predictors [115,116]. For example, in the Caucasian population, the likelihood of a SVR in CC genotype carriers is about 80% (i.e., twofold higher than in carriers of CT and TT genotypes) [115]. However, only 30% of Caucasian patients carry the CC genotype. Moreover, up to 40% of the patients carrying the disadvantageous T allele also achieve a SVR to PEG-IFN/RBV and this frequency even increases in triple therapy [117,118]. Thus, the *IL28B* genotype alone is not a sufficient reliable baseline predictor for personalizing antiviral treatment regimens. To reduce costs and side-effects associated with triple therapy, new predictors of therapy outcome are needed.

Prediction of Treatment Outcome Using miRNA Expression Patterns Several recent observations suggest that miRNA-based assays might provide new prognostic markers for predicting HCV therapy success and help build personalized anti-HCV treatment schemes. (i) miRNAs can directly influence positively or negatively the HCV life cycle by binding to the viral genome. For example, the liver-specific miR-122 facilitates HCV RNA replication [53] and translation [54], whereas miR-199a inhibits HCV RNA replication [119], and miR-196 [120] as well as miR-let-7b [121] suppress HCV RNA expression. (ii) miRNAs can also indirectly influence the HCV life cycle by modulating required host factors. Positive regulators are miR-491 [122], miR-141 [123], and miR-130A [124], and a negative regulator is the miR-196 [125]. (iii) Type I IFN signaling leads to modifications of miRNA expression levels within liver cells, some of which, such as miR-196, miR-296, miR-351, miR-431, and miR-448, have anti-HCV activity [43]. (iv) miRNAs play an important role in regulating immune system development and function [126], and treatment responses have been significantly correlated with the baseline expression level of type I IFN-stimulated genes in the liver [127,128]. As IFN-treatment affects antigen presentation as well as effector lymphocyte development and activity, and all these help to combat HCV *in vivo*, it is likely that these beneficial effects are mirrored in miRNA expression patterns.

The first attempts to correlate baseline expression levels of miRNAs with treatment outcome in chronic HCV-1-infected patients are encouraging. Sarasin-Filipowicz *et al.* showed that non-responders to treatment have a low baseline expression level of hepatic miR-122 [129]. However, liver biopsy is an invasive procedure that is not performed with all patients. Therefore, the identification of blood markers will be of great interest. The use of the miR-122 detected in serum to predict treatment outcome has been assessed; however, contradictory results have been obtained. While Waidmann *et al.* did not associate the non-response with the expression level of miR-122 in serum [130], a recent study that included a larger population size suggested that serum miR-122 could be used as a surrogate of hepatic miR-122 to predict treatment outcome [131]. As genome-wide analyses of miRNA expression patterns could give more accurate predictions in comparison with single miRNA expression, it would be interesting to use "omics" techniques for this approach. In this context, the miRNA expression pattern in liver biopsies derived from over 400 miRNAs using microarrays was significantly associated with PEG-IFN/RBV treatment outcome [132]. Nevertheless, at the moment, no study has performed an exhaustive analysis of miRNA expression patterns in blood in order to predict the treatment outcome in HCV-1-infected patients.

10.3.2.4 Future Perspectives

"Omics" approaches have completely changed our way of studying chronic HCV infections and fascinating progress has been made. However, it is of note that no study has analyzed miRNA expression patterns using NGS and, thus, novel miRNAs mechanistically involved in liver fibrogenesis or in therapy response might

have not been identified. Moreover, no study has made an exhaustive analysis of miRNA expression patterns from peripheral blood cells. As (i) circulating peripheral blood cells including monocytes and lymphocytes have been previously involved in liver fibrogenesis [133–135], and (ii) treatment response has been significantly correlated with the proportion and activation state of regulatory T cells [136,137] and T helper cells [138–140], there is a need for a more exhaustive analysis of the miRNAs expressed in these cells *in vivo* and a further characterization of the molecular pathways involved.

10.4 miRNA-Targeting Therapeutics

miRNAs are potential therapeutic targets in chronic HCV infection. For example, a miR-122 inhibitor decreased HCV load in chimpanzees [141] and in humans [55] with no evidence of severe side-effects or viral resistance development. Depending on the role exerted by the targeted miRNA in a given infection, it will be necessary to inhibit or overexpress the respective miRNA. However, the use of miRNA therapeutics should be a cautious approach especially because a single miRNA can regulate hundreds of mRNAs and consequently multiple cellular pathways, leading to the potential induction of unwanted side-effects. For example, it is known that miR-122 also acts as a suppressor of hepatocarcinogenesis [142] and, thus, it has been suggested that the administration of miR-122 inhibitory therapy against HCV infection should be strictly monitored [143]. Another challenge is which strategy to use in order to specifically deliver the miRNA inhibitors into the diseased tissue without losing their stability. To date, the strategy used is the administration of a single-stranded oligonucleotide complementary to the target miRNA (antagomiR) or a double-stranded oligoribonucleotide corresponding to the miRNA to be overexpressed (mimic). In the case of the antagomiR against miR-122, researchers used a chemically modified antisense oligonucleotide that was able to enter into hepatic cells and was resistant to nucleases [141]. More research on miRNA regulatory pathways associated with different infections and diseases is needed in order to exploit miRNA-targeting therapeutic options.

10.5 Concluding Remarks

In recent years, knowledge in the field of miRNAs and infectious diseases has increased enormously. Deregulated miRNA expression patterns have been associated with some infectious diseases and outcomes. Furthermore, the high stability of miRNAs in serum opens the attractive possibility to use host miRNAs as potential disease biomarkers in easily accessible samples. However, the molecular mechanisms of the roles played by these miRNAs at different levels of

pathogenesis are still poorly understood. Addressing this fundamental issue will be essential in order to design novel specific therapeutic strategies that modulate such miRNAs. In addition, further work is required for successful delivery of antagomiRs or miRNA mimics.

Acknowledgments

This work was supported by grants from the Spanish Ministerio de Ciencia e Innovación BFU2010-20803 and SAF2010-21336. A.M. is a researcher funded from the Institució Catalana de Recerca i Estudis Avançats (ICREA). I.L. is the recipient of a Joint ERS/SEPAR Long-Term Fellowship (LTRF 82-2012).

References

1 O'Connell, R.M., Rao, D.S., Chaudhuri, A.A., and Baltimore, D. (2010) Physiological and pathological roles for microRNAs in the immune system. *Nature Reviews Immunology*, **10** (2), 111–122.

2 Contreras, J. and Rao, D.S. (2011) MicroRNAs in inflammation and immune responses. *Leukemia*, **26** (3), 404–413.

3 Cortez, M.A., Bueso-Ramos, C., Ferdin, J., Lopez-Berestein, G., Sood, A.K., and Calin, G.A. (2011) MicroRNAs in body fluids – the mix of hormones and biomarkers. *Nature Reviews Clinical Oncology*, **8** (8), 467–477.

4 Friedman, R.C., Farh, K.K., Burge, C.B., and Bartel, D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*, **19** (1), 92–105.

5 Varnholt, H., Drebber, U., Schulze, F., Wedemeyer, I., Schirmacher, P., Dienes, H.P., and Odenthal, M. (2008) MicroRNA gene expression profile of hepatitis C virus-associated hepatocellular carcinoma. *Hepatology*, **47** (4), 1223–1232.

6 Bala, S., Marcos, M., and Szabo, G. (2009) Emerging role of microRNAs in liver diseases. *World Journal of Gastroenterology*, **15** (45), 5633–5640.

7 Jiang, X., Tsitsiou, E., Herrick, S.E., and Lindsay, M.A. (2010) MicroRNAs and the regulation of fibrosis. *FEBS Journal*, **277** (9), 2015–2021.

8 Shah, A.A., Leidinger, P., Blin, N., and Meese, E. (2010) miRNA: small molecules as potential novel biomarkers in cancer. *Current Medicinal Chemistry*, **17** (36), 4427–4432.

9 Valadi, H., Ekström, K., Bossios, A., Sjöstrand, M., Lee, J.J., and Lötvall, J.O. (2007) Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nature Cell Biology*, **9** (6), 654–659.

10 Harris, A., Krams, S.M., and Martinez, O.M. (2010) MicroRNAs as immune regulators: implications for transplantation. *American Journal of Transplantation*, **10** (4), 713–719.

11 O'Connell, R.M., Rao, D.S., and Baltimore, D. (2012) microRNA regulation of inflammatory responses. *Annual Review of Immunology*, **30**, 295–312.

12 Taganov, K.D., Boldin, M.P., Chang, K.J., and Baltimore, D. (2006) NF-kappaB-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses. *Proceedings of the National Academy of Sciences of the United States of America*, **103** (33), 12481–12486.

13 Muljo, S.A., Ansel, K.M., Kanellopoulou, C., Livingston, D.M., Rao, A., and Rajewsky, K. (2005) Aberrant T cell

differentiation in the absence of Dicer. *Journal of Experimental Medicine*, **202** (2), 261–269.

14 Koralov, S.B., Muljo, S.A., Galler, G.R., Krek, A., Chakraborty, T., Kanellopoulou, C., Jensen, K., Cobb, B.S., Merkenschlager, M., Rajewsky, N., and Rajewsky, K. (2008) Dicer ablation affects antibody diversity and cell survival in the B lymphocyte lineage. *Cell*, **132** (5), 860–874.

15 Zabaleta, J. (2012) MicroRNA: a bridge from *H. pylori* infection to gastritis and gastric cancer development. *Frontiers in Genetics*, **3**, 294.

16 Xiao, B., Liu, Z., Li, B.S., Tang, B., Li, W., Guo, G., Shi, Y., Wang, F., Wu, Y., Tong, W.D., Guo, H., Mao, X.H., and Zou, Q.M. (2009) Induction of microRNA-155 during *Helicobacter pylori* infection and its negative regulatory role in the inflammatory response. *Journal of Infectious Diseases*, **200** (6), 916–925.

17 Fassi Fehri, L., Koch, M., Belogolova, E., Khalil, H., Bolz, C., Kalali, B., Mollenkopf, H.J., Beigier-Bompadre, M., Karlas, A., Schneider, T., Churin, Y., Gerhard, M., and Meyer, T.F. (2010) *Helicobacter pylori* induces miR-155 in T cells in a cAMP-Foxp3-dependent manner. *PLoS One*, **5** (3), e9500.

18 Oertli, M., Engler, D.B., Kohler, E., Koch, M., Meyer, T.F., and Muller, A. (2011) MicroRNA-155 is essential for the T cell-mediated control of *Helicobacter pylori* infection and for the induction of chronic gastritis and colitis. *Journal of Immunology*, **187** (7), 3578–3586.

19 Lario, S., Ramirez-Lazaro, M.J., Aransay, A.M., Lozano, J.J., Montserrat, A., Casalots, A., Junquera, F., Alvarez, J., Segura, F., Campo, R., and Calvet, X. (2012) microRNA profiling in duodenal ulcer disease caused by *Helicobacter pylori* infection in a Western population. *Clinical Microbiology and Infection*, **18** (8), E273–282.

20 Eulalio, A., Schulte, L., and Vogel, J. (2012) The mammalian microRNA response to bacterial infections. *RNA Biology*, **9** (6), 742–750.

21 Schulte, L.N., Eulalio, A., Mollenkopf, H.J., Reinhardt, R., and Vogel, J. (2011) Analysis of the host microRNA response to *Salmonella* uncovers the control of major cytokines by the let-7 family. *EMBO Journal*, **30** (10), 1977–1989.

22 Sharbati, S., Sharbati, J., Hoeke, L., Bohmer, M., and Einspanier, R. (2012) Quantification and accurate normalisation of small RNAs through new custom RT-qPCR arrays demonstrates *Salmonella*-induced microRNAs in human monocytes. *BMC Genomics*, **13**, 23.

23 Schnitger, A.K., Machova, A., Mueller, R.U., Androulidaki, A., Schermer, B., Pasparakis, M., Kronke, M., and Papadopoulou, N. (2011) *Listeria monocytogenes* infection in macrophages induces vacuolar-dependent host miRNA response. *PLoS One*, **6** (11), e27435.

24 Izar, B., Mannala, G.K., Mraheil, M.A., Chakraborty, T., and Hain, T. (2012) microRNA response to *Listeria monocytogenes* infection in epithelial cells. *International Journal of Molecular Sciences*, **13** (1), 1173–1185.

25 Lind, E.F., Elford, A.R., and Ohashi, P.S. (2013) Micro-RNA 155 is required for optimal $CD8^+$ T cell responses to acute viral and intracellular bacterial challenges. *Journal of Immunology*, **190** (3), 1210–1216.

26 Ma, F., Xu, S., Liu, X., Zhang, Q., Xu, X., Liu, M., Hua, M., Li, N., Yao, H., and Cao, X. (2011) The microRNA miR-29 controls innate and adaptive immune responses to intracellular bacterial infection by targeting interferon-gamma. *Nature Immunology*, **12** (9), 861–869.

27 Schmidt, W.M., Spiel, A.O., Jilma, B., Wolzt, M., and Muller, M. (2009) *In vivo* profile of the human leukocyte microRNA response to endotoxemia. *Biochemical and Biophysical Research Communications*, **380** (3), 437–441.

28 Stern-Ginossar, N., Elefant, N., Zimmermann, A., Wolf, D.G., Saleh, N., Biton, M., Horwitz, E., Prokocimer, Z., Prichard, M., Hahn, G., Goldman-Wohl, D., Greenfield, C., Yagel, S., Hengel, H., Altuvia, Y., Margalit, H., and Mandelboim, O. (2007) Host immune system gene targeting by a viral miRNA. *Science*, **317** (5836), 376–381.

29 Kim, S., Lee, S., Shin, J., Kim, Y., Evnouchidou, I., Kim, D., Kim, Y.K., Kim, Y.E., Ahn, J.H., Riddell, S.R., Stratikos, E., Kim, V.N., and Ahn, K. (2011) Human cytomegalovirus microRNA miR-US4-1 inhibits $CD8^+$ T cell responses by targeting the aminopeptidase ERAP1. *Nature Immunology*, **12** (10), 984–991.

30 Kim, Y., Lee, S., Kim, S., Kim, D., Ahn, J.H., and Ahn, K. (2012) Human cytomegalovirus clinical strain-specific microRNA miR-UL148D targets the human chemokine RANTES during infection. *PLoS Pathogens*, **8** (3), e1002577.

31 Nachmani, D., Stern-Ginossar, N., Sarid, R., and Mandelboim, O. (2009) Diverse herpesvirus microRNAs target the stress-induced immune ligand MICB to escape recognition by natural killer cells. *Cell Host & Microbe*, **5** (4), 376–385.

32 Abend, J.R., Uldrick, T., and Ziegelbauer, J.M. (2010) Regulation of tumor necrosis factor-like weak inducer of apoptosis receptor protein (TWEAKR) expression by Kaposi's sarcoma-associated herpesvirus microRNA prevents TWEAK-induced apoptosis and inflammatory cytokine expression. *Journal of Virology*, **84** (23), 12139–12151.

33 Abend, J.R., Ramalingam, D., Kieffer-Kwon, P., Uldrick, T.S., Yarchoan, R., and Ziegelbauer, J.M. (2012) Kaposi's sarcoma-associated herpesvirus microRNAs target IRAK1 and MYD88, two components of the toll-like receptor/interleukin-1R signaling cascade, to reduce inflammatory-cytokine expression. *Journal of Virology*, **86** (21), 11663–11674.

34 Lei, X., Zhu, Y., Jones, T., Bai, Z., Huang, Y., and Gao, S.J. (2012) A Kaposi's sarcoma-associated herpesvirus microRNA and its variants target the transforming growth factor beta pathway to promote cell survival. *Journal of Virology*, **86** (21), 11698–11711.

35 Suffert, G., Malterer, G., Hausser, J., Viiliainen, J., Fender, A., Contrant, M., Ivacevic, T., Benes, V., Gros, F., Voinnet, O., Zavolan, M., Ojala, P.M., Haas, J.G., and Pfeffer, S. (2011) Kaposi's sarcoma herpesvirus microRNAs target caspase 3 and regulate apoptosis. *PLoS Pathogens*, **7** (12), e1002405.

36 Gottwein, E. and Cullen, B.R. (2010) A human herpesvirus microRNA inhibits p21 expression and attenuates p21-mediated cell cycle arrest. *Journal of Virology*, **84** (10), 5229–5237.

37 Choy, E.Y., Siu, K.L., Kok, K.H., Lung, R. W., Tsang, C.M., To, K.F., Kwong, D.L., Tsao, S.W., and Jin, D.Y. (2008) An Epstein–Barr virus-encoded microRNA targets PUMA to promote host cell survival. *Journal of Experimental Medicine*, **205** (11), 2551–2560.

38 Nikitin, P.A. and Luftig, M.A. (2011) At a crossroads: human DNA tumor viruses and the host DNA damage response. *Future Virology*, **6** (7), 813–830.

39 Skalsky, R.L., Corcoran, D.L., Gottwein, E., Frank, C.L., Kang, D., Hafner, M., Nusbaum, J.D., Feederle, R., Delecluse, H.J., Luftig, M.A., Tuschl, T., Ohler, U., and Cullen, B.R. (2012) The viral and cellular microRNA targetome in lymphoblastoid cell lines. *PLoS Pathogens*, **8** (1), e1002484.

40 Marquitz, A.R., Mathur, A., Nam, C.S., and Raab-Traub, N. (2011) The Epstein–Barr virus BART microRNAs target the pro-apoptotic protein Bim. *Virology*, **412** (2), 392–400.

41 Riley, K.J., Rabinowitz, G.S., Yario, T.A., Luna, J.M., Darnell, R.B., and Steitz, J.A. (2012) EBV and human microRNAs co-target oncogenic and apoptotic viral and human genes during latency. *EMBO Journal*, **31** (9), 2207–2221.

42 Bauman, Y., Nachmani, D., Vitenshtein, A., Tsukerman, P., Drayman, N., Stern-Ginossar, N., Lankry, D., Gruda, R., and Mandelboim, O. (2011) An identical miRNA of the human JC and BK polyoma viruses targets the stress-induced ligand ULBP3 to escape immune elimination. *Cell Host & Microbe*, **9** (2), 93–102.

43 Pedersen, I.M., Cheng, G., Wieland, S., Volinia, S., Croce, C.M., Chisari, F.V., and David, M. (2007) Interferon modulation of cellular microRNAs as an antiviral mechanism. *Nature*, **449** (7164), 919–922.

44 Wang, P., Hou, J., Lin, L., Wang, C., Liu, X., Li, D., Ma, F., Wang, Z., and Cao, X. (2010) Inducible microRNA-155 feedback promotes type I IFN signaling in antiviral innate immunity by targeting suppressor of cytokine signaling 1. *Journal of Immunology*, **185** (10), 6226–6233.

45 Wiesen, J.L. and Tomasi, T.B. (2009) Dicer is regulated by cellular stresses and interferons. *Molecular Immunology*, **46** (6), 1222–1228.

46 Song, L., Liu, H., Gao, S., Jiang, W., and Huang, W. (2010) Cellular microRNAs inhibit replication of the H1N1 influenza A virus in infected cells. *Journal of Virology*, **84** (17), 8849–8860.

47 Nathans, R., Chu, C.Y., Serquina, A.K., Lu, C.C., Cao, H., and Rana, T.M. (2009) Cellular microRNA and P bodies modulate host–HIV-1 interactions. *Molecular Cell*, **34** (6), 696–709.

48 Lecellier, C.H., Dunoyer, P., Arar, K., Lehmann-Che, J., Eyquem, S., Himber, C., Saib, A., and Voinnet, O. (2005) A cellular microRNA mediates antiviral defense in human cells. *Science*, **308** (5721), 557–560.

49 Hou, J., Wang, P., Lin, L., Liu, X., Ma, F., An, H., Wang, Z., and Cao, X. (2009) MicroRNA-146a feedback inhibits RIG-I-dependent Type I IFN production in macrophages by targeting TRAF6, IRAK1, and IRAK2. *Journal of Immunology*, **183** (3), 2150–2158.

50 Rosato, P., Anastasiadou, E., Garg, N., Lenze, D., Boccellato, F., Vincenti, S., Severa, M., Coccia, E.M., Bigi, R., Cirone, M., Ferretti, E., Campese, A.F., Hummel, M., Frati, L., Presutti, C., Faggioni, A., and Trivedi, P. (2012) Differential regulation of miR-21 and miR-146a by Epstein–Barr virus-encoded EBNA2. *Leukemia*, **26** (11), 2343–2352.

51 Lagos, D., Pollara, G., Henderson, S., Gratrix, F., Fabani, M., Milne, R.S., Gotch, F., and Boshoff, C. (2010) miR-132 regulates antiviral innate immunity through suppression of the p300 transcriptional co-activator. *Nature Cell Biology*, **12** (5), 513–519.

52 Rosenberger, C.M., Podyminogin, R.L., Navarro, G., Zhao, G.W., Askovich, P.S., Weiss, M.J., and Aderem, A. (2012) miR-451 regulates dendritic cell cytokine responses to influenza infection. *Journal of Immunology*, **189** (12), 5965–5975.

53 Jopling, C.L., Yi, M., Lancaster, A.M., Lemon, S.M., and Sarnow, P. (2005) Modulation of hepatitis C virus RNA abundance by a liver-specific microRNA. *Science*, **309** (5740), 1577–1581.

54 Henke, J.I., Goergen, D., Zheng, J., Song, Y., Schuttler, C.G., Fehr, C., Junemann, C., and Niepmann, M. (2008) microRNA-122 stimulates translation of hepatitis C virus RNA. *EMBO Journal*, **27** (24), 3300–3310.

55 Janssen, H.L., Reesink, H.W., Lawitz, E.J., Zeuzem, S., Rodriguez-Torres, M., Patel, K., van der Meer, A.J., Patick, A.K., Chen, A., Zhou, Y., Persson, R., King, B.D., Kauppinen, S., Levin, A.A., and Hodges, M.R. (2013) Treatment of HCV infection by targeting microRNA. *New England Journal of Medicine*, **368** (18), 1685–1694.

56 Huang, J., Wang, F., Argyris, E., Chen, K., Liang, Z., Tian, H., Huang, W., Squires, K., Verlinghieri, G., and Zhang, H. (2007) Cellular microRNAs contribute to HIV-1 latency in resting primary $CD4^+$ T lymphocytes. *Nature Medicine*, **13** (10), 1241–1247.

57 Lu, F., Weidmer, A., Liu, C.G., Volinia, S., Croce, C.M., and Lieberman, P.M. (2008) Epstein–Barr virus-induced miR-155 attenuates NF-kappaB signaling and stabilizes latent virus persistence. *Journal of Virology*, **82** (21), 10436–10443.

58 Ho, B.C., Yu, S.L., Chen, J.J., Chang, S.Y., Yan, B.S., Hong, Q.S., Singh, S., Kao, C.L., Chen, H.Y., Su, K.Y., Li, K.C., Cheng, C. L., Cheng, H.W., Lee, J.Y., Lee, C.N., and Yang, P.C. (2011) Enterovirus-induced miR-141 contributes to shutoff of host protein translation by targeting the translation initiation factor eIF4E. *Cell Host & Microbe*, **9** (1), 58–69.

59 Gupta, A., Gartner, J.J., Sethupathy, P., Hatzigeorgiou, A.G., and Fraser, N.W. (2006) Anti-apoptotic function of a microRNA encoded by the HSV-1 latency-associated transcript. *Nature*, **442** (7098), 82–85.

60 Umbach, J.L., Kramer, M.F., Jurak, I., Karnowski, H.W., Coen, D.M., and Cullen, B.R. (2008) MicroRNAs expressed by herpes simplex virus 1 during latent infection regulate viral mRNAs. *Nature*, **454** (7205), 780–783.

61 Grey, F., Meyers, H., White, E.A., Spector, D.H., and Nelson, J. (2007) A human cytomegalovirus-encoded microRNA regulates expression of multiple viral genes involved in replication. *PLoS Pathog*, **3** (11), e163.

62 Samols, M.A., Skalsky, R.L., Maldonado, A.M., Riva, A., Lopez, M.C., Baker, H.V., and Renne, R. (2007) Identification of cellular genes targeted by KSHV-encoded microRNAs. *PLoS Pathog*, **3** (5), e65.

63 Mack, U., Migliori, G.B., Sester, M., Rieder, H.L., Ehlers, S., Goletti, D., Bossink, A., Magdorf, K., Holscher, C., Kampmann, B., Arend, S.M., Detjen, A., Bothamley, G., Zellweger, J.P., Milburn, H., Diel, R., Ravn, P., Cobelens, F., Cardona, P.J., Kan, B., Solovic, I., Duarte, R., and Cirillo, D.M. (2009) LTBI: latent tuberculosis infection or lasting immune responses to *M. tuberculosis*? A TBNET consensus statement. *European Respiratory Journal*, **33** (5), 956–973.

64 World Health Organization (2012) *Global Tuberculosis Report (WHO/HTM/TB/2012.6)*, WHO, Geneva.

65 Walzl, G., Ronacher, K., Hanekom, W., Scriba, T.J., and Zumla, A. (2011) Immunological biomarkers of tuberculosis. *Nature Reviews Immunology*, **11** (5), 343–354.

66 Urbanczik, R. (1985) Present position of microscopy and of culture in diagnostic mycobacteriology. *Zentralblatt fur Bakteriologie Mikrobiologie und Hygiene A*, **260** (1), 81–87.

67 Steingart, K.R., Ng, V., Henry, M., Hopewell, P.C., Ramsay, A., Cunningham, J., Urbanczik, R., Perkins, M.D., Aziz, M. A., and Pai, M. (2006) Sputum processing methods to improve the sensitivity of smear microscopy for tuberculosis: a systematic review. *Lancet Infectious Diseases*, **6** (10), 664–674.

68 Edwards, L.B., Acquaviva, F.A., Livesay, V.T., Cross, F.W., and Palmer, C.E. (1969) An atlas of sensitivity to tuberculin, PPD-B, and histoplasmin in the United States. *American Review of Respiratory Disease*, **99** (4 Suppl.), 1–132.

69 Berkel, G.M., Cobelens, F.G., de Vries, G., Draayer-Jansen, I.W., and Borgdorff, M. W. (2005) Tuberculin skin test: estimation of positive and negative predictive values from routine data. *International Journal of Tuberculosis and Lung Disease*, **9** (3), 310–316.

70 Ruiz-Manzano, J., Blanquer, R., Calpe, J. L., Caminero, J.A., Cayla, J., Dominguez, J.A., Garcia, J.M., and Vidal, R. (2008) Diagnosis and treatment of tuberculosis. *Archivos de Bronconeumologia*, **44** (10), 551–566.

71 Boehme, C.C., Nabeta, P., Hillemann, D., Nicol, M.P., Shenai, S., Krapp, F., Allen, J., Tahirli, R., Blakemore, R., Rustomjee, R., Milovic, A., Jones, M., O'Brien, S.M., Persing, D.H., Ruesch-Gerdes, S., Gotuzzo, E., Rodrigues, C., Alland, D., and Perkins, M.D. (2010) Rapid molecular detection of tuberculosis and rifampin resistance. *New England Journal of Medicine*, **363** (11), 1005–1015.

72 Dheda, K., Ruhwald, M., Theron, G., Peter, J., and Yam, W.C. (2013) Point-of-care diagnosis of tuberculosis: past, present and future. *Respirology*, **18** (2), 217–232.

73 Wang, C., Yang, S., Sun, G., Tang, X., Lu, S., Neyrolles, O., and Gao, Q. (2011) Comparative miRNA expression profiles in individuals with latent and active tuberculosis. *PLoS One*, **6** (10), e25832.

74 Fu, Y., Yi, Z., Wu, X., Li, J., and Xu, F. (2011) Circulating microRNAs in patients with active pulmonary tuberculosis. *Journal of Clinical Microbiology*, **49** (12), 4246–4251.

75 Liu, Y., Wang, X., Jiang, J., Cao, Z., Yang, B., and Cheng, X. (2011) Modulation of T cell cytokine production by miR-144* with elevated expression in patients with pulmonary tuberculosis. *Molecular Immunology*, **48** (9–10), 1084–1090.

76 Wu, J., Lu, C., Diao, N., Zhang, S., Wang, S., Wang, F., Gao, Y., Chen, J., Shao, L., Lu, J., Zhang, X., Weng, X., Wang, H., Zhang, W., and Huang, Y. (2011) Analysis of microRNA expression profiling identifies miR-155 and

miR-155* as potential diagnostic markers for active tuberculosis: a preliminary study. *Human Immunology*, **73** (1), 31–37.

77 Rajaram, M.V., Ni, B., Morris, J.D., Brooks, M.N., Carlson, T.K., Bakthavachalu, B., Schoenberg, D.R., Torrelles, J.B., and Schlesinger, L.S. (2011) Mycobacterium tuberculosis lipomannan blocks TNF biosynthesis by regulating macrophage MAPK-activated protein kinase 2 (MK2) and microRNA miR-125b. *Proceedings of the National Academy of Sciences of the United States of America*, **108** (42), 17408–17413.

78 Maertzdorf, J., Weiner, J. 3rd, Mollenkopf, H.J., Bauer, T., Prasse, A., Muller-Quernheim, J., and Kaufmann, S.H. (2012) Common patterns and disease-related signatures in tuberculosis and sarcoidosis. *Proceedings of the National Academy of Sciences of the United States of America*, **109** (20), 7853–7858.

79 Yi, Z., Fu, Y., Ji, R., Li, R., and Guan, Z. (2012) Altered microRNA signatures in sputum of patients with active pulmonary tuberculosis. *PLoS One*, **7** (8), e43184.

80 Kumar, R., Halder, P., Sahu, S.K., Kumar, M., Kumari, M., Jana, K., Ghosh, Z., Sharma, P., Kundu, M., and Basu, J. (2012) Identification of a novel role of ESAT-6-dependent miR-155 induction during infection of macrophages with Mycobacterium tuberculosis. *Cellular Microbiology*, **14** (10), 1620–1631.

81 Kleinsteuber, K., Heesch, K., Schattling, S., Kohns, M., Sander-Julch, C., Walzl, G., Hesseling, A., Mayatepek, E., Fleischer, B., Marx, F.M., and Jacobsen, M. (2013) Decreased expression of miR-21, miR-26a, miR-29a, and miR-142-3p in $CD4^+$ T cells and peripheral blood from tuberculosis patients. *PLoS One*, **8** (4), e61609.

82 Singh, Y., Kaul, V., Mehra, A., Chatterjee, S., Tousif, S., Dwivedi, V.P., Suar, M., Van Kaer, L., Bishai, W.R., and Das, G. (2013) *Mycobacterium tuberculosis* controls microRNA-99b (miR-99b) expression in infected murine dendritic cells to modulate host immunity. *Journal of Biological Chemistry*, **288** (7), 5056–5061.

83 Abd-El-Fattah, A.A., Sadik, N.A., Shaker, O.G., and Aboulftouh, M.L. (2013) Differential microRNAs expression in serum of patients with lung cancer, pulmonary tuberculosis, and pneumonia. *Cell Biochemistry and Biophysics*, **67** (3), 875–884.

84 Bloom, C.I., Graham, C.M., Berry, M.P., Wilkinson, K.A., Oni, T., Rozakeas, F., Xu, Z., Rossello-Urgell, J., Chaussabel, D., Banchereau, J., Pascual, V., Lipman, M., Wilkinson, R.J., and O'Garra, A. (2012) Detectable changes in the blood transcriptome are present after two weeks of antituberculosis therapy. *PLoS One*, **7** (10), e46191.

85 Mohd Hanafiah, K., Groeger, J., Flaxman, A.D., and Wiersma, S.T. (2013) Global epidemiology of hepatitis C virus infection: new estimates of age-specific antibody to HCV seroprevalence. *Hepatology*, **57** (4), 1333–1342.

86 Craxi, A. (2011) EASL clinical practice guidelines: management of hepatitis C virus infection. *Journal of Hepatology*, **55** (2), 245–264.

87 World Health Organization (2011) *Hepatitis C (WHO Fact Sheet 164)*, WHO, Geneva.

88 Mormone, E., George, J., and Nieto, N. (2011) Molecular pathogenesis of hepatic fibrosis and current therapeutic approaches. *Chemico-Biological Interactions*, **193** (3), 225–231.

89 Esteban, J.I., Sauleda, S., and Quer, J. (2008) The changing epidemiology of hepatitis C virus infection in Europe. *Journal of Hepatology*, **48** (1), 148–162.

90 Jacobson, I.M., McHutchison, J.G., Dusheiko, G., Di Bisceglie, A.M., Reddy, K.R., Bzowej, N.H., Marcellin, P., Muir, A.J., Ferenci, P., Flisiak, R., George, J., Rizzetto, M., Shouval, D., Sola, R., Terg, R.A., Yoshida, E.M., Adda, N., Bengtsson, L., Sankoh, A.J., Kieffer, T.L., George, S., Kauffman, R.S., and Zeuzem, S. (2011) Telaprevir for previously untreated chronic hepatitis C virus infection. *New England Journal of Medicine*, **364** (25), 2405–2416.

91 Poordad, F., McCone, J.Jr., Bacon, B.R., Bruno, S., Manns, M.P., Sulkowski, M.S., Jacobson, I.M., Reddy, K.R., Goodman, Z. D., Boparai, N., DiNubile, M.J., Sniukiene, V., Brass, C.A., Albrecht, J.K., and

Bronowicki, J.P. (2011) Boceprevir for untreated chronic HCV genotype 1 infection. *New England Journal of Medicine*, **364** (13), 1195–1206.

92 Sherman, K.E., Flamm, S.L., Afdhal, N.H., Nelson, D.R., Sulkowski, M.S., Everson, G.T., Fried, M.W., Adler, M., Reesink, H. W., Martin, M., Sankoh, A.J., Adda, N., Kauffman, R.S., George, S., Wright, C.I., and Poordad, F. (2011) Response-guided telaprevir combination treatment for hepatitis C virus infection. *New England Journal of Medicine*, **365** (11), 1014–1024.

93 Ghany, M.G., Nelson, D.R., Strader, D.B., Thomas, D.L., and Seeff, L.B. (2011) An update on treatment of genotype 1 chronic hepatitis C virus infection: 2011 practice guideline by the American Association for the Study of Liver Diseases. *Hepatology*, **54** (4), 1433–1444.

94 Wang, X.W., Heegaard, N.H., and Orum, H. (2012) MicroRNAs in liver disease. *Gastroenterology*, **142** (7), 1431–1443.

95 Friedman, S.L. (2008) Hepatic stellate cells: protean, multifunctional, and enigmatic cells of the liver. *Physiological Reviews*, **88** (1), 125–172.

96 Murakami, Y., Toyoda, H., Tanaka, M., Kuroda, M., Harada, Y., Matsuda, F., Tajima, A., Kosaka, N., Ochiya, T., and Shimotohno, K. (2011) The progression of liver fibrosis is related with overexpression of the miR-199 and 200 families. *PLoS One*, **6** (1), e16081.

97 Guo, C.J., Pan, Q., Li, D.G., Sun, H., and Liu, B.W. (2009) miR-15b and miR-16 are implicated in activation of the rat hepatic stellate cell: an essential role for apoptosis. *Journal of Hepatology*, **50** (4), 766–778.

98 Roderburg, C., Mollnow, T., Bongaerts, B., Elfimova, N., Vargas Cardenas, D., Berger, K., Zimmermann, H., Koch, A., Vucur, M., Luedde, M., Hellerbrand, C., Odenthal, M., Trautwein, C., Tacke, F., and Luedde, T. (2012) Micro-RNA profiling in human serum reveals compartment-specific roles of miR-571 and miR-652 in liver cirrhosis. *PLoS One*, **7** (3), e32999.

99 Li, J., Ghazwani, M., Zhang, Y., Lu, J., Fan, J., Gandhi, C.R., and Li, S. (2012) miR-122 regulates collagen production via targeting hepatic stellate cells and suppressing P4HA1 expression. *Journal of Hepatology*, **58** (3), 522–528.

100 Ji, J., Zhang, J., Huang, G., Qian, J., Wang, X., and Mei, S. (2009) Over-expressed microRNA-27a and 27b influence fat accumulation and cell proliferation during rat hepatic stellate cell activation. *FEBS Letters*, **583** (4), 759–766.

101 Kwiecinski, M., Noetel, A., Elfimova, N., Trebicka, J., Schievenbusch, S., Strack, I., Molnar, L., von Brandenstein, M., Tox, U., Nischt, R., Coutelle, O., Dienes, H.P., and Odenthal, M. (2011) Hepatocyte growth factor (HGF) inhibits collagen I and IV synthesis in hepatic stellate cells by miRNA-29 induction. *PLoS One*, **6** (9), e24568.

102 Roderburg, C., Luedde, M., Vargas Cardenas, D., Vucur, M., Mollnow, T., Zimmermann, H.W., Koch, A., Hellerbrand, C., Weiskirchen, R., Frey, N., Tacke, F., Trautwein, C., and Luedde, T. (2013) miR-133a mediates TGF-beta-dependent derepression of collagen synthesis in hepatic stellate cells during liver fibrosis. *Journal of Hepatology*, **58** (4), 736–742.

103 Venugopal, S.K., Jiang, J., Kim, T.H., Li, Y., Wang, S.S., Torok, N.J., Wu, J., and Zern, M.A. (2010) Liver fibrosis causes downregulation of miRNA-150 and miRNA-194 in hepatic stellate cells, and their overexpression causes decreased stellate cell activation. *American Journal of Physiology*, **298** (1), G101–106.

104 He, Y., Huang, C., Sun, X., Long, X.R., Lv, X.W., and Li, J. (2012) MicroRNA-146a modulates TGF-beta1-induced hepatic stellate cell proliferation by targeting SMAD4. *Cellular Signalling*, **24** (10), 1923–1930.

105 Lakner, A.M., Steuerwald, N.M., Walling, T.L., Ghosh, S., Li, T., McKillop, I.H., Russo, M.W., Bonkovsky, H.L., and Schrum, L.W. (2012) Inhibitory effects of microRNA 19b in hepatic stellate cell-mediated fibrogenesis. *Hepatology*, **56** (1), 300–310.

106 Roderburg, C., Urban, G.W., Bettermann, K., Vucur, M., Zimmermann, H., Schmidt, S., Janssen, J., Koppe, C., Knolle,

P., Castoldi, M., Tacke, F., Trautwein, C., and Luedde, T. (2011) Micro-RNA profiling reveals a role for miR-29 in human and murine liver fibrosis. *Hepatology*, **53** (1), 209–218.

107 Marquez, R.T., Bandyopadhyay, S., Wendlandt, E.B., Keck, K., Hoffer, B.A., Icardi, M.S., Christensen, R.N., Schmidt, W.N., and McCaffrey, A.P. (2010) Correlation between microRNA expression levels and clinical parameters associated with chronic hepatitis C viral infection in humans. *Laboratory Investigation*, **90** (12), 1727–1736.

108 Mann, J., Chu, D.C., Maxwell, A., Oakley, F., Zhu, N.L., Tsukamoto, H., and Mann, D.A. (2010) MeCP2 controls an epigenetic pathway that promotes myofibroblast transdifferentiation and fibrosis. *Gastroenterology*, **138** (2), 705–714.

109 Shrivastava, S., Petrone, J., Steele, R., Lauer, G.M., Bisceglie, A.M., and Ray, R.B. (2013) Upregulation of circulating miR-20a is correlated with hepatitis C virus mediated liver disease progression. *Hepatology*, **58** (3), 863–871.

110 Ogawa, T., Enomoto, M., Fujii, H., Sekiya, Y., Yoshizato, K., Ikeda, K., and Kawada, N. (2012) MicroRNA-221/222 upregulation indicates the activation of stellate cells and the progression of liver fibrosis. *Gut*, **61** (11), 1600–1609.

111 Murakami, Y., Toyoda, H., Tanahashi, T., Tanaka, J., Kumada, T., Yoshioka, Y., Kosaka, N., Ochiya, T., and Taguchi, Y.H. (2012) Comprehensive miRNA expression analysis in peripheral blood can diagnose liver disease. *PLoS One*, **7** (10), e48366.

112 Cermelli, S., Ruggieri, A., Marrero, J.A., Ioannou, G.N., and Beretta, L. (2011) Circulating microRNAs in patients with chronic hepatitis C and non-alcoholic fatty liver disease. *PLoS One*, **6** (8), e23937.

113 Kau, A., Vermehren, J., and Sarrazin, C. (2008) Treatment predictors of a sustained virologic response in hepatitis B and C. *Journal of Hepatology*, **49** (4), 634–651.

114 Akuta, N., Suzuki, F., Sezaki, H., Suzuki, Y., Hosaka, T., Someya, T., Kobayashi, M., Saitoh, S., Watahiki, S., Sato, J., Matsuda, M., Arase, Y., Ikeda, K., and Kumada, H. (2005) Association of amino acid substitution pattern in core protein of hepatitis C virus genotype 1b high viral load and non-virological response to interferon–ribavirin combination therapy. *Intervirology*, **48** (6), 372–380.

115 Ge, D., Fellay, J., Thompson, A.J., Simon, J.S., Shianna, K.V., Urban, T.J., Heinzen, E.L., Qiu, P., Bertelsen, A.H., Muir, A.J., Sulkowski, M., McHutchison, J.G., and Goldstein, D.B. (2009) Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature*, **461** (7262), 399–401.

116 Thomas, D.L., Thio, C.L., Martin, M.P., Qi, Y., Ge, D., O'Huigin, C., Kidd, J., Kidd, K., Khakoo, S.I., Alexander, G., Goedert, J.J., Kirk, G.D., Donfield, S.M., Rosen, H.R., Tobler, L.H., Busch, M.P., McHutchison, J.G., Goldstein, D.B., and Carrington, M. (2009) Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. *Nature*, **461** (7265), 798–801.

117 Poordad, F., Bronowicki, J., Gordon, S., Zeuzem, S., Jacobson, I., Sulkowski, M., Poynard, T., Morgan, T., Burroughs, M., Sniukiene, V., Boparai, N., and Brass, C. (2011) IL28B polymorphism predicts virologic response in patients with hepatitis C genotype 1 treated with boceprevir (BOC) combination therapy. *Journal of Hepatology*, **54**, S1–S24.

118 Jacobson, I., Catlett, I., Marcellin, P., Bzowej, N., Muir, A., Adda, N., Bengtsson, L., George, S., Seepersaud, S., Ramachandran, R., Sussky, K., Kauffman, R., and Botfield, M. (2011) Telaprevir substantially improved SVR rates across all IL28B genotypes in the Advance trial. *Journal of Hepatology*, **54**, S535–S546.

119 Murakami, Y., Aly, H.H., Tajima, A., Inoue, I., and Shimotohno, K. (2009) Regulation of the hepatitis C virus genome replication by miR-199a. *Journal of Hepatology*, **50** (3), 453–460.

120 Hou, W., Tian, Q., Zheng, J., and Bonkovsky, H.L. (2010) MicroRNA-196 represses Bach1 protein and hepatitis C virus gene expression in human hepatoma cells expressing hepatitis C

viral proteins. *Hepatology*, **51** (5), 1494–1504.

121 Cheng, J.C., Yeh, Y.J., Tseng, C.P., Hsu, S.D., Chang, Y.L., Sakamoto, N., and Huang, H.D. (2012) Let-7b is a novel regulator of hepatitis C virus replication. *Cellular and Molecular Life Sciences*, **69** (15), 2621–2633.

122 Ishida, H., Tatsumi, T., Hosui, A., Nawa, T., Kodama, T., Shimizu, S., Hikita, H., Hiramatsu, N., Kanto, T., Hayashi, N., and Takehara, T. (2011) Alterations in microRNA expression profile in HCV-infected hepatoma cells: involvement of miR-491 in regulation of HCV replication via the PI3 kinase/Akt pathway. *Biochemical and Biophysical Research Communications*, **412** (1), 92–97.

123 Banaudha, K., Kaliszewski, M., Korolnek, T., Florea, L., Yeung, M.L., Jeang, K.T., and Kumar, A. (2011) MicroRNA silencing of tumor suppressor DLC-1 promotes efficient hepatitis C virus replication in primary human hepatocytes. *Hepatology*, **53** (1), 53–61.

124 Bhanja Chowdhury, J., Shrivastava, S., Steele, R., Di Bisceglie, A.M., Ray, R., and Ray, R.B. (2012) Hepatitis C virus infection modulates expression of interferon stimulatory gene IFITM1 by upregulating miR-130A. *Journal of Virology*, **86** (18), 10221–10225.

125 Bihrer, V., Friedrich-Rust, M., Kronenberger, B., Forestier, N., Haupenthal, J., Shi, Y., Peveling-Oberhag, J., Radeke, H.H., Sarrazin, C., Herrmann, E., Zeuzem, S., Waidmann, O., and Piiper, A. (2011) Serum miR-122 as a biomarker of necroinflammation in patients with chronic hepatitis C virus infection. *American Journal of Gastroenterology*, **106** (9), 1663–1669.

126 Baltimore, D., Boldin, M.P., O'Connell, R. M., Rao, D.S., and Taganov, K.D. (2008) MicroRNAs: new regulators of immune cell development and function. *Nature Immunology*, **9** (8), 839–845.

127 Sarasin-Filipowicz, M., Oakeley, E.J., Duong, F.H., Christen, V., Terracciano, L., Filipowicz, W., and Heim, M.H. (2008) Interferon signaling and treatment outcome in chronic hepatitis C. *Proceedings of the National Academy of Sciences of the United States of America*, **105** (19), 7034–7039.

128 Selzner, N., Chen, L., Borozan, I., Edwards, A., Heathcote, E.J., and McGilvray, I. (2008) Hepatic gene expression and prediction of therapy response in chronic hepatitis C patients. *Journal of Hepatology*, **48** (5), 708–713.

129 Sarasin-Filipowicz, M., Krol, J., Markiewicz, I., Heim, M.H., and Filipowicz, W. (2009) Decreased levels of microRNA miR-122 in individuals with hepatitis C responding poorly to interferon therapy. *Nature Medicine*, **15** (1), 31–33.

130 Waidmann, O., Bihrer, V., Kronenberger, B., Zeuzem, S., Piiper, A., and Forestier, N. (2011) Pretreatment serum microRNA-122 is not predictive for treatment response in chronic hepatitis C virus infection. *Digestive and Liver Disease*, **44** (5), 438–441.

131 Su, T.H., Liu, C.H., Liu, C.J., Chen, C.L., Ting, T.T., Tseng, T.C., Chen, P.J., Kao, J. H., and Chen, D.S. (2013) Serum microRNA-122 level correlates with virologic responses to pegylated interferon therapy in chronic hepatitis C. *Proceedings of the National Academy of Sciences of the United States of America*, **110** (19), 7844–7849.

132 Murakami, Y., Tanaka, M., Toyoda, H., Hayashi, K., Kuroda, M., Tajima, A., and Shimotohno, K. (2010) Hepatic microRNA expression is associated with the response to interferon treatment of chronic hepatitis C. *BMC Medical Genomics*, **3**, 48.

133 Zimmermann, H.W., Seidler, S., Nattermann, J., Gassler, N., Hellerbrand, C., Zernecke, A., Tischendorf, J.J., Luedde, T., Weiskirchen, R., Trautwein, C., and Tacke, F. (2010) Functional contribution of elevated circulating and hepatic non-classical CD14CD16 monocytes to inflammation and human liver fibrosis. *PLoS One*, **5** (6), e11049.

134 Zimmermann, H.W., Seidler, S., Gassler, N., Nattermann, J., Luedde, T., Trautwein, C., and Tacke, F. (2011) Interleukin-8 is activated in patients with chronic liver diseases and associated with hepatic macrophage accumulation in

human liver fibrosis. *PLoS One*, **6** (6), e21381.

135 Doi, H., Iyer, T.K., Carpenter, E., Li, H., Chang, K.M., Vonderheide, R.H., and Kaplan, D.E. (2012) Dysfunctional B-cell activation in cirrhosis resulting from hepatitis C infection associated with disappearance of CD27-positive B-cell population. *Hepatology*, **55** (3), 709–719.

136 Akiyama, M., Ichikawa, T., Miyaaki, H., Motoyoshi, Y., Takeshita, S., Ozawa, E., Miuma, S., Shibata, H., Taura, N., and Nakao, K. (2010) Relationship between regulatory T cells and the combination of pegylated interferon and ribavirin for the treatment of chronic hepatitis type C. *Intervirology*, **53** (3), 154–160.

137 Soldevila, B., Alonso, N., Martinez-Arconada, M.J., Morillas, R.M., Planas, R., Sanmarti, A.M., and Martinez-Caceres, E. M. (2010) A prospective study of T- and B-lymphocyte subpopulations, CD81 expression levels on B cells and regulatory $CD4^+$ $CD25^+$ $CD127^{low/-}$ $FoxP3^+$ T cells in patients with chronic HCV infection during pegylated interferon-alpha2a plus ribavirin treatment. *Journal of Viral Hepatitis*, **18** (6), 384–392.

138 Lee, S., Hammond, T., Watson, M.W., Flexman, J.P., Cheng, W., Fernandez, S., and Price, P. (2010) Could a loss of memory T cells limit responses to hepatitis C virus (HCV) antigens in blood leucocytes from patients chronically infected with HCV before and during pegylated interferon-alpha and ribavirin therapy? *Clinical and Experimental Immunology*, **161** (1), 118–126.

139 Trapero-Marugan, M., Garcia-Buey, L., Munoz, C., Quintana, N.E., Moreno-Monteagudo, J.A., Borque, M.J., Fernandez, M.J., Salvanes, F.R., Medina, J., and Moreno-Otero, R. (2006) Sustained virological response to peginterferon plus ribavirin in chronic hepatitis C genotype 1 patients is associated with a persistent T_h1 immune response. *Alimentary Pharmacology & Therapeutics*, **24** (1), 117–128.

140 Vrolijk, J.M., Kwekkeboom, J., Janssen, H.L., Hansen, B.E., Zondervan, P.E., Osterhaus, A.D., Schalm, S.W., and Haagmans, B.L. (2003) Pretreatment intrahepatic $CD8^+$ cell count correlates with virological response to antiviral therapy in chronic hepatitis C virus infection. *Journal of Infectious Diseases*, **188** (10), 1528–1532.

141 Lanford, R.E., Hildebrandt-Eriksen, E.S., Petri, A., Persson, R., Lindow, M., Munk, M.E., Kauppinen, S., and Orum, H. (2010) Therapeutic silencing of microRNA-122 in primates with chronic hepatitis C virus infection. *Science*, **327** (5962), 198–201.

142 Hsu, S.H., Wang, B., Kota, J., Yu, J., Costinean, S., Kutay, H., Yu, L., Bai, S., La Perle, K., Chivukula, R.R., Mao, H., Wei, M., Clark, K.R., Mendell, J.R., Caligiuri, M.A., Jacob, S.T., Mendell, J.T., and Ghoshal, K. (2012) Essential metabolic, anti-inflammatory, and anti-tumorigenic functions of miR-122 in liver. *Journal of Clinical Investigation*, **122** (8), 2871–2883.

143 Zeisel, M.B., Lupberger, J., Fofana, I., and Baumert, T.F. (2013) Host-targeting agents for prevention and treatment of chronic hepatitis C - perspectives and challenges. *Journal of Hepatology*, **58** (2), 375–384.

11
Towards the Identification of Condition-Specific Microbial Populations from Human Metagenomic Data

Cédric C. Laczny and Paul Wilmes

11.1
Introduction

Mixed microbial communities are ubiquitous and contribute essential functionalities to all ecosystems. One particular ecosystem – the human body – is inhabited by diverse sets of microorganisms in various sites such as the oral cavity or the gastrointestinal tract (GIT, herein we use the term to refer to all the structures of the human digestive tract ranging from the mouth to the anus). Recent in-depth surveys of the microbial diversity of individual body habitats, such as the mouth, nose, or skin, have highlighted different abundances of taxa per body site leading to pronounced intraindividual variability [1]. Furthermore, interindividual differences in terms of either microbiome composition and/or functional potential have also been described using high-throughput metagenomic sequencing data [2–4].

Some of the essential functionalities carried out by commensal or mutualistic microorganisms in humans include carbohydrate metabolism [5], nutrient absorption [6], and epithelial barrier function maintenance [7,8]. The importance of the human microbiome (the collective of microorganisms inhabiting the human body) in relation to human physiology is reflected by the around 10 : 1 ratio of the number of microbial cells to human cells [9–11]. On the genetic level, the functional potential encoded within the microbial metagenome is 100-fold greater than that encoded by the human genome [12].

The human microbiome is characterized by extensive complexity, but the aforementioned symbiotic relationships between the host and microorganisms are typically in equilibrium. Disruption of this balance can lead to a state referred to as microbial dysbiosis in which parasitic pathobionts and pathogens may dominate the communities [13]. Such imbalances may arise from different perturbations, such as increase in the relative numbers of obligate and opportunistic pathogens [14], inappropriate and/or repeated administration of antibiotics [15–17], mental stress [18], or can even have a dietary origin [19]. Dysbiosis has been associated with a range of different human diseases [20]. Examples include human immunodeficiency virus (HIV) transmission and infection in women

Nucleic Acids as Molecular Diagnostics, First Edition. Edited by Andreas Keller and Eckart Meese.

[21], colorectal cancer [22,23], metabolic syndrome [24,25], and type 2 diabetes (T2DM) [26]. A recent review by Pflughoeft and Versalovic highlights that the complex world of endogenous mixed microbial communities needs to be strongly considered in the context of human systems biology [27]. Consequently, within the context of "systems biomedicine," high-resolution tools are required that allow the unbiased and accurate detection of changes in microbial community structure and function, especially in relation to dysbiotic states. This is all the more important as changes in the microbiome towards a dysbiotic state dominated by particular microbial populations may have important prognostic value and may allow for early therapeutic intervention.

11.2 Nucleic Acid-Based Methods in Diagnostic Microbiology

Nucleic acids (and specifically DNA as it encodes the "blueprint" of organisms) represent appropriate analytes for the precise characterization or identification of microorganisms in samples obtained from human subjects. Microbiological culture remains the primary diagnostic method in medical microbiology because it forms the foundation of Koch's postulates. Obtained cultures are nowadays subjected to in-depth characterizations that may involve the use of molecular tools. Genome-based multilocus sequence analysis (MLSA; or multilocus sequence typing (MLST)) [28] and whole-genome sequencing [29] provide increased or absolute resolution, respectively, for taxonomic classification of isolated bacterial cultures [30]. Combined culture and DNA-based diagnostic tests are readily available (e.g., the Verigene Gram-Positive Blood Culture Test) [31]. By design, such tests offer very targeted results with regard to the supported assay panel, which typically only includes a small number of specific microorganisms known to be relevant from an infection point of view because of previous extensive characterization efforts. In this section, we will consider why culture-independent approaches are desirable and discuss various aspects of DNA-based characterization of entire microbial communities for the detection of specific microorganisms relevant to human health and disease (i.e., condition-specific microbial populations), without the need to subject samples to known cultivation bottlenecks.

11.2.1 Limitations of Culture-Dependent Approaches

The emergence of molecular culture-independent techniques (primarily developed for the characterization of microbial communities within environmental samples) over the past quarter century resulted from the fact that many microorganisms are difficult or impossible to culture in isolation under standard laboratory conditions [32–34]. Various reasons for "unculturability" are known (for a recent review, see [35]). Importantly, the application of standard

culture-dependent approaches may lead to the retrieval of non-representative microbial populations also referred to as "laboratory weeds" [36]. Therefore, in the clinical context, culture-dependent techniques often fail to retrieve disease-relevant cultures [37–39], especially as strain-level resolution should typically be achieved for accurate diagnosis [40]. Furthermore, significantly reduced time allotments for culture-independent molecular assays should exist as these can typically be directly applied to samples obtained *in situ* without the need for extensive sample handling, processing, and work-up.

11.2.2
Culture-Independent Characterization of Microbial Communities

A commonly used means for the characterization of microbial community structure is the targeted sequencing of particular marker genes, most commonly the 16S rRNA gene. This gene shows a comparatively low intragenomic heterogeneity [41], which can therefore be harnessed for the indiscriminate amplification of the gene across the members of a community, while the hypervariable regions allow for the identification of individual taxa at different levels of taxonomic resolution (genus, species, etc.) [42,43]. Theoretically, amplification and deep sequencing of marker genes should allow qualitative and quantitative differences in microbial community structure to be resolvable. However, uncertainties in associating specific sequences with taxa are commonplace and phylogenetic overviews do not convey any information about functional differences that may result from changes in community structure. Fluorescence *in situ* hybridization (FISH) combined with epifluorescence microscopy [44] and microarray-based [45,46] assays have also been developed for the culture-independent taxonomic characterization of microbial communities, mostly using specific discriminatory marker gene sequences (primarily targeting the 16S rRNA gene). Despite the possibility of culture-independent application, marker gene-based approaches suffer from the requirement of prior knowledge (i.e., for the definition of particular *a priori* known target sequences, similar to culture-dependent techniques where particular growth conditions must be known).

11.2.3
Metagenomics

As a result of the advent of next-generation high-throughput random shotgun sequencing (RSS) techniques, the characterization of microbial community structure via 16S rRNA gene–targeted sequencing may be supplemented or substituted by RSS of the entire genomic complement of a sample [47]. The application of untargeted high-throughput genomics to mixed microbial communities without prior isolated culturing is generally referred to as "metagenomics" [48]. The metagenomic approach has recently been used for the in-depth description of the healthy human microbiome [49,50] and for the definition of "enterotypes," which allow classification of the human GIT microbiome into

three distinct categories reflecting particular microbial community compositions [51]. Metagenomic sequencing of the entire genomic complements of microbial communities offers a more comprehensive picture of human microbiota compared with marker gene-based approaches. It allows for much more information to be captured when compared with sequencing of 16S rRNA genes alone, especially as sequences of individual loci (e.g., ribosomal genes) can be recovered from the genomic complements [52]. Therefore, metagenomic sequencing of extracted microbial community DNA offers the ability to resolve information on microbial community composition as well as overall functional potential.

11.2.4
Fecal Samples as Proxies to Evaluate Human Microbiome-Related Health Status

Prominent examples of metagenomics-based characterizations of human-associated microbial populations include those carried out on samples derived from the GIT. The GIT plays a central role in the human body, notably for nutrient and energy uptake. In particular, the colon includes a large number of microbial cells (3.2×10^{11} cells/g or cells/ml) [53]. As a result, a high microbial activity is associated with processes along the entire GIT. In-depth characterizations have revealed overall medial diversity in the GIT, but this increases from the stomach to the colon (summarized in [54]) leading to specific biogeographical patterns along the GIT [55]. However, due to the restricted accessibility of the GIT, fecal samples are usually collected and act as proxies for processes occurring along the entire GIT [49,50,56]. While the microbial community structure of fecal samples is at best representative of the lower part of the digestive tract (descending colon), the data derived from such samples have nonetheless provided invaluable insights into microbiome-associated health and disease status [26,57–60].

11.3
Need for Comprehensive Microbiome Characterization in Medical Diagnostics

The rapid and correct identification of disease-associated microbial populations is of key importance for the effective diagnosis and treatment of diseased individuals [61,62]. However, diagnosis is still mostly performed through traditional culturing techniques, themselves developed in the nineteenth century, before the isolates obtained are then subjected to advanced molecular characterization. Well-publicized examples of life-threatening infections for which the disease-causing agent was identified using this approach include methicillin-resistant *Staphylococcus aureus* (MRSA) strains infecting individuals due to prolonged stays in hospitals or healthcare facilities [63] or the unusual serotype of Shiga toxin-producing *Escherichia coli* (O104 : H4) linked to the large outbreak of diarrhea and hemolytic/uremic syndrome in Germany in 2011 [64]. Obtaining isolate cultures (e.g., from patient samples) is labor- and time-consuming (e.g., up

to 3 days for blood culture to become positive and up to 2 days for identification [62]; Figure 11.1a), aside from the above-described possible complications of "unculturability" or non-representative populations due to artificial culture conditions. Targeted molecular approaches, such as FISH [65], generally improve on the time aspect but they focus on specific and small sets of microorganisms [31,62]. Mass spectrometry-based approaches also offer improved turnaround times but require previously characterized mass spectra of isolated cultures for comparison [66]. Metagenomics offers a means to directly obtain details of microbial community compositions and/or functions linked to the respective communities and their members. Hence, culture-independent techniques, in particular untargeted metagenomic sequencing, are of increased interest for the comprehensive characterization of microbiomes in diagnostics (Figure 11.1b). Effectively, while the study by Rasko *et al.* [64] on the Shiga toxin-producing *E. coli* was based on isolated cultures, Loman *et al.* [59] have demonstrated that genomic reconstruction and identification of the respective outbreak strains are achievable from metagenomic sequence data [59]. Therefore, modern culture-independent workflows, although not yet part of clinical routine, offer a pronounced reduction in turnaround times over culture-dependent techniques (Figure 11.1). As highlighted in the study by Loman *et al.* [59], improved sensitivity and decreasing procedural costs (e.g., for reagents) are likely to support the adaption of metagenomics-based culture-independent approaches in clinical diagnostic practice, specifically in cases of infectious outbreaks of pathogens that elude standard diagnostics.

Periodontal diseases are prime examples of diseases in which imbalances in the microbial ecology, specifically within the oral cavity, play an important role for disease onset and progression. Dysbiosis within the oral cavity has been shown to cause adverse systemic effects in humans with the most compelling links established to cardiovascular disease [67]. Recent studies have demonstrated the potential of metagenomics-based approaches to reveal microbial community compositions associated with periodontitis [68,69].

Furthermore, clear shifts in GIT microbial consortia have also been reported for different cancers, such as oral squamous cell carcinoma [70] and colon cancer [22,71–73]. Most convincingly, evidence for the association of GIT microbial community shifts and disease has been established for both type 1 [74,75] and type 2 [26,76] diabetes mellitus (T1DM or T2DM), respectively. In these comparative studies involving diseased versus normal cohorts, the ratios of *Bacteroidetes* to *Firmicutes* (the two most dominant gastrointestinal phyla) are higher in individuals suffering from diabetes mellitus compared with controls [74–76]. Following experimental validation, such insights may be harnessed for the definition of condition-specific biomarkers (e.g., condition-specific microbial populations) with high diagnostic value analogous to microbial isolates fulfilling Koch's postulates.

The majority of the above-cited studies have involved characterization of the taxonomic structure of the microbial population by 16S rRNA gene sequencing. However, a more in-depth characterization of the metabolic functions carried

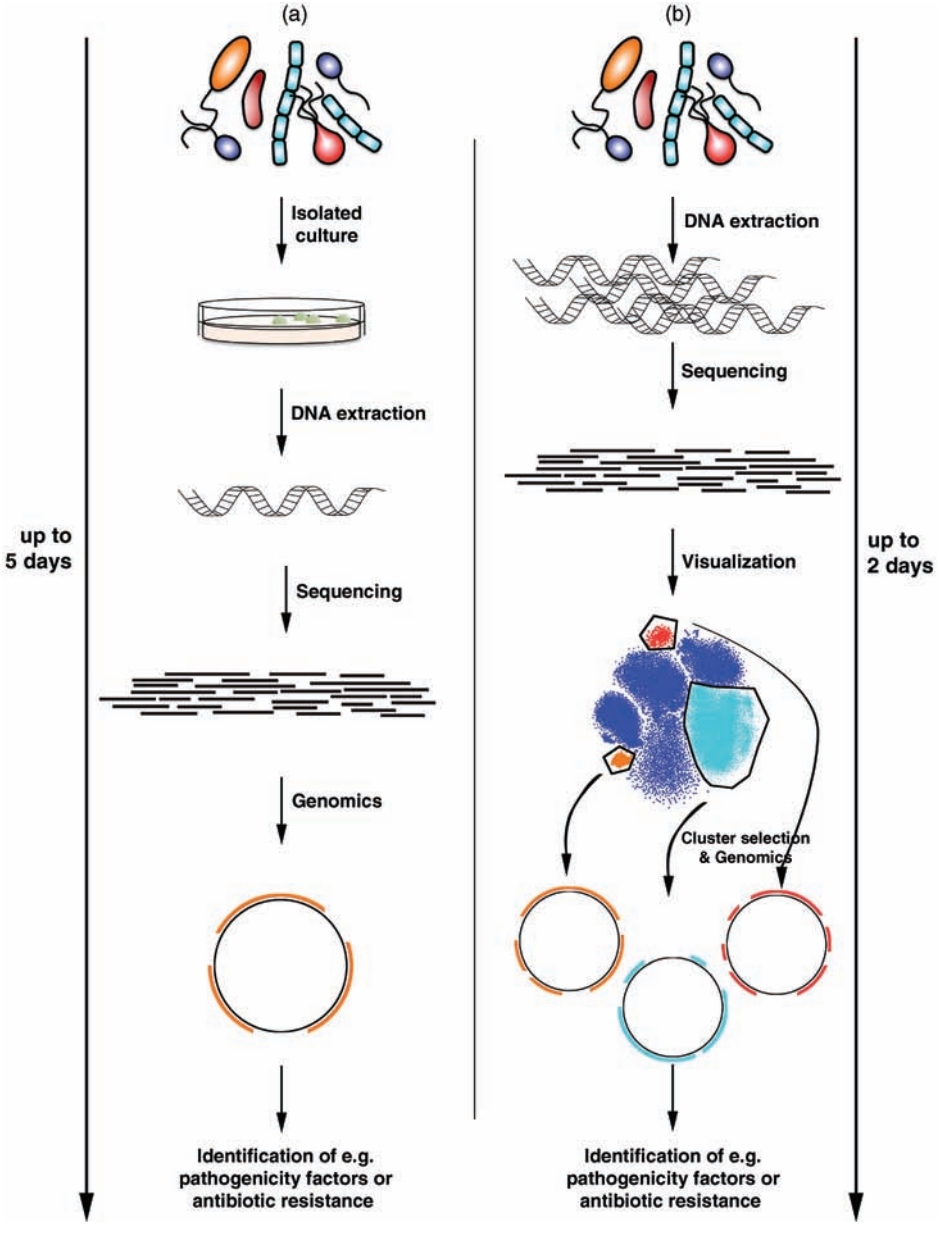


Figure 11.1 Illustration of workflows for the identification of a disease-causing microorganism. (a) Culture-dependent versus (b) culture-independent approaches. Different time requirements for both approaches are indicated based on turnaround times for Illumina MiSeq-based DNA sequencing. The culture-dependent approach mainly leads to the genomic reconstruction of a single population, ideally the one that is causally linked to the infection, whereas the culture-independent approach provides genomic information of the entire microbial community at once given appropriate sequencing depth.

out by individual community members may highlight pathways of particular interest that may explain putative causal links between dysbiosis and pathogenesis. For example, Morgan *et al.* [77] reported a change of 12% in functional genes associated with particular metabolic pathways (e.g., "decreased short-chain fatty acid production" and "increased biosynthesis and transport of compounds advantageous for oxidative stress" in the disease case), while the number of genera differed only by 2% between individuals diagnosed with inflammatory bowel disease and healthy individuals [77]. This work used metagenomic data for the analysis of gene and pathway composition. In a more recent comprehensive metagenome-wide association study involving 345 Chinese subjects, 60 000 T2DM markers were identified and validated for their potential to classify T2DM [26]. Patients with T2DM exhibited a moderate degree of gastrointestinal microbial dysbiosis that was characterized by a decrease in the abundance of some universal butyrate-producing bacteria and an increase in various opportunistic pathogens, such as *Bacteroides caccae*, *Clostridium hathewayi*, *Clostridium ramosum*, *Clostridium symbiosum*, *Eggerthella lenta*, and *E. coli*, as well as an enrichment in functional genes involved in sulfate reduction and oxidative stress resistance [26]. Causal relationships between the dysbiotic states and T2DM have yet to be established. Nonetheless, the study by Qin *et al.* [26] clearly demonstrates the power of microbiome-derived diagnostic markers for diseases associated with microbial dysbiosis. Examples of such biomarkers include KEGG orthologous groups associated with L-aspartate oxidase or starch synthase, as well as drug resistance genes that are found to be greatly enriched in T2DM GIT microbial communities. Such markers are of course of high relevance for effective therapeutic interventions.

The identification of recently acquired mobile genetic elements (e.g., pathogenicity islands and/or antibiotic resistance genes) is particularly pertinent in infectious disease cases. Numerous putative antibiotic resistance genes have been found in a wide range of commensal microbial organisms through the application of metagenomic approaches [78–80]. These genes bear the potential of acquisition by pathogens, which may in turn lead to decreased efficiency of antibiotic therapies in the case of pathogenic infections. In such cases, it is important that acquired genes are rapidly identified for clinicians to tailor antibiotic therapies accordingly. This highlights the current need to link specific gene complements to specific populations, which can only be achieved through genomic linkage. Furthermore, it underlines the importance of the comprehensive characterization of the human microbiome in medical diagnostics. Consequently, a

Detection of distinct populations from metagenomic data occurs here by "binning" (selections of exemplary sequences are highlighted in black). Cluster selection is followed by population-level genomic reconstructions, in turn leading to the simultaneous reconstruction of a plurality of population-level genomes. The in-depth characterization of the genomic reconstructions for the identification of specific disease-relevant genomic regions is carried out in both approaches as a downstream process.

gene-centric metagenomic approach (e.g., [81]) will not be fruitful for metagenomics-based diagnostics since such an approach may be able to identify specific antibiotic resistance genes, but it will not be able to link these to a specific population encoding them (e.g., a specific pathogenic strain).

We would like to highlight that most of the studies discussed above have unraveled putative links between microbial dysbiosis and human diseases. However, it remains to be established whether the shifts in microbial populations that characterize the imbalances are causal or consequential. Nonetheless, the apparent differences in microbial community structure and function offer exciting prospects for diagnosis or, if causal links are demonstrated, future prognosis.

11.4 Challenges for Metagenomics-Based Diagnostics: Read Lengths, Sequencing Library Sizes, and Microbial Community Composition

Microbial genomes, in particular those of bacteria, are of lengths in the range of a few megabases [82]. The reconstruction of previously uncharacterized complete or near-complete microbial population-level genomes from short genomic fragment-based (short read-based) RSS experiments on microbial/environmental samples requires the assembly of the short and randomly distributed sequence fragments into longer, contiguous sequences (contigs) [83–85]. Ultimately, these contigs should be joined to resolve the complete genomic information. Confident genomic reconstructions are of particular interest in diagnostics as entire genomes provide absolute resolution for taxonomic classification [29,30] and also cover the entire functional potential of microbial populations [59,64].

Although individual microbial genomes are rather small [82], technical aspects related to the short-read next-generation sequencing (NGS) techniques, the sheer number of different taxa (complexity), and their relative abundance differences (evenness) render any metagenomic assembly and population-level genome reconstruction challenging [83–85]. At the time of writing, the most commonly used NGS technology is Illumina, yielding at most 2×250-bp long, paired-end reads for the MiSeq system with a throughput usually in the order of several million reads [86]. While sequencing technologies that produce reads of lengths in the order of a few kilobases are available, they are not (yet) commonly used for metagenomic sequencing even though they have been found to be very useful for the sequencing of isolate microbial genomes [87]. These technologies have not been widely applied in metagenomics because, among other reasons, they do not yet provide the throughput of sequencers which generate short reads. Nevertheless, long genomic reads are (generally) desirable due to their extended congruent information content.

Regarding complexity and evenness of microbial communities, samples of different origins will exhibit varying complexities, ranging from around five dominant taxa in acid mine drainage biofilms to an estimated 1 million in soil [88]. In general, a log-normal distribution has been found to be a good approximation

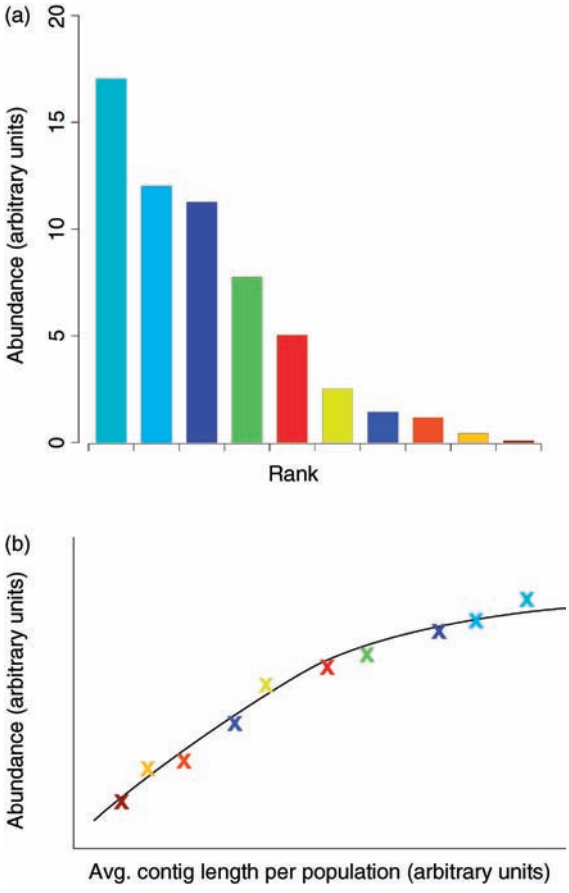


Figure 11.2 Illustration of the relationship between taxon abundances and resulting contig lengths derived from metagenomic data. (a) In microbial communities, taxon abundances follow a rank-abundance distribution. Each bar represents a distinct taxon. (b) Lengths of recoverable contigs are a function of population size: taxa with higher abundance provide more genomic information in the form of obtained reads leading to longer, high-quality contigs.

for species abundance in a wide variety of communities [89] (not only microbial communities) with evidence having been specifically reported for microbial communities in soil samples [90] (Figure 11.2a). Accordingly, the most abundant organisms are likely to be well represented by high-quality, long assembled genomic fragments due to strong consensus (high coverage) in the assembly and suitable coverage over long distances in the host genomes [91]. On the contrary, for the vast majority of organisms, neither of these two is likely to occur as these organisms' abundances will be low, which will manifest in low recovery of genomic information, hence low coverage and, thus, short contigs (Figure 11.2b). Furthermore, recombination events are rampant in microbial communities [78].

These events promote the exchange of genetic material between different community members and continuously lead to the emergence of locally adapted hybrids [92].

Given the characteristics of microbial communities with regard to organismal distribution, any application of metagenomic sequencing requires a specific amount of sequencing depth that, in turn, is based on the underlying question. Concretely, characterizing a targeted subset of genes (e.g., via conserved and hypervariable regions of the 16S rRNA gene) requires fewer amounts of sequencing data than when aiming at full genomic recovery of the constituent microorganisms within a sample. Factors such as unknown overall number of taxa and high amounts of low abundance species in environmental samples [89,93] typically necessitate deep sequencing depth for the recovery of comprehensive genomic information from populations, in particular from those of low abundance [94]. Greatly improved sequencing depth will eventually be feasible due to further reductions in sequencing costs [95]. However, despite deep sequencing, uniform coverage of population-level genomes is unlikely to be obtainable using current sequencing technologies. For example, a 188 times coverage of the dominant *Enterococcus faecium* genome (around 5.5×10^6 reads of size 100 nucleotides), will leave around 1.25 kbp uncovered for an artificial community of 12 species. Furthermore, Ni *et al.* [96] have estimated the required amount for achieving at least 20 times genome coverage of species-level genomes with a relative community-wide species abundance of more than 1% to be at least 7 Gbp for a fecal sample [97]. Manual examination of such large metagenomic datasets is prohibitive; thus, efficient computational approaches for the analysis and comparison of metagenomic data are required.

If we combine all the above-mentioned points with the fact that the human genome is yet to be officially finalized despite over a decade of intense curation, metagenomic assembly, and data analysis is and will remain a major challenge.

11.5 Deconvolution of Population-Level Genomic Complements from Metagenomic Data

While the analytical challenges of metagenomics remain daunting, some of these may be curbed by the fact that analyses of human-associated microbiota in health and disease can be performed using incomplete genomic fragments (e.g., contigs) and therefore finished genomes are not necessary for many applications, in particular in diagnostics where knowledge of sets of specific genomic regions may be sufficient. Diagnostics-relevant information may be obtained, for example, from separating obtained contigs according to their taxon of origin and approximating taxon abundances based on the coverage values deduced following the mapping of metagenomic reads.

The process of grouping genomic sequences according to their taxonomic origin is commonly referred to as "binning." In its most basic form, binning involves the use of different data-inherent characteristics (e.g., %GC or read

coverage) to cluster genomic fragments with similar characteristics that, in turn, are hypothesized to originate from the same microbial population. Binning holds great promise for a variety of applications in metagenomic data analysis as the separation into independent subsets renders parallel workflows possible, decreases the complexity for the downstream analyses, and increases signal-to-noise ratios. Moreover, improved population-level assemblies may be achieved through iterative binning and re-assembly of binned genomic fragments leading to improved genome recovery (Figure 11.3) [98,99]. Finally, the deconvolution of microbial community composition plays an essential role for the identification of condition-specific microbial populations from human metagenomic data.

In the following subsections, we highlight recent human microbiome-focused studies that have employed a subset of available metagenomic data analysis tools for the resolution of the taxonomic structure and/or functional analysis of community members. These tools can generally be divided into (i) reference-dependent approaches relying on prior knowledge (i.e., microbial reference genomes) or (ii) reference-independent approaches that are entirely based on the characteristics of given input sequences and, thus, do not require *a priori* information.

11.5.1 Reference-Dependent Metagenomic Data Analysis

Reference-dependent approaches require *a priori* information for classifying and analyzing newly generated metagenomic data. These approaches mostly use sequence alignment or sequence composition similarity measures for assigning sequence fragments to specific organismal groups. Alignments allow direct sequence comparisons whereas sequence composition-based approaches involve the transformation of the genomic sequences into numerical representations. These representations may be one-dimensional (e.g., %GC) or high-dimensional (e.g., frequencies of oligonucleotide patterns).

11.5.1.1 Alignment-Based Approaches

Reference-dependent approaches based on sequence alignments have recently been applied in multiple large-scale studies, such as the Human Microbiome Project [49,50] and the MetaHIT project [100], for the designation of genomic sequences to different phylogenetic ranks (e.g., genus, species) in order to resolve the taxonomic structures of the microbial communities. However, the identification of specific groups reflecting distinct taxa is often restricted by limited or ambiguous homology between the query sequences and reference sequences. Limited homology may be, among others, due to genuine differences in the genomic sequences (e.g., mutations, insertions/deletions) or it may originate from genome plasticity (e.g., recombination events), the latter causing extensive variations in gene content between closely related strains of the same species [78,101]. Ambiguous homology will be especially pronounced in evolutionary strongly conserved genomic regions, such as rRNA genes.

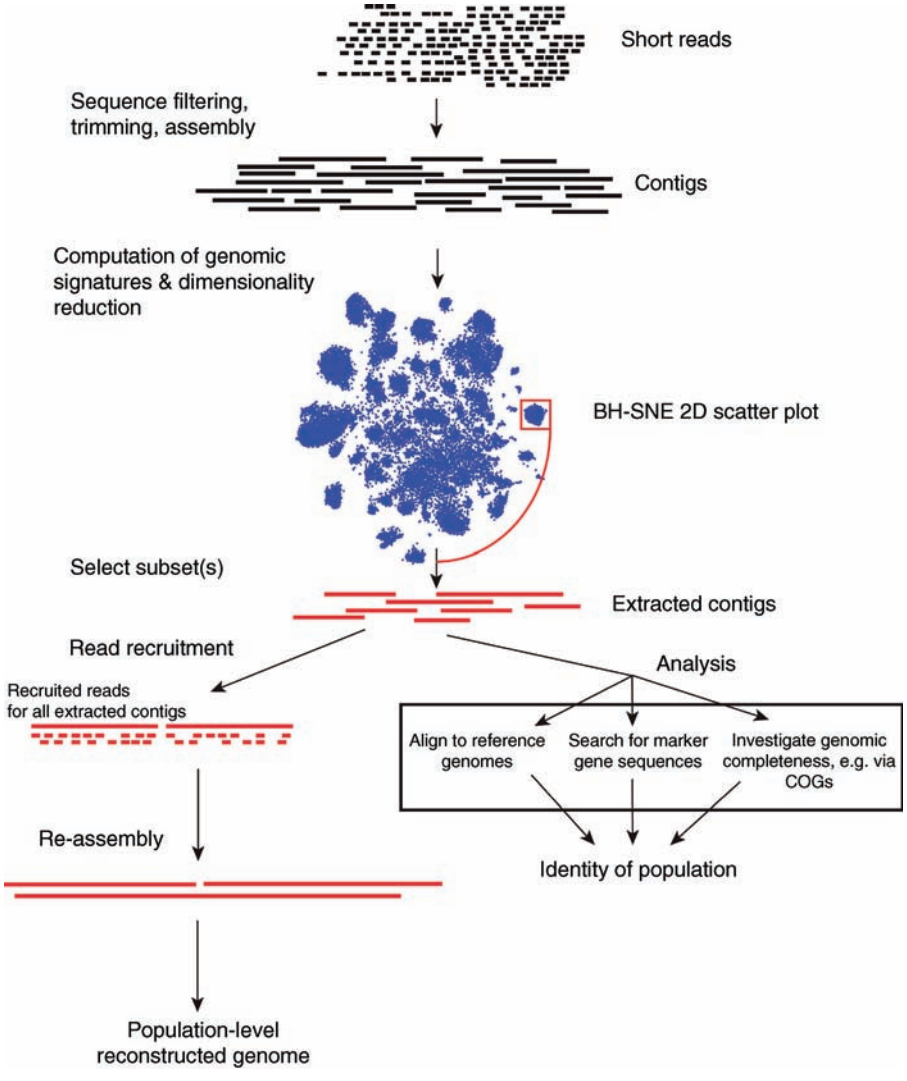


Figure 11.3 Exemplary workflow for population-level genomic reconstruction and population identification. The input (short reads) is processed according to custom quality criteria and the processed reads are then subjected to assembly. Based on the assembled contigs, genomic signatures (e.g., oligonucleotide frequency-based signatures) are computed and transformed with a centered log-ratio transformation. These transformed signatures are then reduced to lower dimensions (here two dimensions (2D)) and the taxonomic structure is visualized for discriminative selection by human input. Based on the selection (bin), the corresponding contigs provide genomic information that can be used to refine the initial assembly or for the analysis of the extracted contigs. The results of the analysis step depend on existing information. In the case of a novel microbial population represented by a bin, genomic reconstruction would be attempted. Novel microbial populations may be characterized by limited homology to existing references, given the assembly step has not introduced errors that may manifest themselves in the identification step. However, high consensus in the assembly indicates likely genuine sequences. COGs, Clusters of Orthologous Groups.

An advantage of reference-dependent alignment-based approaches is that they allow comparably short genomic fragments of metagenomic data, around 75 bp, to be used for analysis. Analysis of the metagenomic data generated as part of the MetaHIT project included gene prediction via MetaGene [102], thereby allowing functional characteristics of the GIT microbial communities to be resolved. MetaGene employs a stochastic approach to annotate metagenomic fragments based on scoring predicted open reading frames (ORFs; genomic regions between a start and a stop codon) and the computation of an optimal (high-scored) combination of ORFs that is then used as the annotated gene set. Other metagenomic data analysis tools include MG-RAST, which is a Web service that provides integrated phylogenetic and functional annotation of metagenomic sequence fragments based on the mining of existing protein and nucleotide databases [103]. In particular, the phylogenetic annotation involves the SEED "nr" database and a collection of rRNA databases [104].

In the context of the human microbiome, the MG-RAST service was used for the characterization of the human skin microbiome from metagenomic sequences of *in situ* obtained samples [79]. This human skin study revealed various staphylococci that are putatively intrinsically methicillin resistant as many of the reads aligned to methicillin-resistance genes according to SEED level 3 subsystems.

11.5.1.2 Sequence Composition-Based Approaches

The composition of a sequence is often defined as the count or frequency of oligonucleotide patterns or *k*-mers (e.g., %GC or tetranucleotide frequency) along a given sequence. An overview of commonly used genomic signature definitions is given in Gori *et al.* [105]. Sequence compositions are typically conserved over entire genomes, including strain-variant regions. Hence, specific genomic signatures can be defined for entire genomes or genomic fragments based on sequence compositions alone and these signatures can then be used to compare them to the signatures of known reference genome sequences. A measure of compositional similarity (e.g., Pearson correlation or Euclidean distance) is then applied for pairwise comparisons instead of a sequence similarity measure as it is used in sequence alignments. Sequence composition-based approaches generally provide superior computational efficiency but may be less sensitive than direct sequence comparisons via alignments. Nevertheless, a detailed phylogenetic resolution may be achievable (i.e., down to species-level resolution) [106–108].

Sommer *et al.* [80] used Phylopythia [109] for metagenomic data analysis in a study with the aim of investigating the presence of antibiotic-resistance genes in the human microbiota of saliva and fecal samples. The selection of the gene-containing fragments for the analysis was based on functional screening for antibiotic resistance of genomic fragments isolated from the individual microbiota. Phylopythia is a composition-based approach employing support vector machines to train taxon classifiers from the genomic signatures of existing reference genomes. New sequences are assigned to specific taxa using these previously trained classifiers.

Irrespective of the application of alignment-based or composition-based approaches, reference-dependent approaches involve searching generated metagenomic data against available genome reference sequences. This process is, however, limiting as current estimates suggest that only about 1% of microbial taxa are characterized [110]. While the proportion of isolated and fully characterized taxa in the human-oriented setting is probably much greater, it is highly questionable if all the taxa have been characterized to their full extent.

11.5.2 Reference-Independent Metagenomic Data Analysis

To account for some of the highlighted limitations of reference-based approaches, alternative methods have been developed that do not require any *a priori* information and, thus, should allow for the generation of new knowledge about microbial community structure and functional relationships. Concretely, these approaches require as input only (assembled) genomic fragments from metagenomic sequencing experiments and their objective is to exploit the data-inherent taxonomic structure so as to reveal clusters of sequences (i.e., bins) that are likely to originate from the same taxon or a set of closely related taxa. Importantly, the sequence information within bins may be curated and exploited for deriving information about the associated functional potential by various downstream genomic analyses. Fundamentally, reference-independent binning approaches are for the most part based on sequence compositional features, which are inherent to any sequence, and it has been shown that the sequence composition enables the differentiation between different microbial taxa [106–108].

Although the bulk of a microbial genome exhibits a highly conserved and specific genomic signature (e.g., 4-mer-based composition) [108], some regions are evolutionary strongly conserved between genomes of different phyla (e.g., rRNA gene sequences), and thereby result in similar and hard-to-differentiate genomic signatures. In contrast, genomic regions that have been recently acquired (e.g., as the result of lateral gene transfer) may also exhibit confounding signatures. In such cases, mate-pair or paired-end information from the metagenomic sequencing enables retracing of individual genomic linkages for the accurate reconstruction of population-level genomes. Subsequently, the estimation of the individual taxon abundance is possible by mapping the original reads back onto the binned genomic sequence fragments. Especially for diagnostic purposes, the ability to asses and compare the relative abundances of microbial community members is important (e.g., to detect overgrowth of indigenous pathobionts) [13]. However, exact estimations of the relative abundances from read coverage may be challenging due to uneven coverage of genomic regions/genomes resulting from biases associated with current sequencing technologies [111,112]. Furthermore, insufficient population-level genome coverage resulting from small to very small populations sizes may further confound data analysis efforts. Such

low abundance organisms may, however, be important to be resolved from a diagnostics point of view because of the existence of keystone species within ecosystems that have a disproportionately large effect on community structure and function relative to their abundance [12,113]. Functional analyses of the genomic information in individual bins (e.g., via gene annotation by MetaGene [102]) allows for the formulation of population-level functional hypotheses that have to be tested using *in vitro*, *in vivo*, or *ex vivo* experiments [114].

Recently, an approach based on the Lander–Waterman model for sequencing [115] has been proposed for reference-independent binning [116]. The Lander–Waterman model was originally developed to mathematically characterize the properties of "islands" (contiguous groups of reads that are connected by overlaps of a certain minimal length) in genomic mapping projects and provides mathematical formulae to, among others, estimate the number of reads that are required to close gaps in genomic sequencing projects, hence, to cover a genome as comprehensively as possible, ultimately leading to "finished genomes." In the binning approach developed by Wu and Ye [116], the abundances of constituent taxa are estimated based on an expectation–maximization algorithm that fits the parameters (abundance level λ_i for the ith bin) for a mixture of Poisson distributions and bins are formed based on the similarity of the estimated abundances. The performance of such an approach is inherently linked to the differences in abundance and the error rate of the binning (i.e., mixing unrelated taxa into common bins) increases strongly when abundance ratios approach unity. This can prove to be a serious limitation as abundances in microbial communities often follow a logarithmic rank-abundance curve and exhibit high numbers of low-abundance organisms (a long tail in the abundance distribution) (Figure 11.2a) [89,93].

The predominant approach for reference-independent binning is based on the exploitation of similarity of short oligonucleotide-based signatures (oligonucleotide patterns are usually between 4 and 6 nucleotides in length). A popular method relies on the calculation of oligonucleotide frequency-based genomic signatures and their visualization using a U-Matrix [117] based on self-organizing maps (SOMs) [98,118–121]. Generally, the visualization allows leveraging the prominent pattern recognition capabilities of the human eye–brain system [122]. A major limitation of this binning approach is that input sequence lengths need to be in the order of 5 kbp in order for signals to be well resolvable [119,120].

Motivated by the power of SOM-based approaches and use of human-based augmentation for the binning of metagenomic data, we have recently developed a sequence composition-based approach for the fast and intuitive visualization of microbial community structure from metagenomic data at genomic fragment lengths around 1 kbp (Figure 11.3). More specifically, the method combines the centered log-ratio transformation of the genomic signatures with Barnes–Hut stochastic neighbor embedding (BH-SNE) – a recent non-linear dimension reduction (NLDR) technique from the machine learning literature [123].

BH-SNE was specifically designed for the low-dimensional (two- or three-dimensional) visualization of high-dimensional data (e.g., oligonucleotide frequency-based signatures) by preserving local neighborhood structure. This preservation is of particular importance for the confident low-dimensional representation of clusters from high-dimensional data and, thus, should allow accurate binning of metagenomic data based on genomic signatures. Our approach improves on the sensitivity, specificity, and speed of existing reference-independent binning approaches employing SOMs, and thus allows for intuitive and efficient data exploration by visualizing the metagenomic data-inherent taxonomic structure [99]. Based on human-augmented input, clusters of interest can be selected for further analyses and/or population-level genomic reconstructions [124] (Figure 11.3). Due to currently necessary assembly steps, the clustered sequences have to be validated with respect to their assembly quality. Such steps may be omitted in future by the use of long read-based sequencing (e.g., [87]) as reads resulting from the use of such technologies provide sufficient signal for our human-augmented binning approach and, thus, the initial assembly step is not required.

11.6 Need for Comparative Metagenomic Data Analysis Tools

The detailed resolution of microbial community structure through the aforementioned methods enables the analysis of condition-specific microbiota, such as identification of specific microbial populations associated with either health or disease states. In particular, condition-specific taxa may be identified through comparison of representative microbial community compositions, associated with either condition (e.g., disease versus control), resolved using, for example, NLDR-based visualization of genomic signatures. In particular, taxa that are exclusive to one condition (e.g., occurring in either case or control datasets) naturally serve as ideal diagnostic markers. Disease-relevant information encoded by specific genomic regions (e.g., those encoding antibiotic resistance genes) will in particular support therapeutic decision making.

As highlighted in Section 11.4, events of genomic transfer are commonplace in microbial communities and, thus, functional differences can possibly be found in similar taxa. Hence, if no or only very minor differences in microbial community structure are to be found, a more in-depth characterization of the functional potential with respect to the individual conditions may be essential. The detection and identification of a prophage that encodes Shiga toxin 2, and a distinct set of additional virulence and antibiotic resistance factors encoded within the genomes of a particularly lethal *E. coli* strain [64], illustrate the need for the identification of specific disease-relevant genomic regions and the importance of strain-resolved comparative metagenomic data analysis tools for diagnostic applications.

11.6.1 Reference-Based Comparative Tools

While a detailed manual comparison of individual taxa between conditions is possible on small datasets and represents the dominating *modus operandi*, this seems prohibitive when considering the comparison of a larger collection of metagenomic datasets with the aim of identification of condition-specific, that is, discriminative, features, such as microbial populations or specific genes. Hence, computational tools have recently been developed to support the detection of particular differences in microbial composition and/or functional potential when considering large numbers of metagenomes. The existing tools are primarily based on the identification of constituent taxa using *a priori* known reference sequences thereby allowing the comparison of taxonomic compositions. Furthermore, some tools offer comparisons of functional potential, for the detection of discriminatory features, mostly via annotation of the respective genomic sequences by homology-based methods. For example, MEGAN can be used for the comparison of taxonomic, SEED, or KEGG annotations, and results in visual and computational comparisons [125]. In the visual comparison, the abundance of each annotation (number of reads that have been assigned to a given annotation) is presented. Detailed differences between the datasets can then be identified by visual inspection. The computational comparison offers a more coarse-grained view of the differences between datasets. METAREP offers statistical tests for the comparison of multiple metagenomic samples based on metagenomic annotation profiles to detect statistically significant differences [126]. Metastats can be used to detect particular features that characterize the discovered differences, such as significantly deregulated genes [127], thus providing means to prioritize further analysis based on the generated hypotheses. LEFse is based on non-parametric tests for the comparison of input features, thus making it more robust to potentially non-normally distributed features, such as organism abundance [128]. Furthermore, it estimates the respective effect size of statistically significant features. The rationale behind using the effect size is that features with high effect sizes are thought to explain more of the observed taxonomic/phenotypic differences and thus are likely to be of increased biological interest, including for diagnostic purposes. The particular strengths of these tools are mainly that they enable targeted follow-up analyses based on identified discriminative features. This is only possible as long as the references or annotations contain sequences providing reasonable homology. However, this may not necessarily be the case as we have detailed above.

11.6.2 Reference-Independent Identification of Condition-Specific Microbial Populations from Human Metagenomic Data

We have shown in Section 11.5.2 that our human-augmented binning approach is an efficient tool for the resolution of microbial community structure without

the requirement of *a priori* information. Here, we introduce an additional functionality of our approach: the identification of condition-specific microbial populations from human metagenomic data. We first illustrate this using in-house metagenomic datasets obtained from fecal samples of 10 healthy individuals. The visual superimposition of individual microbial community structures allows for the identification of condition-specific clusters derived from specific microbial populations (Figure 11.4a). Discrete unique sequence clusters are apparent for a single individual who follows a vegetarian diet (P03), whereas all other individuals follow an omnivorous diet. We are currently performing further analysis of the identified sequences, in particular as it has been previously shown that a vegetarian diet influences the human colonic fecal microbiota as characterized by fecal samples [129]. Preliminary comparisons against existing publicly available microbial reference genomes have shown limited homology for the genomic fragments in the condition-specific cluster structure of our study of 10 healthy individuals, illustrating the utility of efficient unsupervised binning approaches in the detection of, potentially previously uncharacterized, condition-specific microbial populations.

Analogous to our study of 10 healthy individuals, we examined a subsample of the T2DM dataset from Qin *et al.* [26] for condition-specific microbial populations. We compared the microbial communities of five lean males diagnosed with T2DM to those derived from five lean males not known to have the disease. Numerous condition-specific groups of microbial sequences (condition-specific bins) are apparent for the diseased individuals (Figure 11.4b). The majority of the condition-specific bins are shared by two or more T2DM individuals and different condition-specific bins are shared by different T2DM individuals. However, no condition-specific bins appear to be shared by all individuals diagnosed with T2DM, suggesting that differences may be subtle as already described by Qin *et al.* [26].

For similar studies, natural choices for wet-lab validation techniques include polymerase chain reaction-based approaches to specifically confirm the presence/higher abundances of target sequences in the samples of one condition and the lack/lower abundances of these sequences in the other condition(s), such as found exclusively in fecal samples from the vegetarian individual in one of our studies.

11.7
Future Perspectives in Microbiome-Enabled Diagnostics

There is growing interest in determining if compositional and associated functional changes in endogenous microbial communities play defining roles in human health and disease [130]. Actually, there are strong suggestions that microbial dysbiotic states may be causally related to a range of diseases rendering microbiome diagnostics an absolute necessity in the contexts of systems biomedicine and personalized medicine [131–133].

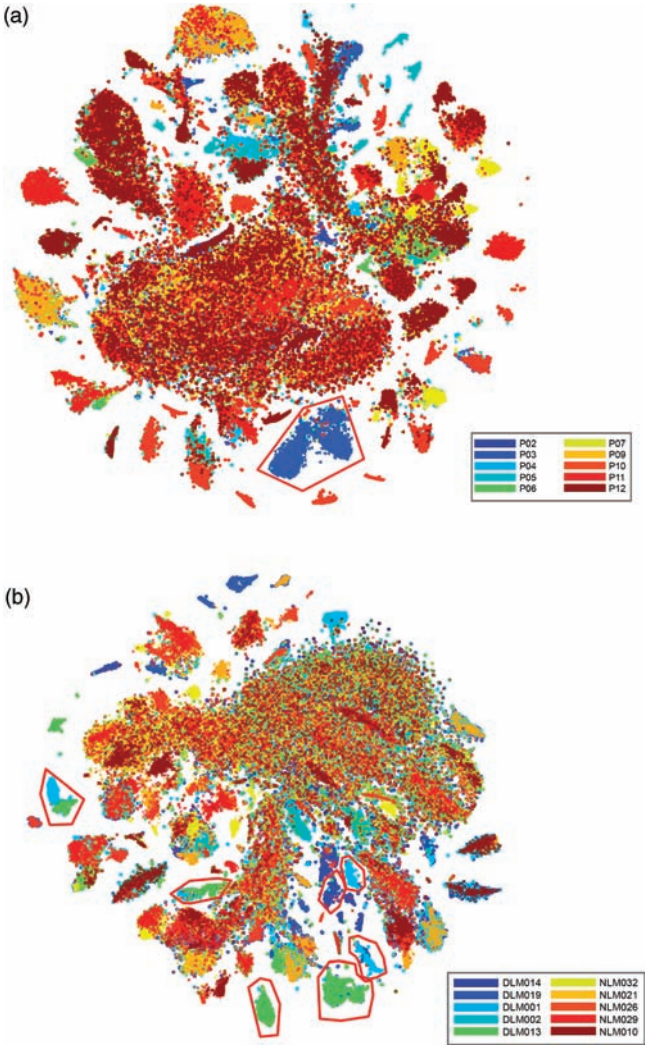


Figure 11.4 Identification of condition-specific microbial populations from metagenomic datasets by superimposition BH-SNE scatter plots. (a) Fecal metagenomic datasets derived from 10 healthy individuals are superimposed for the identification of condition-specific microbial populations. Selected (red polygon) genomic sequences are specific to P03, the only vegetarian in the cohort. (b) Diabetic ($n = 5$) versus non-diabetic ($n = 5$) lean males (DLM and NLM, respectively). Selections (red) represent a subset of clusters with genomic signatures distinctly found in diabetic individuals.

The identification of condition-specific sequences encoded by microbial populations provides the basis for the targeted identification of the respective microbial organism(s) or may serve as *de novo* condition-specific marker sequences when no taxonomic identification is achieved due to missing reference sequences.

Specifically, the examples highlighted in this chapter (i.e., the study of 10 healthy individuals and the T2DM comparative study) demonstrate the applicability and usefulness of our recently developed approach for the identification of condition-specific microbial populations from human metagenomic data. Furthermore, we can imagine that our approach is useful for the comparison of individual patients' microbiota at different time points (e.g., during a diet or therapy), thereby having applications for monitoring therapeutic regimes.

Although fecal samples have been shown to be practical proxies for the evaluation of human GIT microbiome-related health, blood samples are more desirable because of the established routine sampling procedures at basically any time, the decreased risk of sample contamination (fecal samples may come into contact with non-sterile, sanitary surfaces) and constant blood circulation throughout the entire human body [134]. As such, the blood represents a multipurpose body-wide proxy. Wang *et al.* [135] have recently identified a plethora of exogenously derived RNAs in human blood plasma samples, including RNAs from microbial organisms such as bacteria, archaea, and fungi [135]. Microbial RNAs appear to be mostly derived from the GIT microbiome based on the taxonomic affiliation of detected RNAs (Figure 11.5a) [135]. Furthermore, circulating RNAs have the potential to allow the determination of whole-body microbial genomic transcription activity.

Motivated by these recent findings, we are exploring the potential exogenous small RNA complement of human blood particularly with respect to microbial small RNAs in order to exploit this for detecting changes in human microbiome composition and function in particular in the GIT.

By mapping blood-borne small RNAs onto a fecal metagenomic sequence composition-based taxonomic structure map (Figure 11.5b), it becomes clear that there appears to be an over-representation of certain GIT taxa in the exogenous small RNA spectrum in blood and this subset may therefore be harnessed for diagnostic purposes for these specific populations. Such analyses allow the identification of clusters that show a high level of transcription, thus supplementing the analysis of functional potential that is enabled by the use of metagenomics by information on transcriptional activity (metatranscriptomics). This is of particular interest for the detection of lower-abundance organisms that might exhibit a relatively high transcriptional activity, possibly representing keystone species. Similar to approaches that have focused on circulating microRNAs [136,137], the identified sequences represent potential blood-borne diagnostic markers that may enable the monitoring of the activity of particular microorganisms. Naturally, follow-up studies are required to find minimally redundant sets or possibly even single sequences that may provide suitable diagnostic potential.

It follows from the recent discovery of exogenous circulating RNAs that further studies are required to answer open questions such as if, how and why these RNAs, including small RNAs, enter into circulation or what their functions or consequences might be on the host, or on host microbiomes, such as found at

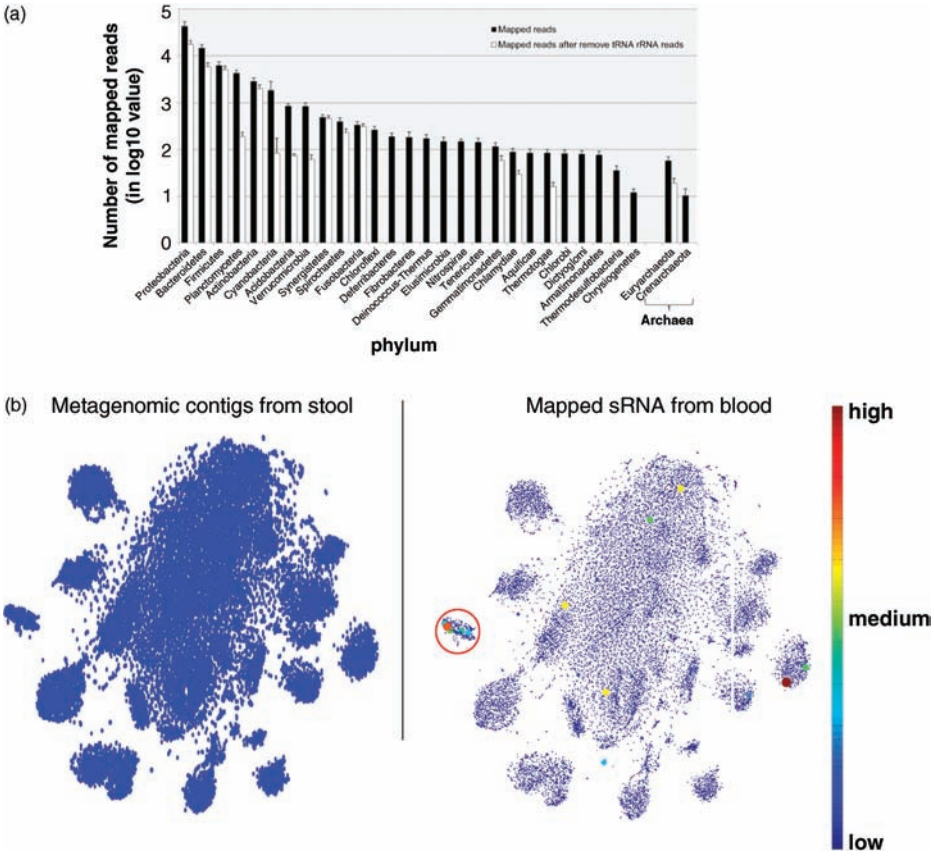


Figure 11.5 Use of exogenous small RNA in samples for microbiome diagnostics. (a) Taxonomic distribution of potential prokaryotic RNA in human blood. The *y*-axis indicates the numbers of reads (log_{10} value) and individual phyla are indicated on the *x*-axis. The number of reads represents the average of all nine plasma samples used in the study. The solid bars represent the total number of processed reads mapped to specific phyla while open bars are the number after removing rRNA and tRNA reads. (Reproduced from [135]. © 2012 Wang *et al.*) (b) Taxonomic distribution of the prokaryotic small RNA (sRNA) in human blood. The two-dimensional structure (reduced from the original high dimensions via BH-SNE) is defined by the contigs obtained from fecal sample metagenomic sequencing and assembly. In the right pane, the color (see color bar) as well as the size of individual points indicates abundance levels of potential small RNAs found in blood that map to assembled contigs from fecal samples. A single cluster exhibiting overall high abundance is highlighted.

distant body sites. Nevertheless, their presence in circulation would represent an invaluable opportunity from a diagnostic point of view and it will be very interesting to see which health-related associations will be revealed from the application of improved analytical techniques.

Acknowledgments

We would like to thank Patrick May, Emilie Muller, Nikos Vlassis, and Dilmurat Yusuf from the Luxembourg Centre for Systems Biomedicine for their assistance and support. The present work was supported by an European Union Joint Programming in Neurodegenerative Diseases grant INTER/JPND/12/01 to P.W. ATTRACT Programme grant to P.W. (A09/03) and an Aide à la Formation Recherche (AFR) grant to C.C.L. (PHD/4964712), funded by the Luxembourg National Research Fund (FNR).

References

1 Zhou, Y., Gao, H., Mihindukulasuriya, K., La Rosa, P., Wylie, K., Vishnivetskaya, T. *et al.* (2013) Biogeography of the ecosystems of the healthy human body. *Genome Biology*, **14** (1), R1.
2 Lin, A., Bik, E., Costello, E., Dethlefsen, L., Haque, R., Relman, D. *et al.* (2013) Distinct distal gut microbiome diversity and composition in healthy children from Bangladesh and the United States. *PLoS One*, **8** (1), e53838.
3 Lazarevic, V., Whiteson, K., Hernandez, D., François, P., and Schrenzel, J. (2010) Study of inter- and intra-individual variations in the salivary microbiota. *BMC Genomics*, **11**, 523.
4 Schloissnig, S., Arumugam, M., Sunagawa, S., Mitreva, M., Tap, J., Zhu, A. *et al.* (2013) Genomic variation landscape of the human gut microbiome. *Nature*, **493** (7430), 45–50.
5 Wong, J., de Souza, R., Kendall, C., Emam, A., and Jenkins, D.J. (2006) Colonic health: fermentation and short chain fatty acids. *Journal of Clinical Gastroenterology*, **40** (3), 235–243.
6 Walter, J. and Ley, R. (2011) The human gut microbiome: ecology and recent evolutionary changes. *Annual Review of Microbiology*, **65**, 411–429.
7 Guarner, F. and Malagelada, J. (2003) Gut flora in health and disease. *Lancet*, **361** (9356), 512–519.
8 Hooper, L., Wong, M., Thelin, A., Hansson, L., Falk, P., and Gordon, J. (2001) Molecular analysis of commensal host-microbial relationships in the intestine. *Science*, **291** (5505), 881–884.
9 Savage, D. (1977) Microbial ecology of the gastrointestinal tract. *Annual Review of Microbiology*, **31**, 107–133.
10 Wilson, M. (2008) *Bacteriology of Humans: An Ecological Perspective*, Blackwell, Malden, MA.
11 Costello, E., Lauber, C., Hamady, M., Fierer, N., Gordon, J., and Knight, R. (2009) Bacterial community variation in human body habitats across space and time. *Science*, **326** (5960), 1694–1697.
12 Bäckhed, F., Ley, R., Sonnenburg, J., Peterson, D., and Gordon, J. (2005) Host–bacterial mutualism in the human intestine. *Science*, **307** (5717), 1915–1920.
13 Kamada, N., Chen, G., Inohara, N., and Núñez, G. (2013) Control of pathogens and pathobionts by the gut microbiota. *Nature Immunology*, **14** (7), 685–690.
14 Packey, C. and Sartor, R. (2009) Commensal bacteria, traditional and opportunistic pathogens, dysbiosis and bacterial killing in inflammatory bowel diseases. *Current Opinion in Infectious Diseases*, **22** (3), 292–301.
15 Cho, I., Yamanishi, S., Cox, L., Methé, B., Zavadil, J., Li, K. *et al.* (2012) Antibiotics in early life alter the murine colonic microbiome and adiposity. *Nature*, **488** (7413), 621–626.
16 Hernández, E., Bargiela, R., Diez, M., Friedrichs, A., Pérez-Cobas, A., Gosalbes, M. *et al.* (2013) Functional consequences of microbial shifts in the human gastrointestinal tract linked to antibiotic treatment and obesity. *Gut Microbes*, **4** (4), 306–315.
17 Kim, H., Borewicz, K., White, B., Singer, R., Sreevatsan, S., Tu, Z. *et al.* (2012)

Microbial shifts in the swine distal gut in response to the treatment with antimicrobial growth promoter, tylosin. *Proceedings of the National Academy of Sciences of the United States of America*, **109** (38), 15485–15490.

18 Knowles, S., Nelson, E., and Palombo, E. (2008) Investigating the role of perceived stress on bacterial flora activity and salivary cortisol secretion: a possible mechanism underlying susceptibility to illness. *Biological Psychology*, **77** (2), 132–137.

19 Brown, K., DeCoffe, D., Molcan, E., and Gibson, D. (2012) Diet-induced dysbiosis of the intestinal microbiota and the effects on immunity and disease. *Nutrients*, **4** (8), 1095–1119.

20 Frank, D., Zhu, W., Sartor, R., and Li, E. (2011) Investigating the biological and clinical significance of human dysbioses. *Trends in Microbiology*, **19** (9), 427–434.

21 Petrova, M., van den Broek, M., Balzarini, J., Vanderleyden, J., and Lebeer, S. (2013) Vaginal microbiota and its role in HIV transmission and infection. *FEMS Microbiology Reviews*, **37** (5), 762–792.

22 Wu, N., Yang, X., Zhang, R., Li, J., Xiao, X., Hu, Y. *et al.* (2013) Dysbiosis signature of fecal microbiota in colorectal cancer patients. *Microbial Ecology*, **66** (2), 462–470.

23 Sobhani, I., Amiot, A., Le Baleur, Y., Levy, M., Auriault, M., Van Nhieu, J. *et al.* (2013) Microbial dysbiosis and colon carcinogenesis: could colon cancer be considered a bacteria-related disease? *Therapeutic Advances in Gastroenterology*, **6** (3), 215–229.

24 D'Aversa, F., Tortora, A., Ianiro, G., Ponziani, F., Annicchiarico, B., and Gasbarrini, A. (2013) Gut microbiota and metabolic syndrome. *Internal and Emergency Medicine*, **8** (Suppl. 1), S11–S15.

25 Sanz, Y., Santacruz, A., and Gauffin, P. (2010) Gut microbiota in obesity and metabolic disorders. *Proceedings of the Nutrition Society*, **69** (3), 434–441.

26 Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F. *et al.* (2012) A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*, **490** (7418), 55–60.

27 Pflughoeft, K.J. and Versalovic, J. (2012) Human microbiome in health and disease. *Annual Review of Pathology*, **7**, 99–122.

28 Gevers, D., Cohan, F., Lawrence, J., Spratt, B., Coenye, T., Feil, E. *et al.* (2005) Re-evaluating prokaryotic species. *Nature Reviews Microbiology*, **3** (9), 733–739.

29 Goris, J., Konstantinidis, K., Klappenbach, J., Coenye, T., Vandamme, P., and Tiedje, J.M. (2007) DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology*, **57** (1), 81–91.

30 Rajendhran, J. and Gunasekaran, P. (2011) Microbial phylogeny and diversity: small subunit ribosomal RNA sequence analysis and beyond. *Microbiological Research*, **166** (2), 99–110.

31 Buchan, B., Ginocchio, C.C., Manii, R., Cavagnolo, R., Pancholi, P., Swyers, L., Thomson, R.B., Anderson, C., Kaul, K., and Ledeboer, N.A. (2013) Multiplex identification of gram-positive bacteria and resistance determinants directly from positive blood culture broths: evaluation of an automated microarray-based nucleic acid test. *PLoS Medicine*, **10** (7), e1001478.

32 Staley, J. and Konopka, A. (1985) Measurement of *in situ* activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Review of Microbiology*, **39**, 321–346.

33 Amann, R., Ludwig, W., and Schleifer, K. (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiological Reviews*, **59** (1), 143–169.

34 Oberauner, L., Zachow, C., Lackner, S., Högenauer, C., Smolle, K., and Berg, G. (2013) The ignored diversity: complex bacterial communities in intensive care units revealed by 16S pyrosequencing. *Scientific Reports*, **3**, 1413.

35 Vartoukian, S., Palmer, R., and Wade, W. (2010) Strategies for culture of 'unculturable' bacteria. *FEMS Microbiology Letters*, **309** (1), 1–7.

36 Amann, R. and Ludwig, W. (2000) Ribosomal RNA-targeted nucleic acid probes for studies in microbial ecology. *FEMS Microbiology Reviews*, **24** (5), 555–565.

37 Richardson, D., Burrows, L., Korithoski, B., Salit, I., Butany, J., David, T. *et al.* (2003) *Tropheryma whippelii* as a cause of afebrile culture-negative endocarditis: the evolving spectrum of Whipple's disease. *Journal of Infection*, **47** (2), 170–173.

38 Dreier, J., Vollmer, T., Freytag, C., Bäumer, D., Körfer, R., and Kleesiek, K. (2008) Culture-negative infectious endocarditis caused by *Bartonella* spp.: 2 case reports and a review of the literature. *Diagnostic Microbiology and Infectious Disease*, **61** (4), 476–483.

39 de Jong, E., Rentenaar, R.J., van Pelt, R., de Lange, W., Schreurs, W., van Soolingen, D. *et al.* (2009) Two cases of *Mycobacterium microti*-induced culture-negative tuberculosis. *Journal of Clinical Microbiology*, **47** (9), 3038–3040.

40 Ngwa, G., Schop, R., Weir, S., León-Velarde, C., and Odumeru, J. (2013) Detection and enumeration of *E. coli* O157: H7 in water samples by culture and molecular methods. *Journal of Microbiological Methods*, **92** (2), 164–172.

41 Coenye, T. and Vandamme, P. (2003) Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. *FEMS Microbiology Letters*, **228** (1), 45–49.

42 Woo, P., Lau, S., Teng, J., Tse, H., and Yuen, K. (2008) Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. *Clinical Microbiology and Infection*, **14** (10), 908–934.

43 Rhoads, D., Cox, S., Rees, E., Sun, Y., and Wolcott, R. (2012) Clinical identification of bacteria in human chronic wound infections: culturing vs. 16S ribosomal DNA sequencing. *BMC Infectious Diseases*, **12**, 321.

44 Schramm, A., Fuchs, B., Nielsen, J., Tonolla, M., and Stahl, D. (2002) Fluorescence *in situ* hybridization of 16S rRNA gene clones (Clone-FISH) for probe validation and screening of clone libraries. *Environmental Microbiology*, **4** (11), 713–720.

45 Ahn, J., Yang, L., Paster, B., Ganly, I., Morris, L., Pei, Z. *et al.* (2011) Oral microbiome profiles: 16S rRNA pyrosequencing and microarray assay comparison. *PLoS One*, **6** (7), e22788.

46 DeSantis, T., Brodie, E., Moberg, J., Zubieta, I., Piceno, Y., and Andersen, G. (2007) High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. *Microbial Ecology*, **53** (3), 371–383.

47 Manichanh, C., Chapple, C., Frangeul, L., Gloux, K., Guigo, R., and Dore, J. (2008) A comparison of random sequence reads versus 16S rDNA sequences for estimating the biodiversity of a metagenomic library. *Nucleic Acids Research*, **36** (16), 5180–5188.

48 Handelsman, J., Rondon, M., Brady, S., Clardy, J., and Goodman, R. (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology*, **5** (10), R245–R249.

49 Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. *Nature*, **486** (7402), 207–214.

50 Human Microbiome Project Consortium (2012) A framework for human microbiome research. *Nature*, **486** (7402), 215–221.

51 Arumugam, M., Raes, J., Pelletier, E., Le, P.D., Yamada, T., Mende, D. *et al.* (2011) Enterotypes of the human gut microbiome. *Nature*, **473** (7346), 174–180.

52 Miller, C., Baker, B., Thomas, B., Singer, S., and Banfield, J.F. (2011) EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biology*, **12** (5), R44.

53 Whitman, W., Coleman, D., and Wiebe, W. (1998) Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences of the United States of America*, **95** (12), 6578–6583.

54 Wilmes, P. (2011) Microbial community proteomics, in *Handbook of Molecular Microbial Ecology I: Metagenomics and Complementary Approaches* (ed. F.J. Bruijn), Wiley-Blackwell, Oxford, pp. 627–636.

55 Stearns, J., Lynch, M., Senadheera, D., Tenenbaum, H., Goldberg, M., Cvitkovitch, D. *et al.* (2011) Bacterial biogeography of the human digestive tract. *Scientific Reports*, **1**, 170.

56 Schwartz, S., Friedberg, I., Ivanov, I., Davidson, L., Goldsby, J., Dahl, D. *et al.* (2012) A metagenomic study of diet-dependent interaction between gut microbiota and host in infants reveals differences in immune response. *Genome Biology*, **13** (4), r32.

57 Chen, H., Yu, Y., Wang, J., Lin, Y., Kong, X., Yang, C. *et al.* (2013) Decreased dietary fiber intake and structural alteration of gut microbiota in patients with advanced colorectal adenoma. *American Journal of Clinical Nutrition*, **97** (5), 1044–1052.

58 Michail, S., Durbin, M., Turner, D., Griffiths, A., Mack, D., Hyams, J. *et al.* (2012) Alterations in the gut microbiome of children with severe ulcerative colitis. *Inflammatory Bowel Diseases*, **18** (10), 1799–1808.

59 Loman, N., Constantinidou, C., Christner, M., Rohde, H., Chan, J., Quick, J. *et al.* (2013) A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of Shiga-toxigenic *Escherichia coli* O104: H4. *JAMA*, **309** (14), 1502–1510.

60 Finegold, S., Dowd, S., Gontcharova, V., Liu, C., Henley, K., Wolcott, R. *et al.* (2010) Pyrosequencing study of fecal microflora of autistic and control children. *Anaerobe*, **16** (4), 444–453.

61 Greninger, A., Chen, E., Sittler, T., Scheinerman, A., Roubinian, N., Yu, G. *et al.* (2010) A metagenomic analysis of pandemic influenza A (2009 H1N1) infection in patients from North America. *PLoS One*, **5** (10), e13381.

62 Tissari, P., Zumla, A., Tarkka, E., Mero, S., Savolainen, L., Vaara, M. *et al.* (2010) Accurate and rapid identification of bacterial species from positive blood cultures with a DNA-based microarray platform: an observational study. *Lancet*, **375** (9710), 224–230.

63 Baraboutis, I., Tsagalou, E., Papakonstantinou, I., Marangos, M., Gogos, C., Skoutelis, A. *et al.* (2011) Length of exposure to the hospital environment is more important than antibiotic exposure in healthcare associated infections by methicillin-resistant *Staphylococcus aureus*: a comparative study. *Brazilian Journal of Infectious Diseases*, **15** (5), 426–435.

64 Rasko, D., Webster, D., Sahl, J., Bashir, A., Boisen, N., Scheutz, F. *et al.* (2011) Origins of the E. coli strain causing an outbreak of hemolytic-uremic syndrome in Germany. *New England Journal of Medicine*, **365** (8), 709–717.

65 Deck, M., Anderson, E., Buckner, R., Colasante, G., Coull, J., Crystal, B. *et al.* (2012) Multicenter evaluation of the *Staphylococcus* QuickFISH method for simultaneous identification of *Staphylococcus aureus* and coagulase-negative staphylococci directly from blood culture bottles in less than 30 minutes. *Journal of Clinical Microbiology*, **50** (6), 1994–1998.

66 Sauer, S., Freiwald, A., Maier, T., Kube, M., Reinhardt, R., Kostrzewa, M. *et al.* (2008) Classification and identification of bacteria by mass spectrometry and computational analysis. *PLoS One*, **3** (7), e2843.

67 Beck, J. and Offenbacher, S. (2005) Systemic effects of periodontitis: epidemiology of periodontal disease and cardiovascular disease. *Journal of Periodontology*, **76** (11 Suppl.), 2089–2100.

68 Liu, B., Faller, L., Klitgord, N., Mazumdar, V., Ghodsi, M., Sommer, D. *et al.* (2012) Deep sequencing of the oral microbiome reveals signatures of periodontal disease. *PLoS One*, **7** (6), e37919.

69 Wang, J., Qi, J., Zhao, H., He, S., Zhang, Y., Wei, S. *et al.* (2013) Metagenomic sequencing reveals microbiota and its functional potential associated with periodontal disease. *Scientific Reports*, **3**, 1843.

70 Meurman, J.H. (2010) Oral microbiota and cancer. *Journal of Oral Microbiology*, **10** (2). doi: 10.3402/jom.v2i0.5195

71 Cario, E. (2013) Microbiota and innate immunity in intestinal inflammation and neoplasia. *Current Opinion in Gastroenterology*, **29** (1), 85–91.

72 Pushalkar, S., Ji, X., Li, Y., Estilo, C., Yegnanarayana, R., Singh, B. *et al.* (2012) Comparison of oral microbiota in tumor and non-tumor tissues of patients with oral squamous cell carcinoma. *BMC Microbiology*, **12**, 144.

73 Sanapareddy, N., Legge, R., Jovov, B., McCoy, A., Burcal, L., Araujo-Perez, F. *et al.* (2012) Increased rectal microbial richness is associated with the presence of colorectal adenomas in humans. *ISME Journal*, **6** (10), 1858–1868.

74 Brown, C., Davis-Richardson, A., Giongo, A., Gano, K., Crabb, D., Mukherjee, N. *et al.* (2011) Gut microbiome metagenomics analysis suggests a functional model for the development of autoimmunity for type 1 diabetes. *PLoS One*, **6** (10), e25792.

75 Giongo, A., Gano, K., Crabb, D., Mukherjee, N., Novelo, L., Casella, G. *et al.* (2011) Toward defining the autoimmune microbiome for type 1 diabetes. *ISME Journal*, **5** (1), 82–91.

76 Larsen, N., Vogensen, F., van den Berg, F., Nielsen, D., Andreasen, A., Pedersen, B. *et al.* (2010) Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLoS One*, **5** (2), e9085.

77 Morgan, X., Tickle, T., Sokol, H., Gevers, D., Devaney, K., Ward, D. *et al.* (2012) Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biology*, **13** (9), R79.

78 Smillie, C., Smith, M., Friedman, J., Cordero, O., David, L., and Alm, E.J. (2011) Ecology drives a global network of gene exchange connecting the human microbiome. *Nature*, **480** (7376), 241–244.

79 Mathieu, A., Delmont, T., Vogel, T., Robe, P., Nalin, R., and Simonet, P. (2013) Life on human surfaces: skin metagenomics. *PLoS One*, **8** (6), e65288.

80 Sommer, M., Dantas, G., and Church, G. (2009) Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science*, **325** (5944), 1128–1131.

81 Tringe, S., von Mering, C., Kobayashi, A., Salamov, A., Chen, K., Chang, H. *et al.* (2005) Comparative metagenomics of microbial communities. *Science*, **308** (5721), 554–557.

82 Raes, J., Korbel, J., Lercher, M., von Mering, C., and Bork, P. (2007) Prediction of effective genome size in metagenomic samples. *Genome Biology*, **8** (1), R10.

83 Peng, Y., Leung, H., Yiu, S., and Chin, F. (2012) IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, **28** (11), 1420–1428.

84 Kultima, J., Sunagawa, S., Li, J., Chen, W., Chen, H., Mende, D. *et al.* (2012) MOCAT: a metagenomics assembly and gene prediction toolkit. *PLoS One*, **7** (10), e47656.

85 Ruby, J., Bellare, P., and Derisi, J. (2013) PRICE: software for the targeted assembly of components of (Meta) genomic sequence data. *G3 (Bethesda)*, **3** (5), 865–880.

86 Illumina (2013) MiSeq Specifications; http://www.illumina.com/systems/miseq/performance_specifications.ilmn.

87 Chin, C., Alexander, D., Marks, P., Klammer, A., Drake, J., Heiner, C. *et al.* (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods*, **10** (6), 563–569.

88 Wilmes, P., Simmons, S.L., Denef, V.J., and Banfield, J.F. (2009) The dynamic genetic repertoire of microbial communities. *FEMS Microbiology Reviews*, **33**, 109–132.

89 McGill, B., Maurer, B., and Weiser, M. (2006) Empirical evaluation of neutral theory. *Ecology*, **87** (6), 1411–1423.

90 Doroghazi, J. and Buckley, D. (2008) Evidence from GC-TRFLP that bacterial communities in soil are lognormally distributed. *PLoS One*, **3** (8), e2910.

91 Treangen, T., Koren, S., Sommer, D., Liu, B., Astrovskaya, I., Ondov, B. *et al.* (2013)

MetAMOS: a modular and open source metagenomic assembly and analysis pipeline. *Genome Biology*, **14**, R2.

92 Denef, V., Kalnejais, L., Mueller, R., Wilmes, P., Baker, B., Thomas, B. *et al.* (2010) Proteogenomic basis for ecological divergence of closely related bacteria in natural acidophilic microbial communities. *Proceedings of the National Academy of Sciences of the United States of America*, **107** (6), 2383–2390.

93 Haegeman, B., Hamelin, J., Moriarty, J., Neal, P., Dushoff, J., and Weitz, J. (2013) Robust estimation of microbial diversity in theory and in practice. *ISME Journal*, **7** (6), 1092–1101.

94 Wendl, M., Kota, K., Weinstock, G., and Mitreva, M. (2012) Coverage theories for metagenomic DNA sequencing based on a generalization of Stevens' theorem. *Journal of Mathematical Biology*, **67** (5), 1141–1161.

95 Stein, L. (2010) The case for cloud computing in genome informatics. *Genome Biology*, **11**, 207.

96 Ni, J., Yan, Q., and Yu, Y. (2013) How much metagenomic sequencing is enough to achieve a given goal? *Scientific Reports*, **3**, 1968.

97 Eckburg, P., Bik, E., Bernstein, C., Purdom, E., Dethlefsen, L., Sargent, M. *et al.* (2005) Diversity of the human intestinal microbial flora. *Science*, **308** (5728), 1635–1638.

98 Sharon, I., Morowitz, M., Thomas, B., Costello, E., Relman, D., and Banfield, J.F. (2013) Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. *Genome Research*, **23** (1), 111–120.

99 Laczny, C.C., Pinel, N., Vlassis, N., and Wilmes, P. (2014) Alignment-free visualization of metagenomic data by nonlinear dimension reduction. *Scientific Reports*, **4**, 4516.

100 Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K., Manichanh, C. *et al.* (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, **464** (7285), 59–65.

101 Medini, D., Donati, C., Tettelin, H., Masignani, V., and Rappuoli, R. (2005) The microbial pan-genome. *Current Opinion in Genetics & Development*, **15** (6), 589–594.

102 Noguchi, H., Park, J., and Takagi, T. (2006) MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Research*, **34** (19), 5623–5630.

103 Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E., Kubal, M. *et al.* (2008) The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*, **9**, 386.

104 Overbeek, R., Begley, T., Butler, R., Choudhuri, J., Chuang, H., Cohoon, M. *et al.* (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Research*, **33** (17), 5691–5702.

105 Gori, F., Mavroedis, D., Jetten, M., and Marchiori, E. (2011) Genomic signatures for metagenomic data analysis: exploiting the reverse complementarity of tetranucleotides, in *Proceedings of the IEEE International Conference on Systems Biology (ISB)*, IEEE, New York, pp. 149–154.

106 Karlin, S., Ladunga, I., and Blaisdell, B. (1994) Heterogeneity of genomes: measures and values. *Proceedings of the National Academy of Sciences of the United States of America*, **91** (26), 12837–12841.

107 Karlin, S. (1998) Global dinucleotide signatures and analysis of genomic heterogeneity. *Current Opinion in Microbiology*, **1** (5), 598–610.

108 Noble, P., Citek, R., and Ogunseitan, O. (1998) Tetranucleotide frequencies in microbial genomes. *Electrophoresis*, **19** (4), 528–535.

109 McHardy, A., Martín, H., Tsirigos, A., Hugenholtz, P., and Rigoutsos, I. (2007) Accurate phylogenetic classification of variable-length DNA fragments. *Nature Methods*, **4** (1), 63–72.

110 Hongoh, Y. and Toyoda, A. (2011) Whole-genome sequencing of unculturable bacterium using whole-

genome amplification. *Methods in Molecular Biology*, **733**, 25–33.
111 Minoche, A., Dohm, J., and Himmelbauer, H. (2011) Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biology*, **12** (11), R112.
112 Quail, M., Smith, M., Coupland, P., Otto, T., Harris, S., Connor, T. *et al.* (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, **13**, 341.
113 Ze, X., Duncan, S., Louis, P., and Flint, H. (2012) *Ruminococcus bromii* is a keystone species for the degradation of resistant starch in the human colon. *ISME Journal*, **6** (8), 1535–1543.
114 Fritz, J., Desai, M., Shah, P., Schneider, J., and Wilmes, P. (2013) From meta-omics to causality: experimental models for human microbiome research. *Microbiome*, **1**, 14.
115 Lander, E. and Waterman, M. (1988) Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics*, **2** (3), 231–239.
116 Wu, Y. and Ye, Y. (2011) A novel abundance-based algorithm for binning metagenomic sequences using *l*-tuples. *Journal of Computational Biology*, **18** (3), 523–534.
117 Ultsch, A. (1993) Self-organizing neural networks for visualisation and classification, in *Information and Classification*, Springer, Berlin, pp. 307–313.
118 Dick, G., Andersson, A., Baker, B., Simmons, S., Thomas, B., Yelton, A. *et al.* (2009) Community-wide analysis of microbial genome sequence signatures. *Genome Biology*, **10**, R85.
119 Abe, T., Sugawara, H., Kinouchi, M., Kanaya, S., and Ikemura, T. (2005) Novel phylogenetic studies of genomic sequence fragments derived from uncultured microbe mixtures in environmental and clinical samples. *DNA Research*, **12** (5), 281–290.
120 Wrighton, K., Thomas, B., Sharon, I., Miller, C., Castelle, C., VerBerkmoes, N. *et al.* (2012) Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science*, **337** (6102), 1661–1665.
121 Wilmes, P., Andersson, A., Lefsrud, M., Wexler, M., Shah, M., Zhang, B. *et al.* (2008) Community proteogenomics highlights microbial strain-variant protein expression within activated sludge performing enhanced biological phosphorus removal. *ISME Journal*, **2** (8), 853–864.
122 Sharpee, T., Kouh, M., and Reynolds, J. (2013) Trade-off between curvature tuning and position invariance in visual area V4. *Proceedings of the National Academy of Sciences of the United States of America*, **110** (28), 11618–11623.
123 van der Maaten, L. (2013) Barnes-Hut-SNE, in *International Conference on Learning Representation*; http://arxiv.org/abs/1301.3342.
124 Iverson, V., Morris, R., Frazar, C., Berthiaume, C., Morales, R., and Armbrust, E. (2012) Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. *Science*, **335** (6068), 587–590.
125 Mitra, S., Klar, B., and Huson, D. (2009) Visual and statistical comparison of metagenomes. *Bioinformatics*, **25** (15), 1849–1855.
126 Goll, J., Rusch, D., Tanenbaum, D., Thiagarajan, M., Li, K., Methé, B. *et al.* (2010) METAREP: JCVI metagenomics reports – an open source tool for high-performance comparative metagenomics. *Bioinformatics*, **26** (20), 2631–2632.
127 White, J., Nagarajan, N., and Pop, M. (2009) Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Computational Biology*, **5** (4), e1000352.
128 Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. *et al.* (2011) Metagenomic biomarker discovery and explanation. *Genome Biology*, **12** (6), R60.
129 Zimmer, J., Lange, B., Frick, J., Sauer, H., Zimmermann, K., Schwiertz, A. *et al.* (2012) A vegan or vegetarian diet substantially alters the human colonic

faecal microbiota. *European Journal of Clinical Nutrition*, **66** (1), 53–60.

130 Shanahan, F. (2012) The gut microbiota – a clinical perspective on lessons learned. *Nature Reviews Gastroenterology & Hepatology*, **9** (10), 609–614.

131 Greenblum, S., Turnbaugh, P., and Borenstein, E. (2012) Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. *Proceedings of the National Academy of Sciences of the United States of America*, **109** (2), 594–599.

132 Maurice, C., Haiser, H., and Turnbaugh, P. (2013) Xenobiotics shape the physiology and gene expression of the active human gut microbiome. *Cell*, **152** (1–2), 39–50.

133 Saad, R., Rizkallah, M., and Aziz, R. (2012) Gut Pharmacomicrobiomics: the tip of an iceberg of complex interactions between drugs and gut-associated microbes. *Gut Pathogens*, **4** (1), 16.

134 Amar, J., Lange, C., Payros, G., Garret, C., Chabo, C., Lantieri, O. *et al.* (2013) Blood microbiota dysbiosis is associated with the onset of cardiovascular events in a large general population: the D.E.S.I.R. study. *PLoS One*, **8** (1), e54461.

135 Wang, K., Li, H., Yuan, Y., Etheridge, A., Zhou, Y., Huang, D. *et al.* (2012) The complex exogenous RNA spectra in human plasma: an interface with human gut biota? *PLoS One*, **7** (12), e51009.

136 Keller, A., Leidinger, P., Bauer, A., Elsharawy, A., Haas, J., Backes, C. *et al.* (2011) Toward the blood-borne miRNome of human diseases. *Nature Methods*, **8** (10), 841–843.

137 Williams, Z., Ben-Dov, I., Elias, R., Mihailovic, A., Brown, M., Rosenwaks, Z. *et al.* (2013) Comprehensive profiling of circulating microRNA via small RNA sequencing of cDNA libraries reveals biomarker potential and limitations. *Proceedings of the National Academy of Sciences of the United States of America*, **110** (11), 4255–4260.

12
Genome, Exome, and Gene Panel Sequencing in a Clinical Setting

Claudia Durand and Saskia Biskup

12.1
Introduction

12.1.1
Genetic Inheritance and Sequencing

In the nineteenth century, Gregor Mendel had the idea that distinct factors were responsible for the heredity of specific traits of living organisms and that these factors were passed from one generation to the next. Some decades later, Thomas Hunt Morgan and Oswald Avery demonstrated that chromosomes are carriers of genetic information in a cell and that DNA is the molecule that encodes this information. The first genetic maps were created by targeted crossing of *Drosophila* flies, but for a very long time it seemed impossible to create similar maps for human beings or to retrieve any information about the human genome.

A huge breakthrough was the elucidation of the molecular structure of DNA and eventually the development of the dideoxy sequencing method by Frederick Sanger in 1977 [1]. Using the Sanger sequencing method, it was possible to decode small sections of the human genome step by step and to elucidate the genetic cause of a reasonable number of hereditary diseases. This technique and the resulting genetic knowledge formed the basis for the emergence of human genetic diagnostics. Another milestone for this discipline was the completion of the Human Genome Project in 2003 that had aimed at the complete sequencing of the 3.2 billion base pairs of the human nuclear genome by Sanger sequencing [2]. This project meant a huge effort for the human genetics community and required the concerted action of several large sequencing centers providing hundreds of Sanger sequencing machines over a timeframe of more than 10 years. The cost of this project was about US$3 billion; roughly US$1 per sequenced base. The deciphering of the human genome brought genetic diagnostics to a new era; however, it remained very time-consuming and expensive to analyze larger amounts of a patient's DNA (e.g., several candidate genes for a patient's disease).

Nucleic Acids as Molecular Diagnostics, First Edition. Edited by Andreas Keller and Eckart Meese.

Over recent years, novel innovative high-throughput sequencing methods have been developed that have meant a paradigm shift in human genetics research and clinics. These new technologies are based on the idea of massively parallel sequencing of millions of DNA fragments in a single sequencing run and are traditionally referred to as "next-generation sequencing" (NGS). Several biotechnological companies now offer NGS platforms that are based on different technologies, but all enable rapid and cost-efficient sequencing. At costs of several thousand dollars, current NGS sequencing machines can generate up to 600 Gb per run, which corresponds to the size of 200 human genomes. These huge sequencing capacities have already changed, and furthermore will fundamentally alter, the field of human molecular diagnostics.

12.1.2 Genetic Testing by DNA Sequencing

As technological progress opens up undreamt of possibilities to analyze the human genome, it is vital to be aware of some basic principles about why and when to consider genetic testing in a patient.

The goal of genetic testing should always be to provide help for the individual patient and/or their families. In contrast to other diagnostic procedures, the results of genetic tests do not only affect the patient, but potentially also the entire family. Therefore, genetic counseling is essential before and after genetic testing. As for all diagnostic procedures, a prerequisite for genetic diagnosis is the informed and voluntary consent of the patient and the relatives. Genetic tests should not be performed at the request of members of a patient's family or other third parties (e.g., insurers, employers) without the expressly written consent of the patient. In Germany, human genetic testing and the use of genetic material and data is regulated by a particular genetic diagnostics law – the Gendiagnostikgesetz (GenDG) – that became applicable in 2010.

The most important reason to consider genetic testing would obviously be if the respective test could detect an inherited disorder ideally even before the development of disease manifestations, enabling causative or preventive treatment. Unfortunately, this scenario is fairly uncommon, as disease-modifying or curative treatments are lacking for most genetic diseases. However, genetic testing can and should also be considered for other reasons: patients and their families have a right to know about genetic diseases running in their families in order to have the possibility of taking informed decisions about life and family planning. Moreover, a genetic diagnosis obviates further diagnostic procedures, which are often costly and sometimes invasive or even harmful for the patient. Similarly, futile or unnecessary side-effects of treatment can be avoided.

DNA sequencing has enabled diagnostic testing for a rapidly increasing number of patients suffering from genetic diseases. However, even today, diagnostic sequencing approaches are only clinically helpful if a monogenic disease or a rare monogenic form of an otherwise common disease is suspected. In these cases, it is possible to correlate a mutation in a single gene with the phenotype

of a patient. If only one or a few genes are known to cause a suspected disorder, genetic testing is performed by targeted mutational analysis. Using specific primers to the genomic region of interest and subsequent amplification by polymerase chain reaction (PCR), a single gene of interest can easily be sequenced by conventional Sanger sequencing. However, in genetically heterogeneous syndromes where numerous genes are known to be able to cause the phenotype, such targeted sequencing approaches will very often fail. For these cases, the recent advances in sequencing technologies (NGS) have meant a great leap forward towards a secure genetic diagnosis, since the increased sequencing capacity allows the parallel identification of sequence variants in dozens or even hundreds of genes that might be implicated in the disorder. For most medical genomic analyses, sequences of a set of genes of specific interest (panel sequencing) or all exonic sequences (whole-exome sequencing (WES)) are selectively enriched or amplified to detect potentially causative mutations in these regions of interest. For a comprehensive analysis, it is also feasible to examine the whole genome of a patient (whole-genome sequencing (WGS)), which can be done in less than 1 week using currently available sequencing methods.

Each of the different approaches described has a preferential scope of application in terms of genetic diagnostic testing. The following sections will describe the different applications of NGS in genetic diagnostics and will illustrate the approaches with practical examples from the clinic.

12.2 Genetic Diagnostics from a Laboratory Perspective – From Sanger to NGS

12.2.1 Sanger Sequencing

The chain termination sequencing method was developed in 1977 by Frederick Sanger [1]. It has been the most widely used sequencing method for nearly 30 years and still provides the basis for many modern sequencing techniques. Together with Walter Gilbert, Sanger was awarded with the Nobel Prize for Chemistry in 1980. In clinical settings, Sanger sequencing is still the gold standard for sequencing individual genes and for validation of NGS results. Key to the method is that a new strand is built on the basis of a single-stranded DNA template while adding small amounts of dideoxynucleotides that lead to a chain termination when being incorporated into the nascent strand. Using this technique, it is possible to sequence up to around 1000 bp of DNA with a very high accuracy. Sanger sequencing is always preceded by a PCR reaction, in which specific primers for the region of interest (e.g., the different exons of a candidate gene) are used to specifically amplify genomic DNA (gDNA). This amplification product is subsequently used in the Sanger sequencing reaction. Sequencing a single gene with several exons therefore requires several PCR amplifications followed by the respective number of sequencing reactions, rendering the

sequencing process an elaborate and time-consuming process. Modern Sanger sequencers are nowadays able to analyze up to 96 PCR products in parallel.

12.2.2 NGS

The high demand for more efficient sequencing methods has driven the development of novel sequencing technologies (so-called NGS) that parallelize and accelerate the sequencing process, and overcome the barriers of Sanger sequencing. As in Sanger sequencing, currently marketed NGS technologies use the basic principle that a complementary DNA sequence is synthesized on the basis of a DNA template and that the assembly of the newly synthesized strand is monitored by the incorporation of new nucleotides. However, in contrast to Sanger sequencing, NGS allows the parallel detection of millions to billions of different DNA molecules. In clinical diagnostics, NGS therefore enables the fast and inexpensive sequencing of hundreds of genes in parallel, which has revolutionized detection rates for many human genetic diseases.

There exist multiple different NGS technologies, all with advantages and disadvantages in clinical use (for technological descriptions of the different platforms, see [3,4]). There are four main factors distinguishing the different platforms with regard to clinical application:

- *Read length*. Read lengths of present platforms range from 35 bp (Life Technologies SOLiD) up to 1000 bp (Roche 454). Longer read lengths are ideal for retrieving haplotype information over a range of hundreds of base pairs, which can eliminate the need for segregation analysis for two recessive mutations. Moreover, long reads facilitate alignment especially in genomic regions with repetitive sequences. For both applications it is often beneficial to subsequently sequence both ends of each analyzed DNA molecule within one NGS run (paired-end sequencing), increasing the overall sequence length and mappability. The first single-molecule sequencers are already on the market that are able to generate average sequences of 4.2–8.5 kb and maximally up to 30 kb (PacBio RSII). This technology and future developments offering even longer read lengths (e.g., nanopore sequencing) will allow us to extend the use of NGS to further clinical applications. Novel long-read techniques (e.g., PacBio RSII, Oxford Nanopores) may further facilitate the reliable identification of insertions, deletions, or translocations, and the identification of compound mutations and haplotyping.
- *Output (and time requirements)*. The maximum output that is produced by an NGS platform is currently 600 Gb of sequence per run (Illumina HiSeq2500). While these high-output instruments were primarily intended to be used for large-scale research projects including WES or WGS, they have also quickly taken over the field of human genetics diagnostics for targeted sequencing of disease-associated gene panels. Depending on the target region size of such panels, tens to hundreds of patient samples may be

Table 12.1 Comparison of the different diagnostic approaches from a technical point of view.

	WGS	WES	Panel sequencing	Single-gene sequencing
Sequencing method used	NGS	NGS	NGS	Sanger
Sequencing length	200–1000 bp (including paired ends)	100–500 bp (including paired ends)	100–500 bp (including paired ends)	100–1000 bp
Sequencer output/ run	up to 600 Gb	up to 600 Gb	10 Mb–600 Gb	<0.1 Mb
Number of analyzed genes	all genes plus all non-coding DNA	all known and transcribed genes (~20 000)	3–500	1
Length of target region	3 Gb	60 Mb	200 kb–1.5 Mb	2 kb (on average)
Average coverage	5–50×	up to 150×	up to 1000×	1×
Required sequencing output	15–150 Gb	9 Gb	0.2–1.5 Gb	2 kb (on average)

sequenced in parallel in one HiSeq2500 run (Table 12.1). However, smaller benchtop sequencers are also available for the lower throughput of clinical samples (MiSeq, Ion Torrent PGM/Proton) allowing the generation of up to 10–15 Gb of sequencing data in one run. The main advantage of such small-size sequencers is their fast processing time when time to result is critical for clinical outcome. Such platforms are able to generate sequencing reports within 1–2 days upon arrival of fresh samples. In comparison, a HiSeq 2500 run alone (without sample preparation) takes 2–6 days depending on overall sequencing output.

- *Accuracy*. An important factor for a reliable clinical diagnostics is the sequencing accuracy, which still is a drawback of most NGS platforms. The average error rates of the most common platforms are within the range of 0.002–0.008 (corresponding to one error per 125–500 bp) [5]. Generally, accuracy problems can be overcome by sequencing patient samples with a high coverage, meaning that each base in the template is covered by tens to hundreds of sequencing reads. This renders an error in a single sequencing read comparatively negligible. Nevertheless, for clinical diagnostics, important findings that have been obtained by NGS are still validated by highly accurate Sanger sequencing.
- *Operating costs*. A major advantage for widespread diagnostic use is a low operating cost. Main cost drivers are investment costs for sequencing machines, costs for chemistry used in the machines, and the man-hours

needed for sample preparation. There is a trend towards simpler sequencing technologies that can reduce costs in terms of sequencing machines and chemistry. For example, Ion Torrent developed a new sequencing technology that uses inexpensive unlabeled nucleotides together with technology from the semiconductor industry instead of expensive fluorescently labeled nucleotides and a complicated and costly optical system (Illumina, PacBio). Moreover, detection systems based on semiconductor chips are promising approaches to easily increase the output with consistent operating costs and time requirements. Lower run times are also contributing to the cost-efficient use of NGS technologies.

The continuous development of novel NGS technologies and the permanent improvement of existing systems opens up great opportunities also in the diagnostic field. However, it is impossible to recommend one specific platform as advantages and disadvantages of the various systems also depend on the desired application.

12.2.3 Practical Workflow: From a Patient's DNA to NGS Sequencing Analysis

Before being analyzed on a sequencing machine, gDNA has to undergo several procedures, such as quality checks, shearing, enrichment of regions of interest, and technology-specific preparation for the respective sequencing method.

A NGS workflow typically consists of distinct parts which will be described step by step in the following (Figure 12.1).

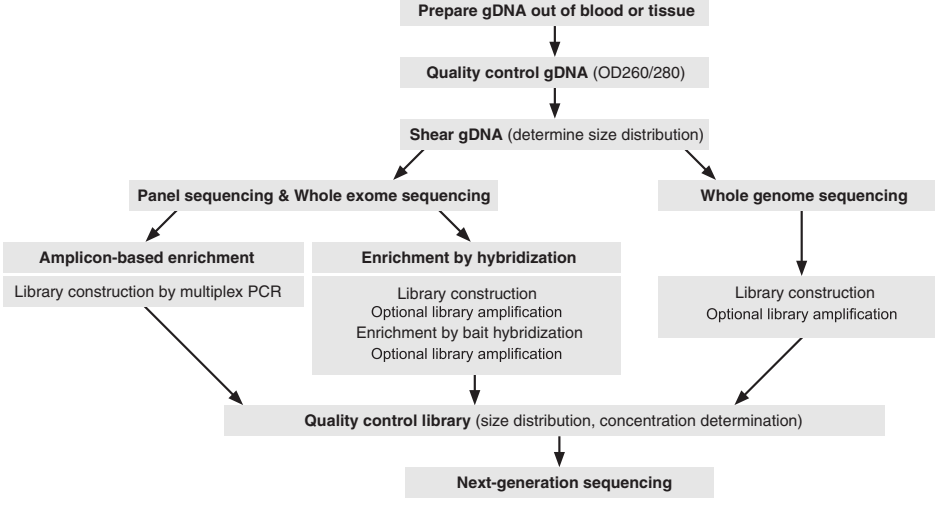


Figure 12.1 Typical workflow for NGS – from blood to sequencing results. For the disease panel and exome sequencing, DNA has to be enriched (amplicon-based or hybridization). This step is omitted in WGS.

12.2.3.1 Preparation of gDNA

As nuclear DNA is identical in all cells of the body, the most easily accessible sources of DNA are usually used for genetic testing, most commonly blood leukocytes. However, hair follicles, buccal swabs, and even Guthrie spots can also be used for the isolation of gDNA for sequencing analyses.

12.2.3.2 Quality Control

For all sequencing applications, DNA samples of high quality and purity are essential. DNA concentration is typically measured photometrically at 260 nm (OD_{260}) and for quality assessment then given as a ratio with concentrations of contaminating proteins (at 280 nm; OD_{280}), and carbohydrates and other impurities such as phenol (both at 230 nm; OD_{230}).

12.2.3.3 Library Preparation and Evaluation

To obtain the desired length of DNA fragments (usually between 100 and 1000 bp), gDNA has to be sheared with the resulting distribution of fragment lengths being as uniform as possible. The shearing can be accomplished using different technologies. The most common shearing method is isothermal sonication. In addition to sonication, other common methods use either endonucleases or a transposase for enzymatic fragmentation. The latter tags the DNA fragments at the same time with library adaptors at both ends.

In all other cases, specific DNA adaptors need to be ligated to both ends of DNA fragments. The sequence and length of these adaptors are platform specific and depend on the sequencing method used. In case of pooling several samples for one sequencing run (which is usually the case in patient diagnostics), the adaptors need to contain patient-specific barcodes in order to later allocate the sequence reads to one specific sample.

12.2.3.4 Enrichment

Even though the basic costs of sequencing have dropped significantly with the emergence of NGS technologies, the cost of sequencing and the effort to clinically evaluate a whole human genome are still prohibitive for most applications. Especially in clinical diagnostics, only specific regions of interest are sequenced. In practice, enrichment is crucial for panel sequencing, where only exons of disease-relevant genes are analyzed, and for exome sequencing, where all exonic regions of the human genome are examined. To enrich the desired regions, several different approaches are commercially available that are either amplicon-based or based on oligonucleotide hybridization (for review, see [6]).

Amplicon-based enrichment strategies rely on the targeted amplification of the desired DNA sequence by PCR. If only a small number of gene-specific amplicons needs to be investigated, it is possible to carry out singleplex PCRs that are afterwards pooled in an equimolar ratio. For the amplification of a larger number of specific sequences, several approaches for target enrichment based on a highly multiplexed, sequence-specific generation of amplicons have been developed. Important factors for these technologies are special algorithms for

primer design enabling simultaneous PCR amplification of multiple target sequences or the combination of PCR with hybridization providing a robust solution for targeting smaller capture regions. Another technology uses picoliter-sized microdroplets that serve as reaction compartments for the PCR. Hundreds to thousands of droplets, each harboring a template molecule, polymerase, dNTPs, and one primer pair for a singleplex PCR reaction, are pooled together in one tube to perform multiple PCR reactions in parallel without direct competition of multiplex PCR primers in the same reagent pool.

Hybridization-based enrichment strategies rely on the explicit design of so-called bait oligonucleotides that specifically hybridize to the desired DNA sequences. Subsequently, hybridized DNA fragments can be separated from the unwanted unhybridized DNA sequences. Hybridization can either be carried out with baits that are immobilized on a microarray chip or in solution where successfully formed bait–target hybrids are captured using streptavidin-coupled magnetic beads recognizing the biotin molecules attached to the bait oligonucleotides. Hybridization-based methods result in a fairly even enrichment of the targeted exons.

12.2.3.5 Quality Control

To evaluate the outcome of the fragmentation steps and to accurately quantify the DNA concentration, sequencing libraries can be analyzed by different methods. The most common instruments are based on capillary electrophoresis and DNA detection by fluorescent dyes, integrated in a compact cartridge or microfluidics chip. The resulting fragment distribution not only allows precise library size and concentration determination, but also detection of small contaminants, such as adapter and primer dimers. Another useful library quantification method is based on quantitative real-time PCR and adapter-specific primers. Here, only library molecules possessing both distinct adaptors on either end are amplified and quantified using a standard dilution series run in parallel. The advantage is that this method only measures such molecules that can subsequently be clonally amplified and sequenced, while fragments with only one adapter or equal adaptors at both ends will not be sequenced and are not detected by real-time PCR. DNA-intercalating fluorescent dyes are also often applied to quantify the concentration of completed libraries using a fluorometer. This method shows a similar dynamic range but less sensitivity compared with real-time PCR. However, both methods cannot discriminate adapter dimer contaminations from library fragments and are therefore often combined with size distribution analysis using, for example, the aforementioned microfluidic chips.

12.2.3.6 Sequencing

The next step consists of the actual sequencing on the particular instruments, with each NGS technology having specific characteristics and caveats (see Section 12.2.2).

12.3 NGS Diagnostics in a Clinical Setting – Comparison Between Genome, Exome, and Panel Diagnostics

12.3.1 Overview

NGS has considerably increased the number of patients who have received a secure clinical diagnosis based on molecular genomic analyses, even if a patient presents with a very ambiguous phenotype. Methodologically, there exist four main approaches for the identification of the molecular cause of rare genetic diseases: WGS, WES, and panel sequencing, all being based on NGS, and single-gene sequencing, based on Sanger sequencing (Figure 12.2).

12.3.2 Clinical Application of WGS

The development of fast and economical high-throughput sequencing technologies very early prompted the idea that in future it should be possible to sequence whole human genomes at a price that should be affordable to all. Years before the advent of NGS technologies, the catchphrase of the "US$1000 genome" was claimed at the very beginning of the twenty-first century, when the scientific community had just witnessed the completion of the first human genome at costs of nearly US$3 billion [7]. With the invention of different NGS technologies, the famous US$1000 genome idea suddenly had the chance to become reality, and has become the ambitious goal of many scientists and biotechnological companies. A little more than a decade later, we are nearly there – raising the expectation that this will open up a new era of predictive and personalized medicine for all patients. While WGS in the first years was (and still is) seen mainly as a tool for academic research projects that deal with the genetic basis of disease, WGS is about to enter clinical diagnostic testing for genetic diseases, slowly but surely.

WGS has a tremendous potential for clinical genetics as it provides information on all 6 billion nucleotides of an individual's DNA. Thus, it is possible to analyze a patient's genome without any bias or prior knowledge of candidate genes for a certain phenotype, whereas traditional genetic tests mainly look at the already known "troublemaker genes." Moreover, it is also possible to include non-coding DNA regions, which are not targeted in regular genetic analyses, although it is known that genetic diseases can be elicited by mutations or copy number variations of regulatory regions situated in intergenic regions. Using long-read platforms, it will also be possible to detect translocations or the exact size of deletions and insertions. The detection of all disease-related genetic variants, independently of the genetic variant's prevalence or frequency, will revolutionize genetic medical diagnostics and will not only permit us to clarify the

	Single Gene Sequencing	Panel Sequencing	Exome Sequencing	Genome Sequencing
Application	Monogenetic syndromes with a strong suspicion for one gene to be causative	Genetically heterogeneous monogenetic syndromes where several genes come into consideration to be causative	Genetically heterogeneous syndromes with an unclear manifestation of symptoms and an unclear suspected diagnosis	Very rare phenotypes of unexplained genetic cause that have found entrance into genetic research
Advantages	• Established method	• Established method • **Time and cost efficient** • High quality of sequencing data due to high coverage • High chance to identify genetic cause of disease, even if phenotype is heterogeneous	• Parallel sequencing of all coding regions • **Identification of novel** disease causing genes possible	• Comprehensive unbiased approach • Analysis of all genomic regions • Detection of mutations outside of coding regions • Detection of translocations, inversions and deletions
Disadvantages	• Sequencing of single genes only • Time consuming and expensive in case more than three genes come into consideration to be causative • Little chance to identify the underlying genetic cause in heterogeneous phenotypes	• Only mutations in genes that have been connected with the disease before **can be identified** • Frequent updates necessary	• **Incidental findings cause** ethical problems • Lower coverage than in panel sequencing • Only detection of mutations in coding regions.	• **Incidental findings cause** ethical problems • Poor quality of sequencing data due to low coverage • Large amount of data • Very expensive

Figure 12.2 Overview of applications and (dis)advantages of the different diagnostic sequencing approaches.

genetic background of an already manifested disease phenotype, but also to make predictions about present and future health risks.

However, although sequencing costs have decreased tremendously during recent years, WGS has not yet found its way into daily clinical practice. One of the main reasons is the high output of generated data as well as their evaluation and interpretation. So far, it has not been possible to deal with the abundance of DNA sequence variants in the human genome in a clinical context. The average human genome contains about 4 million sequence variants that will all need to be detected in a clinical whole-genome analysis. To complicate the situation further, roughly 10 000 of these variants are potentially damaging as they affect the amino acid sequence in encoded proteins [8–11]. However, only a low percentage of these variants is actually pathogenic, which requires a profound and validated prefiltering process to enable a feasible clinical evaluation of variants. To date, the main challenge therefore lies in discriminating the great number of false positives from the true causal allele for a specific phenotype. To transfer WGS into clinical practice, we therefore will first have to overcome the lack of understanding of the enormous number of sequence variants.

Moreover, WGS is still comparatively expensive as the sequencing quality of a US$1000 genome is far from sufficient for diagnostic purposes. In WGS, even small error rates lead to an extremely high number of variations in the resulting sequence that are impossible to assess by a clinician in daily routine practice. For this reason, a whole genome has to be sequenced at a comparatively high coverage to minimize technical errors. However, even then, WGS will never completely cover all regions of a genome. Some difficult-to-sequence areas will always be underrepresented in a whole-genome approach and subsequently will escape clinical analysis or will lead to false-positive or false-negative results.

For these reasons, WGS is currently mainly applied for very rare and selected disease phenotypes that have found their way into genetic research. To accomplish the transition of genome sequencing into routine clinical practice, it is therefore required to set standards for data generation, analysis, interpretation, and reporting. In order to promote this process, a consortium in Boston has initiated an international competition, the so-called CLARITY challenge, to analyze, interpret, and report the results of WES and WGS of selected patients with genetically unsolved phenotypes [12]. The CLARITY challenge was designed to find answers to the most pressing questions and issues about WGS in medical practice:

- What are the best methods to process the massive amounts of data?
- Should the results have an influence on patient care?
- How can information be presented in an understandable and useful way?
- Should unexpected incidental findings be reported?
- How can patient privacy be safeguarded?

In total, three families presenting with different unexplained phenotypes had to be analyzed by the contesting teams. The index patient of Family 1 suffered

from centronuclear myopathy and bilateral sensoneural hearing loss, in Family 2 multiple members suffered from a variety of right-sided cardiac defects, and the index patient of Family 3 suffered from nemaline myopathy. For all families, the phenotype was supposed to have a genetic background; however, no distinct mutation had been detected by standard genetic analyses. In the context of the challenge, all contesting teams were provided with genome sequencing data of the index patients and their participating family members.

It is interesting to note that among the 23 entries of contestants, multiple genes were listed as possibly causative for all families (25 for Family 1, 42 for Family 2, and 29 for Family 3), illustrating the complexity of analyzing and evaluating WGS data. Nonetheless, a consensus of potentially harmful gene variants was achieved for skeletal myopathy (*TTN*; reported as possibly or likely pathogenic by 8/23 groups) and hearing loss (*GJB2*; 6/23) in Family 1, and for cardiac conduction defects in Family 2 (*TRPM4*; 13/23). For Family 3, no convincing pathogenic variants could be identified.

In the end, the challenge highlighted that a holistic view on the genome can help to solve some particular clinical cases. However, it has to be considered that the participating teams spent up to 200 h analysis time per case, which will present a considerable barrier for the widespread use of NGS genome diagnostics in the clinic. It also became obvious that there is not yet a set of best practices that can be applied to the entire "end-to-end" process of genomic measurement and interpretation, while it is clear that genomic medicine will require such consensus and standardization in order to achieve widespread, routine, and reliable clinical use.

In general, one also has to keep in mind that even if WGS will be fully implemented in clinical research, it will not be possible to explain all unclear phenotypes. In many cases, the manifestation of a disease depends on the interaction of different gene variants plus additional environmental factors. It will be a great challenge for the future to get an understanding of these complex traits and to implement diagnostic processes that also take into account risk variants of low penetrance or environmental factors.

12.3.3 Clinical Application of WES

As described above, our ability to generate sequencing data by far outweighs our ability to interpret the data. It is therefore necessary to reasonably constrain the volume of sequencing data generated for the clinical examination of patients. Estimates assume that 85% of the mutations that cause Mendelian diseases are located in the coding sequence of the genome, which in its entirety is called the "exome" [13]. Thus, it stands to reason to narrow down the search for mutations to exonic regions and to carry out a so-called WES approach, which still has a high probability of identifying the underlying genetic cause of a disease.

The advantages of exome sequencing have been deployed recently in numerous studies and have led to the identification of disease-causing gene variants in

many patients. Some of these variants have been found in genes that had been described before in the context of a similar phenotype; others were found in genes that had not previously been connected with the disease. Thus, exome sequencing not only helps to identify already known gene variants in undiagnosed patients, but also contributes to the identification of novel disease-related genes. In order to securely identify a disease causing *de novo* mutation in a patient, it is very helpful to also analyze the patient's parents and then compare the three exomes. This clinically common analysis method is called trio-exome sequencing. The first WES study dedicated to identifying the gene responsible for a rare disease of unknown cause was published in 2010 [14]. Since then, WES has become a success story and it has contributed to the deciphering of the role of more than 150 genes associated with a particular disease in a patient [15]. Today, WES has been successfully implemented in the clinical setting. As it is still a very challenging approach with respect to interpretation of data, it is very important to involve a clinical geneticist at the beginning and also during the analysis process.

Unsolved cases that could neither be diagnosed by a targeted gene by gene analysis nor by a copy number variant analysis are ideal candidates for WES. Recent studies indicate that up to 30% of previously unsolved cases can be diagnosed by WES, which is a considerably high number [16]. This scenario can be exemplified by a 3-year-old girl with developmental delay of unknown cause. By carrying out a trio whole-exome screening, a *SATB2* mutation was identified in the girl who additionally suffered from cleft palate and severely delayed speech development. Previously, the girl had undergone conventional and molecular karyotyping (microarray analysis) as well as targeted analysis for different diseases associated with developmental delay, such as Angelman syndrome, Rett syndrome, or Fragile X syndrome. However, no diagnosis could be established. In the end, a *de novo* nonsense mutation in the *SATB2* gene of the patient (c.715C > T; p.R239X) could be revealed by trio-exome sequencing. In view of the sparse yet available literature, it was concluded that *SATB2* seems to be a key player in craniofacial patterning and cognitive development [17]. This case nicely illustrates the benefits, but also the drawbacks of exome analyses:

Owing to the restricted analysis of only coding regions, a considerably high average coverage of 139-fold could be obtained that would not have been possible in a whole-genome analysis. This successful balancing act between an extensive and unbiased analysis of as many genomic regions as possible and the rational limitation to analyze all exonic regions led to the identification of the genetic cause of a disease that had been unexplained before. Nevertheless, the extensive analysis of a whole exome also required considerable effort in analyzing and evaluating the obtained sequencing data. Even after removing all intronic regions, untranslated regions, and non-synonymous mutations that did not change the protein sequence, 3615 potential single-nucleotide variants were recorded for the index patient. These variants were then further filtered by conducting a trio analysis considering only variants that made sense in the context of parental inheritance. This procedure resulted in 593 mutations that were

suspected of being causative. Only after a further thematic filtering and elimination of false positives, together with use of the available literature, could the particular variant in *SATB2* be pinned down.

Due to the great efforts that are necessary to identify the causative variant out of the huge number of variants, even in cases with a clear clinical manifestation, the more reasonable route in clinical diagnostics is to further narrow down the potential diagnoses by a clinical expert, and by this also to narrow down the number of potentially relevant genes that need to be analyzed and reported.

12.3.4 Clinical Application of Diagnostic Panels

Today, the most reasonable and cost-efficient method of all high-throughput sequencing approaches is a so-called diagnostic panel approach. A diagnostic panel is the targeted and simultaneous analysis of a list of all known genes that have been described as the cause of a particular disease. Panel sequencing is especially applied in genetically heterogeneous diseases, where similar well-characterized phenotypes are caused by mutations in more than one candidate gene. Diagnostic panels are considerably different from the rather scientific and explorative approaches of genome and exome sequencing as exclusively candidate genes are analyzed that have been associated with the relevant disease before. Using one of the enrichment methods described in Section 12.2.3.4, it is possible to sequence only a distinct set of candidate genes very specifically. Due to the parallel sequencing of up to several hundred genes, the genetic cause of a disease is found significantly more frequently than with former single-gene sequencing, and mutations in the genes analyzed can be identified or excluded with a high accuracy. The high accuracy is the most important advantage with respect to WGS or WES and is achieved by the much smaller sequencing volume, which in turn leads to a much higher coverage. At the same time, the method is very time- and cost-efficient. By analyzing only relevant genes, the amount of data generated can be reduced significantly, allowing us to increase the number of patients that can be sequenced per run. Of course, the restricted nature of the analysis has some drawbacks. As only genes that have been known before to be disease-relevant are part of a specific panel, it is possible that the disease-causing gene for a particular case may not have been included in the panel. Therefore, the approach requires regular updates to include all genes currently known to be connected with the disease to minimize the chance of overlooking the disease-causing mutation in a patient. However, with the current state of knowledge and technology, the advantages of the methods clearly outweigh the disadvantages, which is why diagnostic panels have become a success story.

One of the first diagnostic panels that was successfully implemented in clinical use was a panel for the genetic analysis of retinal dystrophies. Retinal dystrophy is a chronic and disabling impairment of visual function and, depending on the subtype and course of disease, may cause blindness. Retinal dystrophy is

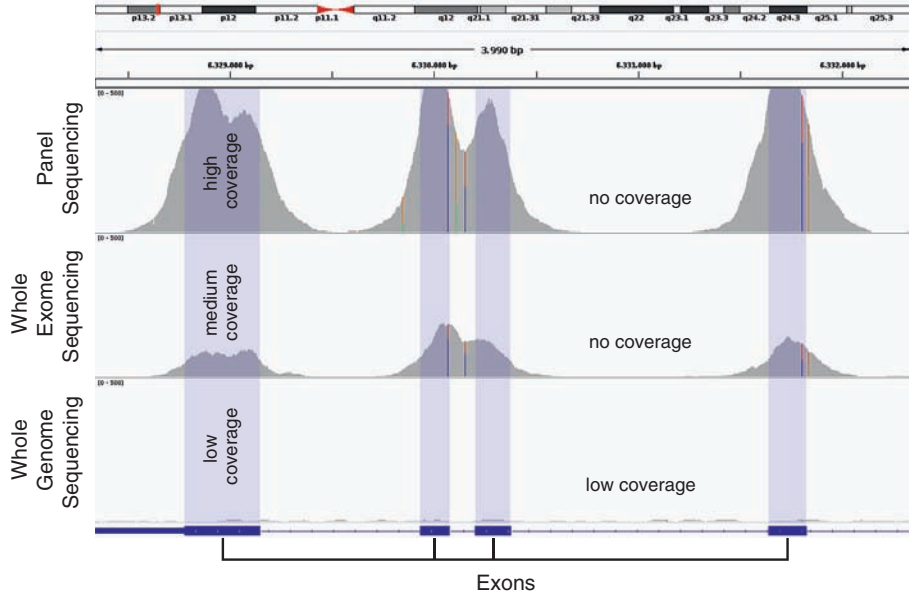


Figure 12.3 Comparison of sequencing coverage for the first four exons of *AIPL1* – a gene involved in the etiopathogenesis of hereditary retinopathies. Exonic regions are depicted at the bottom of the figure by thick bars. Thin lines in between symbolize intronic regions. Coverage is represented by gray curves. In exonic regions, coverage is highest in the panel sequencing approach due to the very targeted enrichment. Coverage is substantially lower in the WES approach. In the WGS approach, the exonic coverage is far from sufficient for clinical evaluation of variants. In contrast, WGS also delivers information about intronic or intergenic regions that are not addressed by WES or panel sequencing.

genetically very heterogenic, which is illustrated by the fact that mutations in more than 100 genes are known to be causative for retinal dystrophies. Due to this enormous genetic heterogeneity, it is a challenging task to identify causative mutations, leaving most patients undiagnosed. Approaches to apply Sanger sequencing analyses to all known retinal dystrophy disease genes led to quite low detection rates coupled with enormous expenses in terms of time and money as hundreds of exons had to be sequenced one by one. Sequencing on an NGS platform allowed the parallel analysis of more than 100 genes (with 1650 exons in total) and yielded an average coverage of more than 750-fold. The extremely high coverage also reduced significantly the error rate, compensating for the only minor disadvantage of NGS compared with Sanger sequencing. It is important to note that a comparative coverage cannot be reached by exome and genome analyses. See Figure 12.3. In the case of retinal dystrophies, the use of the diagnostic NGS panel pipeline led to mutation detection rates of 55–80%, and at the same time was a lot faster and cheaper than Sanger sequencing [18]. This approach has now been adopted by several other groups and diagnostic laboratories, and has enabled clearly determined diagnoses for many patients.

Another clinical scenario where NGS seems particularly appropriate due to the highly heterogeneous nature of the genetic background is epileptic disorders. So far, only 1–2% of all epilepsies have been identified to be monogenetic; the majority are considered to be polygenic or to have a complex genetic background. Very often, the combination of several less-severe genetic variants leads to the development of an epileptic phenotype. Therefore, the establishment of precise clinical diagnosis is a challenging task leading to slightly lower detection rates for the complex field of epilepsies. Nevertheless, with an epilepsy panel containing more than 450 genes, it was possible to securely identify the genetic cause in more than 21% of all analyzed patients with a further 28% of cases being possibly solved. Across the cohort of 800 patients, 69 genes were identified as causative only once, emphasizing the advantage of diagnostic panels for very rare conditions [19]. Especially in the case of epilepsy, a clinical diagnosis of the genetic background is very valuable due to its high relevance for therapy and treatment, and can even guide pharmacotherapy. This is illustrated by the example of a male patient suffering from an unclear dysmorphic retardation syndrome associated with epileptic seizures. The patient had been analyzed for a variety of genetic defects by sequencing the *SCN1A* gene and the *SCL2A1* gene, by carrying out a methylation test for Angelman syndrome, and by array comparative genomic hybridization to exclude chromosomal microaberrations. Only at the age of 22 was the patient analyzed with our diagnostic epilepsy panel in order to finally uncover the genetic basis of the epilepsy. The panel diagnostic revealed a *de novo* hemizygous missense mutation in exon 3 of the *SMS* gene that encodes for spermine synthase. A subsequent biochemical analysis confirmed the extremely reduced activity of spermine synthase. Hence, after many years of fruitless attempts to reveal the genetic cause of the disorder, the use of a diagnostic panel finally led to the diagnosis of Snyder–Robinson syndrome. Despite the very late diagnosis, the patient benefitted from a diagnosis-related change in therapy when he was given an AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid) receptor antagonist that had previously been proven effective in Snyder–Robinson patients. Apart from the therapeutic consequences, an earlier confirmed genetic diagnosis also would have helped in the genetic counseling of the parents about a relatively low recurrence risk in further children and would have enabled targeted prenatal diagnostics in future pregnancies [20].

An early and clearly determined diagnosis is also beneficial for many neurodegenerative diseases (NDDs). Particularly in early-onset NDD, a genetic component can be found in a considerable number of patients. As several genes have been implicated in the pathogenesis of different types of neurodegeneration and each of these genes only accounts for a small fraction of cases, it was crucial to develop an efficient tool for investigating all genes simultaneously. Using a NDD panel containing 242 genes, it was possible to detect known NDD-causing mutations in nearly 20% of all patients analyzed. For many different NDDs, early symptoms of a specific syndrome are not very clear and

commonly show a strong overlap with similar syndromes. This is illustrated by the case of a 36-year-old woman who was clinically diagnosed for dystonia. An initial analysis with a dystonia subpanel did not result in any genetic diagnosis. However, when the complete NDD panel was applied to analyze the patient's DNA, two most likely compound heterozygous mutations in the *PARK2* gene were detected. Mutations in *PARK2* lead to Parkinson's disease 2 – an autosomal recessive disease with a juvenile onset, where dystonia is seen as an early symptom in about 42% of patients. Thus, this diagnosis helped to initiate an appropriate treatment for Parkinson's disease at a very early stage of the disease when no other characteristic Parkinson's symptoms were present (CeGaT, unpublished data).

As already suggested by the name, symptoms of NDDs are caused by the degeneration of neurons that are lost irrecoverably. Therefore, with the appearance of the first symptoms in a patient, the brain or the neurons are already irretrievably damaged, which is why it is desirable to start treatment already before the onset of the disease. It is still a long way off, but the goal in familial NDDs should be to identify carriers of a harmful gene variant early to impede damage to neurons and thus the onset of the disease by initiating a preventive treatment. A very good example of this approach is DIAN (Dominantly Inherited Alzheimer Network; www.dian-info.org) – an international research partnership of leading scientists determined to understand a rare form of Alzheimer's disease that is caused by a gene mutation. The program is funded by a multiple-year research grant from the National Institute on Aging, and currently involves 13 outstanding research institutions in the United States, United Kingdom, Germany, and Australia.

In summary, panel sequencing has revolutionized human genetic diagnostics, as it offers an effective and affordable opportunity to obtain a fast and secure diagnosis for many so-far undiagnosed patients. Panel sequencing approaches will be optimized by further increasing quality whilst simultaneously reducing costs and time. They will also be applied to further disease areas and will be a major genetic testing method for many hereditary diseases.

12.4 Conclusion and Outlook

The application of NGS methods in the clinical setting provides a huge benefit for patients and their families, as it becomes possible to identify the molecular cause of genetic diseases and to provide affected families with a secure clinical diagnosis. In many cases, families have been seeking medical advice for several years, have undergone multiple genetic testing without any success, and are looking back to a long history of misguided therapy consultations. An early identification of the genetic cause of a disease with an ambiguous phenotype facilitates an early therapeutic intervention and provides the basis for new therapeutic strategies in

the long-term. By knowing the genetic cause of the disease, the most effective drug can be prescribed and ineffective drugs as well as related side-effects can be avoided. Moreover, the novel analysis methods will help to reduce the costs for the healthcare system by avoiding expensive and time-consuming gene-by-gene testing or other invasive and non-invasive diagnostics.

In recent years, NGS has successfully developed from a research tool to a diagnostic platform, and for most applications has replaced traditional and expensive Sanger sequencing diagnostics. However, although WGS and WES are powerful unbiased approaches to unravel the underlying cause of genetic diseases, they have not yet found their way into daily clinical routine practice. With the current technologies and our sparse knowledge of the clinical relevance of the vast number of variants that occur in the human genome, panel sequencing is still the preferred diagnostic method. Prospectively, it may be expected, though, that panel sequencing will be compensated by the more comprehensive and unbiased WES and WGS approaches in the future.

However, even advanced large-scale analysis methods such as WGS, WES, or panel sequencing will have their limitations in identifying the underlying cause of a substantial number of diseases as most phenotypes are also influenced by various environmental factors that are not reflected in the patient's genomic information. At least today, we are not able to entirely understand the interaction of many different partly causative components of complex multifactorial diseases.

The vast amount of data that is generated by genomic large-scale analyses also poses another fundamental problem: even if we are able to deal with the human genome in a technical way, we are not able to deal with its interpretation. Each analysis will result in an innumerable number of variants and determining the potential pathogenic nature of these variants will be very challenging. Therefore, there is a huge need for better bioinformatic systems that allow a reasonable classification of all found variants to enable clinicians to provide a clinical report in a reasonable timeframe.

Additionally, each analysis will produce tremendous amounts of sequencing data of genomic regions that are not relevant to the analyzed phenotype. It therefore will be crucial to define how incidental findings not associated with the initial clinical problem should be handled. Should patients be informed about their carrier status of recessive disorders or about their predisposition to develop an incurable late-onset disease? Who will have access to the sequencing data generated? As long as we have not found satisfactory answers to these urgent questions, it will be difficult to implement WGS in daily clinical use. By then, laboratories might want to keep carrying out panel diagnostics or to mask unrelated data in order to not come into ethical conflict.

Nevertheless, we are heading towards personalized medicine that is based on the patient's genome. NGS offers undreamt of possibilities and promises, but also poses problems that have to be solved. However, it is certain that, in the end, our society will not be willing to relinquish the insights and knowledge derived from genetic diagnostics.

References

1 Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, **74**, 5463–5467.

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

3 Metzker, M.L. (2010) Sequencing technologies – the next generation. *Nature Reviews Genetics*, **11** (1), 31–46.

4 Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., and Law, M. (2012) Comparison of next-generation sequencing systems. *Journal of Biomedicine & Biotechnology*, **2012**, 251364.

5 Ross, M.G., Russ, C., Costello, M., Hollinger, A., Lennon, N.J., Hegarty, R., Nusbaum, C., and Jaffe, D.B. (2013) Characterizing and measuring bias in sequence data. *Genome Biology*, **14**, R51.

6 Mamanova, L., Coffey, A.J., Scott, C.E., Kozarewa, I., Turner, E.H., Kumar, A., Howard, E., Shendure, J., and Turner, D.J. (2010) Target-enrichment strategies for next-generation sequencing. *Nature Methods*, **7** (2), 111–118.

7 Mardis, E.R. (2006) Anticipating the $1,000 genome. *Genome Biology*, **7** (7), 112.

8 Ng, P.C., Levy, S., Huang, J., Stockwell, T. B., Walenz, B.P., Li, K., Axelrod, N., Busam, D.A., Strausberg, R.L., and Venter, J.C. (2008) Genetic variation in an individual human exome. *PLoS Genetics*, **4** (8), e1000160.

9 Wang, J., Wang, W., Li, R., Li, Y., Tian, G., Goodman, L., Fan, W., Zhang, J., Li, J., Zhang, J. *et al.* (2008). The diploid genome sequence of an Asian individual. *Nature*, **456** (7218), 60–65.

10 Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.J., Makhijani, V., Roth, G.T. *et al.* (2008) The complete genome of an individual by massively parallel DNA sequencing. *Nature*, **452** (7189), 872–876.

11 Marian, A.J. (2012) Challenges in medical applications of whole exome/genome sequencing discoveries. *Trends in Cardiovascular Medicine*, **22** (8), 219–223.

12 Brownstein, C.A., Beggs, A.H., Homer, N., Merriman, B., Yu, T.W., Flannery, K.C., DeChene, E.T., Towne, M.M., Savage, S.K., Price, E.N. *et al.* (2014) An international effort towards developing standards for best practices in analysis, interpretation and reporting of clinical genome sequencing results in the CLARITY Challenge. *Gene Biology*, **15**, R53.

13 Majewski, J., Schwartzentruber, J., Lalonde, E., Montpetit, A., and Jabado, N. (2011) What can exome sequencing do for you? *Journal of Medical Genetics*, **48** (9), 580–589.

14 Ng, S.B., Buckingham, K.J., Lee, C., Bigham, A.W., Tabor, H.K., Dent, K.M., Huff, C.D., Shannon, P.T., Jabs, E.W., Nickerson, D.A. *et al.* (2010) Exome sequencing identifies the cause of a mendelian disorder. *Nature Genetics*, **42** (1), 30–35.

15 Rabbani, B., Tekin, M., and Mahdieh, N. (2014) The promise of whole-exome sequencing in medical genetics. *Journal of Human Genetics*, **59**, 5–15.

16 Jacob, H.J. (2013) Next-generation sequencing for clinical diagnostics. *New England Journal of Medicine*, **369** (16), 1557–1558.

17 Döcker, D., Schubach, M., Menzel, M., Munz, M., Spaich, C., Biskup, S., and Bartholdi, D. (2013) Further delineation of the SATB2 phenotype. *European Journal of Human Genetics*, doi: 10.1038/ejhg.2013.280 [Epub ahead of print].

18 Glöckle, N., Kohl, S., Mohr, J., Scheurenbrand, T., Sprecher, A., Weisschuh, N., Bernd, A., Rudolph, G., Schubach, M., Poloschek, C. *et al.* (2014) Panel-based next generation sequencing as a reliable and efficient technique to detect mutations in unselected patients with

retinal dystrophies. *European Journal of Human Genetics*, **22** (1), 99–104.

19 Steiner, I., Döcker, M., Russ, A.C., Battke, F., Lemke, J., Lerche, H., Biskup, S., and Hörtnagel, K. (2014) Comprehensive NGS based diagnostics in over 800 patients with epileptic disorders, presented at the 11th European Congress on Epileptology, Stockholm.

20 Steiner, I. and Lorenz, R. (2013) Ein 22-jähriger Patient mit ursächlich ungeklärter schwerer Mehrfachbehinderung und Epilepsie. *Kinderärztliche Praxis*, **84**, 374–378.

13
Analysis of Nucleic Acids in Single Cells

Stefan Kirsch, Bernhard Polzer, and Christoph A. Klein

13.1
Introduction

In recent years, various technologies that allow detection and even isolation of single cells have been developed and are now being incorporated in clinical diagnostic settings, such as the US Food and Drug Administration (FDA)-approved CellSearch System (Veridex) for the detection of circulating tumor cells (CTCs) in the blood of cancer patients. However, for a deeper understanding, reliable molecular methods for single-cell analysis are paramount to overcome the mere counting of tumor cells, especially as recent studies indicate the clinical utility of CTCs to circumvent tissue biopsies of metastatic tumors in patients suffering from systemic disease. Moreover, characterization of CTCs holds the promise for companion diagnostics, in particular the genomic *TMPRSS2* (transmembrane serine protease 2)–*ETS* gene family member rearrangements and the activation status of the androgen receptor pathway in prostate cancer patients. Both potential biomarkers might help to select patients most likely to benefit from targeted agents.

13.2
Isolating Single Cells

In the upcoming era of genome-based diagnostics, reliable technologies for single-cell isolation will be the first key factor in a long line of methods to constantly develop our abilities for personalized medicine. Whatever methods will be used, the ultimate goal is to singularize contamination-free cells into solitary reaction chambers for lysis and subsequent downstream analyses. The first step towards isolation of individual cells is the reliable identification of the cells of interest by morphological characteristics or fluorescence-mediated methods. To facilitate this, solid tissues have to be disaggregated and separated into a suspension of individual cells for all single-cell isolation methods with the exception of

Nucleic Acids as Molecular Diagnostics, First Edition. Edited by Andreas Keller and Eckart Meese.

laser-capture microdissection [1–3]. This method has the invaluable advantage of retaining spatial information, but is currently cumbersome and prone to errors (e.g., copurification of cellular constituents from neighboring cells or partial nuclei dissection). Therefore, the most widely used single-cell isolation methods are semiautomated micromanipulation devices [4–8], including the three-dimensional dynamic dielectrophoretic trap systems associated either with optical tweezers [9] or complementary metal oxide semiconductor (CMOS) technology [10] and fluorescence-activated cell sorting (FACS) [11]. The encouraging developments in microfluidic technology have recently resulted in new devices enabling single-cell isolation by controlled liquid flow in nanoliter-sized reaction chambers [12,13]. Continuous improvement of these microfluidic devices will lead to miniaturized low-cost, high-throughput platforms capable of isolating numerous viable cells with high precision.

13.3
Looking at the DNA of a Single Cancer Cell

For molecular analysis of single cells, the small amount of DNA (6–7 pg in a diploid single human cell [14]) is an obvious obstacle. Although single-cell analysis is feasible with targeted polymerase chain reaction (PCR) or fluorescence *in situ* hybridization (FISH), those assays are only suited for a limited number of targets. However, in the era of personalized therapy, comprehensive analysis is important to deliver the complete picture of the mutational single-cell landscape. To allow genome-wide studies, the genomic DNA of a single cell has to be amplified. For this, whole-genome amplification (WGA) has to overcome three major hurdles: (i) to achieve high genomic coverage, all 3×10^9 nucleotides comprising the human genome have to be amplified; (ii) to allow reliable quantification of copy number variation, all regions of the genome have to be amplified homogeneously; and (iii) to assess sequence variation, the artificial loss of one or even both inherited gene copies (maternal and paternal) has to be avoided.

Researchers have addressed this issue since the early 1990s and several principles have been utilized for WGA of a few or even single cells (Table 13.1). One of the first methods applied on single cells uses degenerate oligonucleotide primers (DOPs) for PCR amplification [15], and relies on a mixture of random core nucleotides flanked by defined 5′ and 3′ sequences. These primers anneal at approximately 1 million sites throughout the DNA template at low temperature and DOP-PCR has been used successfully in preimplantation genetic diagnosis [16,17]. A second principle, which uses highly degenerated oligonucleotides consisting of 15 random bases, is primer extension pre-amplification (PEP)-PCR [18–20]. Both methods using degenerate oligonucleotides are limited by the fact that amplification conditions for the multiple variations of primer combinations (4096 for DOP-PCR and approximately 10^9 for PEP-PCR) vary considerably. Nevertheless, an improved procedure (i-PEP) was successfully used for single-cell analysis of defined genomic loci [21]. However, genomic coverage and allelic

Table 13.1 Methods for WGA of single cells.

	PEP-PCR	DOP-PCR	Displacement DOP-PCR	MDA	LA-PCR	TLAD	MALBAC
Primers	random	random	random and specific	random	specific	specific	specific
Amplification	exponential	exponential	exponential	exponential	exponential	linear after reverse transcription	linear and exponential
Whole-genome representation of single cells	not reported	aCGH (8.3 Mb) [23]; WGS (6%) [24]	aCGH (1.2 Mb) [25]; WGS (36%) [26]	aCGH (34 Mb) [27]; WGS (77%) [26]	aCGH (56 kb) [28]	not reported	WGS (93%) [8]
Allelic drop out in single cells	not reported	30%[a)]	<10%[a)]	10–31% [29,30]	5–10% [31]	not reported	<10%[a)]
Commercially available kits	no	GenomePlex (Sigma)	PicoPLEX (Rubicon Genomics)	RepliG (Qiagen); GenomiPhi (GE Healthcare)	Ampli1 WGA (Silicon Biosystems)	no	single-cell WGA kit (Yikon Genomics)

a) Information as provided by the manufacturer of the commercial kit.

representation are considerably reduced in samples with a low input of genomic DNA (i.e., less than 100 cells) [22].

To minimize the intrinsic sequence amplification error rate of *Taq* polymerase, other groups utilized the highly processive Phi29 DNA polymerase, which possesses proofreading activity [32]. In the approach named multiple strand displacement amplification (MDA), random hexamers bind to denatured DNA followed by an isothermal amplification [33,34], which generates relatively long DNA fragments and therefore relatively high genomic coverage in WGA [26]. Unfortunately, this advantage of long DNA fragments requires high-molecular-weight genomic DNA, and is therefore mainly suitable for analysis of freshly isolated cells and less for clinical samples, for which fixation and transport logistics may lead to considerable degeneration of DNA. Additionally, when used on complex genomes, MDA results in a significant rate of allele dropout [35].

To overcome the limitations of the previously described methods, we developed an adapter-linker PCR (LA-PCR) strategy that homogenously amplifies the whole genome [36] and enabled for the first time reproducible application of WGA of single cells isolated from patient samples. The method relies on a highly specific restriction digest with *Mse*I (restriction site TTAA) of double-stranded DNA, which leads to an average fragment size of 162 bp and is followed by subsequent ligation of one universal adapter to all fragments. As there is no random component in the procedure, next to WGA (e.g., by comparative genomic hybridization (CGH)) it is possible to easily design specific targeted downstream assays that can easily be utilized for validation of the WGA, and somatic mutations as well as gene amplifications of therapeutic targets can be assessed [31,37].

Recently, two new principles have been introduced to single-cell WGA. A T7-based linear amplification (TLAD) method has been adopted for WGA [38]. In principle, TLAD is resistant to sequence- and length-dependent bias, but as the protocol is rather time-consuming it has not been widely used so far [39]. A second promising approach is the multiple-annealing and looping-based amplification cycles (MALBAC) method [8,40] using a quasilinear amplification approach. Briefly, the method utilizes random primers that hybridize to single-stranded template DNA. A polymerase with strand-displacement activity generates 0.5- to 1.5-kb long semiamplicons that are melted off and then processed to full amplicons with complementary ends. These form a loop and are therefore removed from the template pool for the following cycles, leading to a linear amplification of all such amplicons. Finally, the method switches to an exponential amplification of full amplicons by PCR reporting a high genomic coverage of a single-cell genome of 93%. A drawback of the method, however, is the high allelic dropout indicated by the fact that one third of the expected single nucleotide polymorphisms in a cell have not been called [8].

13.4 Molecular DNA Analysis in Single Cells

In principle, all types of genomic analyses applicable to millions of cells can also be used for single-cell analysis. However, one always has to keep in mind the

drawbacks of WGA methods introduced by amplification of a single-cell genome in general (i.e., allelic dropout, biased genome coverage, and sequence errors). As described above, different WGA protocols may have disadvantages over one another and specific intermediate steps frequently have to be introduced to optimize the use of single-cell WGA products to established workflows.

This is especially true for targeted analysis of specific genes of interest, as in classical Sanger sequencing. As many WGA methods rely on a random fragmentation or even random primers for amplification, the design of specific primers for the target of interest is turning into a game of pure chance. For these types of analyses, WGA methods with known fragmentation sites and WGA primers are best suited, as is the case for LA-PCR. By designing primers that are located on the same *Mse*I restriction fragment, allele dropout of single diploid cells could be determined for multiple loci as 5–10% [31]. The same assay was applied on disseminated cancer cells (DCC) of breast cancer patients, showing loss of heterozygosity (LOH) in early disseminated DCCs [31,41]. Additionally, somatic *TP53* mutations in DCCs of patients with various cancers were detected, while 46 diploid cells of healthy donors did not show any mutation in seven analyzed *TP53* exons [37].

The same limitations apply to single-cell analysis by quantitative PCR (qPCR). As WGA methods will show variation between individual samples, quantification should not rely on measurements of single target and single reference loci, but rather use multiple qPCR assays for the target of interest as well as multiple PCRs on reference genes. Additionally, the resulting target/reference ratios should be compared with a cohort of normal diploid cells. By applying these criteria on single WGA products of DCCs obtained by LA-PCR, we could detect human epidermal growth factor receptor (*HER2*) gene copy amplifications in single breast DCCs [31]. In a prospective study on esophageal cancer, we applied the same assay and found that *HER2* gain in a single DCC was the most important risk factor for poor outcome, while *HER2* amplification at the primary site was irrelevant [42].

The first genome-wide studies on single cells were performed by metaphase CGH (mCGH), again mainly with LA-PCR utilized as the WGA method. In a landmark study, we found that several DCCs isolated from the bone marrow of one individual with manifest metastases (stage M1) displayed a high degree of similarity in their chromosomal imbalances. In contrast, when patients were clinically free of metastases (i.e., in clinical stage M0), we observed unexpected genomic heterogeneity of tumor cells from the same individual in most patients [37]. This heterogeneity between individual DCCs and between DCCs and their corresponding primary tumor could be shown for breast cancer [31,41], esophageal cancer [42], and prostate cancer [43]. One drawback of mCGH is the low resolution, limiting the detection to chromosomal gains and losses of 10–20 Mb, which can be overcome by applying array CGH (aCGH) technologies. While first studies using bacterial artificial chromosome (BAC) arrays did not significantly improve the resolution [23,27,44], application of oligonucleotide arrays raised expectations. Most patient-derived data is available for single cells amplified by

GenomePlex (Sigma) and hybridization on Agilent human CGH arrays. This workflow was used for detection of copy number aberrations in single breast cancer DCCs [45] and single CTCs of colorectal cancer patients [46]. Recent application of Agilent CGH arrays after single-cell MDA and LA-PCR reliably detected chromosomal changes of 1 Mb and below in single cells [25,28,47]. Additionally, compared with GenomePlex, noise could be reduced significantly [28]. However, the applicability of these aCGH workflows on a large number of patient samples (e.g., DCCs or CTCs) still needs to be demonstrated.

Over the past few years, whole-genome sequencing (WGS) approaches have been applied to single-cell WGA products. This type of analysis adds another dimension to the quality of obtainable data, as nucleotide sequence and copy number changes can be assessed. However, error rates of polymerases during the WGA procedure reduce the quality of the data, and, especially in PCR-based WGA methods, the introduced amplification bias impedes quantification and reduces genomic coverage (reviewed in [48,49]). The amplification bias is the most likely reason why recently developed MALBAC, which starts with a linear amplification step, has so far demonstrated the highest genomic coverage (93%) [8], compared with MDA-based (77%) [26] and DOP-PCR-based WGA methods (6–36%) [24,26]. In principle, single-molecule sequencing technologies could reduce amplification bias as they do not depend on WGA prior to sequencing. However, these technologies themselves have high error rates and low sequencing efficiency [50,51].

For diagnostic purposes, analyzing the whole genome is still too expensive and in the end not essential. Targeting a subset by genomic enrichment for sequences of interest would enable multiplexing of patient samples, and lead to cost efficiency and higher sensitivity for the targeted sequences. These subsets could be composed of the whole exome [51,52] or specific genes of interest (i.e., specific disease) [53,54]. This strategy has recently been applied on single CTCs of colorectal cancer patients after GenomePlex, uncovering mutations in known oncogenes *APC*, *KRAS*, or *PIK3CA* [46]. Combining such a sequence-enrichment approach with a non-random WGA method with low allelic dropout (e.g., LA-PCR) could open the door to single-cell diagnostics.

13.5 Approaches to Analyze RNA of a Single Cell

Although genomic DNA is more stable than RNA, and therefore better suited for clinical use, especially for sample transport and logistics, important information about the phenotype and functional characteristics of cells cannot be extracted from genomic DNA.

Transcriptome analysis collects information on qualitative and quantitative gene expression of a tissue sample or cell suspensions under certain conditions at a certain time point. Normally gene expression-based investigations require RNA amounts far beyond the RNA content of a single cell. However, comprehensive

analysis of the transcriptome at the single-cell level has been increasingly perceived as the key step to understanding cell diversity and intercellular communication driving developmental processes. Yet, technological developments to analyze a few molecules have only advanced over the last 25 years to the point that they could provide sufficiently accurate measurements of target mRNA levels in single cells.

Single-cell transcriptome technologies can be divided in two categories: direct imaging-based methods, which have the drawback that only a few genes can be analyzed simultaneously, and indirect cDNA synthesis/preamplification-dependent methods, which provide high-dimensional information of gene expression.

The first group comprises the real-time transcription detection systems [55–58], single-molecule RNA-FISH [59–63], and *in situ* padlock probes followed by rolling-circle amplification [64]. The M2 and the living microarray system enable gene expression to be traced over time in living cells; however, the M2 system carries the decisive disadvantage of cloning a bacteriophage M2 stem–loop cassette in each singular gene to be analyzed. The development of the living microarray substantially improved real-time transcription visualization by using a reverse transfection protocol to introduce fluorescent reporter plasmids into independent clusters of monolayer cells to monitor promoter activity in single cells. Recently, single-molecule RNA-FISH has become increasingly important due to its sensitivity in visualizing transcripts in single cells. The possibility to scale-up this method for high-throughput applications with automated computational detection allows scanning the transcription levels of a gene of interest in an unlimited number of cells. The latest techniques [62,63] involve combinatorial barcoding of transcripts through specifically labeled oligonucleotide probes unified with super-resolution imaging. Padlock probes are circularizable oligonucleotide probes that detect short mRNA sequences with single-base resolution at the site of ligation that can be multiplied by rolling-circle amplification. Recently, the versatility of this method was shown by its potential to combine detection of individual mRNA molecules with the detection of protein complexes or post-translational modifications [65]. All these imaging-based techniques represent direct approaches allowing the *in situ* detection of transcript levels of genes in hundreds of cells in intact tissues, thereby preserving the spatial information in the cellular and intercellular context.

The second group comprises reverse transcription (RT)-quantitative PCR [66–68], gene expression profiling [69–76], and RNA-Seq [74,77–80]. The common basis of all these methods is the simultaneous multiplication of a multitude of mRNA species with the well-established RT-PCR technology allowing generation of sufficient starting material for downstream analyses. The fundamental question that needed to be adequately addressed in quantitative single-cell analyses was the influence of the low mRNA amount present in single cells and the highly variable mRNA levels between individual cells on experimental variations. Studies thoroughly investigating experimental noise unraveled the contribution of biological variance and technical variability in quantitative single-cell studies, and provided guidelines to successfully conduct quantifications sensitive enough

to reliably detect single transcripts in one cell [81,82]. Nevertheless, these studies clearly showed that most technical variability is introduced by sample preparation and RT, followed by preamplification – the essential steps to generate sufficient starting material for downstream analyses. This might be explained by the stochastic sampling of small numbers of molecules and stochastic events in the first few cycles of PCR, which are exponentially amplified leading to large quantification errors. Two major principles are distinguished for the amplification of low amounts of mRNA: (i) linear amplification, which is achieved by *in vitro* transcription with T7 RNA polymerase [83], and (ii) exponential PCR amplification, which is part of template switching PCR [84,85], lone-linker PCR [86], and global PCR amplification [87]. In theory, linear amplification should be less biased than exponential amplification, but it has the inherent disadvantage of a stronger 3′ positional bias. As a consequence, two methods were developed applying both amplification principles in a sequential manner [70,73].

To achieve reproducible and reliable gene expression results in single-cell RT-qPCR, each experimental step needs to be optimized and validated very thoroughly in order to minimize transcript losses. Optimized assays should guarantee high sensitivity and provide target specificity over 50 cycles without the generation of unspecific products. The lack of transcript detection with a well-established RT-qPCR assay in a given cell does not necessarily indicate that the gene of interest is not expressed in this cell. The number of cDNA target molecules might be below the detection level of the assay. Additionally, one must be aware of the fact that the RNA detection level is unknown as RT efficiency is strongly dependent on the secondary and tertiary structure of an RNA molecule. Optimal performance of RT-qPCR assays can only be achieved by parallel analysis of serial dilutions of full-length mRNAs of the transcription unit under investigation.

Recent progress in microtranscriptomic technologies [68,88–90] has allowed multiplex RT-qPCR assays in miniaturized volumes for extensive profiling of single cells. Such dynamic arrays for microfluidic real-time PCR allow reproducible and precise handling of liquids on a nanoliter scale, thereby enabling the uniform distribution of reaction components into individual reaction chambers. Significant technological improvement was recently achieved by the introduction of a high-throughput platform performing automated cell isolation, cell lysis, RT, and sequence target preamplification. Direct coupling of the microfluidic single-cell trapping and processing platform to the qPCR platform enables high-precision RT-qPCR experiments to be performed on thousands of individual cells [13].

A semiautomated gene expression system allows amplification-free direct digital multiplexed target profiling of single mRNA molecules of interest [91]. Each target gene is detected by a combination of capture and reporter probe, with the latter carrying a unique color code at the 5′ end. Counting the number of times this color-coded barcode is detected directly correlates with the gene expression level of the target gene. The recent introduction of a single-cell gene expression assay has expanded the utilization options of the platform, but

requires an intermediate RT/cDNA preamplification step, exposing the method to the same technical variability as all other indirect methods.

Gene expression analysis at the single-cell level on high-density oligonucleotide microarrays has the same underlying constraints with respect to the amplification of the starting material as RT-qPCR. However, microarray analyses require relatively large amounts of homogeneously amplified cDNA libraries. Microarray analyses have demonstrated the reproducible and representative amplification of original mRNA samples serially diluted to the single-cell level [70,71,73]. It was concluded that genes expressed at several hundred copies per cell can be quantified, whereas those genes expressed in fewer than 20 copies are less accurately amplified. According to the generation of the single-cell cDNA libraries to be sequenced, RNA-Seq approaches can be divided in three categories: *in vitro* transcription (IVT), homopolymer tailing, and template-switching approaches. CEL-Seq [77] is the only technique based on IVT-generated sequencing libraries having the advantage of less-biased amplification and strand-specific information, while being biased towards the 3′ end of genes. A similar positional bias, albeit less pronounced, is an intrinsic property of the homopolymer-tailing methods of Tang *et al.* [80] and Quartz-Seq [74]. Both methods enable the cDNA strand to be amplified via PCR by addition of a homopolymer tail to the first-strand cDNA. The template-switching approaches STRT [78] and SMART-Seq [79] use the ability of Moloney murine leukemia reverse transcriptases to add a short anchor of cytosines at the 5′ end of the synthesized first-strand cDNA. The STRT approach thereby preserves the strand-specific information. As 5′-capped RNA is the preferred target of the template-switching process, cDNA libraries will be enriched for full-length transcripts with the adverse effect that only those cDNA molecules that were reverse transcribed until the 5′ end will be amplified.

13.6
Expression Analysis in Single Cells and its Biological Relevance in Cancer

Single-cell gene expression analysis has not yet found its way into routine clinical diagnostics, although its potential in tumor diagnostics has been increasingly recognized over the last 10 years. The major problem is the uniqueness of each tumor with respect to cell composition and cell diversity, resulting in highly variable transcript levels for the gene of interest among the different cell subpopulations of a tumor. This makes it currently very difficult to reliably and reproducibly identify gene expression-based biomarkers predicting treatment decisions. As analyses of primary tumors, CTCs, and tumor cells at metastatic sites might suggest that tumor progression could be monitored at all tumor sites before and after treatment, single-cell analysis could be used to identify tumor cells with deterministic value in diagnostic and prognostic applications, even if tumor cells carry seemingly identical genomes. Several single-cell RT-qPCR studies have already clearly demonstrated the heterogeneity of gene expression

in individual cells [12,68,92,93]. The validation of the prognostic and predictive value of, for example, gene expression-based breast cancer prognosis signatures (MammaPrint, Oncotype DX, and PAM50) in single CTCs of patients affected by breast cancer in single-cell RT-qPCR assays executed on high-throughput semiautomated and microtranscriptomic platforms might pave the way to replace clinical diagnostics based on resected tumor biopsies by CTC-based tumor diagnostics ("liquid biopsy") being much more gentle on cancer patients. Single-cell microarray gene expression profiling has also been proven to identify biologically relevant differences among seemingly homogeneous cell populations. Such studies might also help to improve our understanding of what the hallmarks of a cancer cell are on the gene expression level.

13.7 Thoughts on Bioinformatics Approaches

There are numerous software packages – commercial or freeware – available to analyze high-dimensional data generated by RT-qPCR, gene expression profiling, and RNA-Seq. None of them specifically addresses the issue of single-cell analysis. The following considerations summarize briefly the essential effect of single-cell-derived data sets on the main functions of these software packages.

Typically, single-cell RT-qPCR assays have been analyzed by the comparative C_T method [94], albeit the stochastic variation of transcription resulting in highly variable transcripts levels was more and more realized as the inherent challenge in single-cell gene expression profiling. A recent publication takes this into account, and provides a framework for data exploration, quality control, and differential expression testing in order to completely use the salient features of a single-cell RT-qPCR experiment for optimal inference results [95].

To thoroughly analyze the more complex single-cell gene expression profiling data generated by microarray technology, one has additionally to accurately employ the algorithms for three major functions: normalization, grouping, and feature reduction. In order to identify small differences in transcript levels among individual cells, the normalization method for global bias removal must be carefully selected (e.g., normalization based on preselected housekeeping genes might prove to be difficult as such genes show biological expression variances in apparently homogeneous cell populations) [82]. Grouping methods help to identify genes that are coregulated or differentially regulated in individual cells. Roughly, one can differentiate supervised and unsupervised sets. Similar to grouping, feature reduction methods can also be classified in these two sets. Feature reduction is used to remove redundant information in order to identify characteristic features in high-dimensional data. The effect of different grouping and feature reduction algorithms in a single-cell application is nicely illustrated in Raychaudhuri *et al.* [96]. Nevertheless, the authors concluded that expression levels of key genes might still be useful as surrogate markers.

RNA-Seq generates large data sets of great complexity per experiment, requiring even more computational power and sophisticated data analysis pipelines than for microarray technology. As single-cell RNA-Seq analysis is not as mature as single-cell gene expression profiling and the positional bias of the existing methods is quite different, the precise effect of the different algorithms used in standard RNA-Seq pipelines on single-cell RNA-Seq data sets is still unclear. The effect of sequence mapping, transcriptome reconstruction, and expression quantification on single-cell gene expression data therefore needs to be thoroughly investigated and adjusted to enable a more profound understanding of the variations in gene expression among individual cells.

13.8
Future Impact of Single-Cell Analysis in Clinical Diagnosis

As elucidated above, genomic DNA is better suited for diagnostic approaches than RNA. Importantly, several of the previously described WGA methods have been adopted for commercially available single-cell WGA kits (Table 13.1), paving the way for the broad use of single-cell DNA analysis in many research laboratories. In addition to the technical hurdles WGA has to overcome (i.e., allelic dropout, sequence alteration, and homogeneous genome coverage), two additional criteria are of high importance for applying single-cell WGA in a clinical diagnostic setting: (i) sufficient sample volume for test repetitions and (ii) reliable long-term storage without quality loss. Here, at the moment, the most reliable data was collected for LA-PCR-based WGA as established our group. As all steps during the WGA protocol are non-random, reamplification of material is easily feasible [31,97] and stable results after long-term storage of more than 10 years have been demonstrated [37,47].

In the future, an important issue will be to develop assays that meticulously assess the quality of single-cell WGA products. Here, better-suited methods beyond mere measurement of DNA concentration after WGA will have to be developed in order to ensure that the information obtained by single-cell analysis really is reliable and can be used for clinical decision making. As an example, it has been described that CTCs are frequently apoptotic [98], and thus DNA fragmentation and degradation have already initiated.

Basically, two major problems currently prevent the transformation of single-cell-based nucleic acid analysis in routine clinical diagnosis: (i) the unprecedented number and variability of sequence errors derived by amplification from one single cell, and (ii) the need to evaluate RNA reference housekeeping genes in each clinical sample for normalization in gene expression profiling. In the case of RNA, RNase stability requires additional measures for sample logistics.

Combining high-dimensional transcriptomic and genomic information of the same cell [99] would allow us to correct for sequencing errors, and therefore to directly link phenotype and genotype (Figure 13.1). The error-free sequence

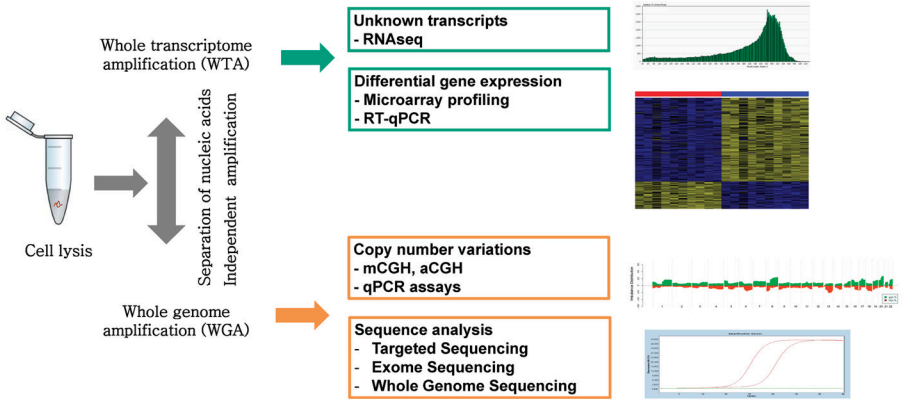


Figure 13.1 General workflow for combined high-resolution analysis of single cells. Isolated single cells are transferred to a lysis buffer followed by separation of the released nucleic acids. mRNA and genomic DNA are independently subjected to whole transcriptome (WTA) and whole genome (WGA) amplification, respectively, in order to generate sufficient amounts of high-quality material to analyze these samples by modern molecular biology techniques. The special molecular properties of WTA and WGA samples require not only the adaptation of the experimental procedures, but also the process of bioinformatic evaluation.

information generated by this method would be limited only to the gene copies expressed in a single cell at the time point of analysis. A further improvement would be the adaptation of the duplex sequencing method [100] to the single-cell level. This method improves the accuracy in next-generation sequencing by a magnitude of 10-million-fold under standard experimental settings by using unique identifiers for each starting molecule. Translating this sequencing accuracy to the single-cell level would allow us to transform routine clinical cancer diagnosis towards single-cell molecular diagnostics.

References

1 Bhattacherjee, V. *et al.* (2004) Laser capture microdissection of fluorescently labeled embryonic cranial neural crest cells. *Genesis*, **39** (1), 58–64.

2 Frumkin, D. *et al.* (2008) Amplification of multiple genomic loci from single cells isolated by laser micro-dissection of tissues. *BMC Biotechnology*, **8**, 17.

3 Yachida, S. *et al.* (2010) Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature*, **467** (7319), 1114–1117.

4 Choi, J.H. *et al.* (2010) Development and optimization of a process for automated recovery of single cells identified by microengraving. *Biotechnology Progress*, **26** (3), 888–895.

5 Kurimoto, K. *et al.* (2007) Global single-cell cDNA amplification to provide a template for representative high-density oligonucleotide microarray analysis. *Nature Protocols*, **2** (3), 739–752.

6 Reizel, Y. *et al.* (2011) Colon stem cell and crypt dynamics exposed by cell lineage reconstruction. *PLoS Genetics*, **7** (7), e1002192.

7 Shlush, L.I. *et al.* (2012) Cell lineage analysis of acute leukemia relapse uncovers the role of replication-rate heterogeneity and microsatellite instability. *Blood*, **120** (3), 603–612.

8 Zong, C. *et al.* (2012) Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science*, **338** (6114), 1622–1626.

9 Zhang, H. and Liu, K.K. (2008) Optical tweezers for single cells. *Journal of the Royal Society Interface*, **5** (24), 671–690.

10 Fuchs, A.B. *et al.* (2006) Electronic sorting and recovery of single live cells from microlitre sized samples. *Lab on a Chip*, **6** (1), 121–126.

11 Dalerba, P. *et al.* (2011) Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nature Biotechnology*, **29** (12), 1120–1127.

12 Guo, G. *et al.* (2010) Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Developmental Cell*, **18** (4), 675–685.

13 White, A.K. *et al.* (2011) High-throughput microfluidic single-cell RT-qPCR. *Proceedings of the National Academy of Sciences of the United States of America*, **108** (34), 13999–4004.

14 Araki, T., Yamamoto, A., and Yamada, M. (1987) Accurate determination of DNA content in single cell nuclei stained with Hoechst 33258 fluorochrome at high salt concentration. *Histochemistry*, **87** (4), 331–338.

15 Telenius, H. *et al.* (1992) Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics*, **13** (3), 718–725.

16 Fragouli, E. *et al.* (2006) Comparative genomic hybridization analysis of human oocytes and polar bodies. *Human Reproduction*, **21** (9), 2319–2328.

17 Wells, D. and Delhanty, J.D. (2000) Comprehensive chromosomal analysis of human preimplantation embryos using whole genome amplification and single cell comparative genomic hybridization.

Molecular Human Reproduction, **6** (11), 1055–1062.

18 Barrett, M.T., Reid, B.J., and Joslyn, G. (1995) Genotypic analysis of multiple loci in somatic cells by whole genome amplification. *Nucleic Acids Research*, **23** (17), 3488–3492.

19 Snabes, M.C. *et al.* (1994) Preimplantation single-cell analysis of multiple genetic loci by whole-genome amplification. *Proceedings of the National Academy of Sciences of the United States of America*, **91** (13), 6181–6185.

20 Zhang, L. *et al.* (1992) Whole genome amplification from a single cell: implications for genetic analysis. *Proceedings of the National Academy of Sciences of the United States of America*, **89** (13), 5847–5851.

21 Dietmaier, W. *et al.* (1999) Multiple mutation analyses in single tumor cells with improved whole genome amplification. *American Journal of Pathology*, **154** (1), 83–95.

22 Kittler, R., Stoneking, M., and Kayser, M. (2002) A whole genome amplification method to generate long fragments from low quantities of genomic DNA. *Analytical Biochemistry*, **300** (2), 237–244.

23 Fiegler, H. *et al.* (2007) High resolution array-CGH analysis of single cells. *Nucleic Acids Research*, **35** (3), e15.

24 Navin, N. *et al.* (2011) Tumour evolution inferred by single-cell sequencing. *Nature*, **472** (7341), 90–94.

25 Bi, W. *et al.* (2012) Detection of ≥1Mb microdeletions and microduplications in a single cell using custom oligonucleotide arrays. *Prenatal Diagnosis*, **32** (1), 10–20.

26 Voet, T. *et al.* (2013) Single-cell paired-end genome sequencing reveals structural variation per cell cycle. *Nucleic Acids Research*, **41** (12), 6119–6138.

27 Le Caignec, C. *et al.* (2006) Single-cell chromosomal imbalances detection by array CGH. *Nucleic Acids Research*, **34** (9), e68.

28 Mohlendick, B. *et al.* (2013) A robust method to analyze copy number alterations of less than 100kb in single cells using oligonucleotide array CGH. *PLoS One*, **8** (6), e67031.

29 Handyside, A.H. *et al.* (2004) Isothermal whole genome amplification from single and small numbers of cells: a new era for preimplantation genetic diagnosis of inherited disease. *Molecular Human Reproduction*, **10** (10), 767–772.

30 Hellani, A. *et al.* (2004) Multiple displacement amplification on single cell and possible PGD applications. *Molecular Human Reproduction*, **10** (11), 847–852.

31 Schardt, J.A. *et al.* (2005) Genomic analysis of single cytokeratin-positive cells from bone marrow reveals early mutational events in breast cancer. *Cancer Cell*, **8** (3), 227–239.

32 Esteban, J.A., Salas, M., and Blanco, L. (1993) Fidelity of phi 29 DNA polymerase. Comparison between protein-primed initiation and DNA polymerization. *Journal of Biological Chemistry*, **268** (4), 2719–2726.

33 Blanco, L. *et al.* (1989) Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. *Journal of Biological Chemistry*, **264** (15), 8935–8940.

34 Dean, F.B. *et al.* (2001) Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Research*, **11** (6), 1095–1099.

35 Murthy, K.K. *et al.* (2005) Assessment of multiple displacement amplification for polymorphism discovery and haplotype determination at a highly polymorphic locus, MC1R. *Human Mutation*, **26** (2), 145–152.

36 Klein, C.A. *et al.* (1999) Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. *Proceedings of the National Academy of Sciences of the United States of America*, **96** (8), 4494–4499.

37 Klein, C.A. *et al.* (2002) Genetic heterogeneity of single disseminated tumour cells in minimal residual cancer. *Lancet*, **360** (9334), 683–689.

38 Liu, C.L., Schreiber, S.L., and Bernstein, B.E. (2003) Development and validation of a T7 based linear amplification for

genomic DNA. *BMC Genomics*, **4** (1), 19.
39 Hughes, S. *et al.* (2005) The use of whole genome amplification in the study of human disease. *Progress in Biophysics and Molecular Biology*, **88** (1), 173–189.
40 Lu, S. *et al.* (2012) Probing meiotic recombination and aneuploidy of single sperm cells by whole-genome sequencing. *Science*, **338** (6114), 1627–1630.
41 Schmidt-Kittler, O. *et al.* (2003) From latent disseminated cells to overt metastasis: genetic analysis of systemic breast cancer progression. *Proceedings of the National Academy of Sciences of the United States of America*, **100** (13), 7737–7742.
42 Stoecklein, N.H. *et al.* (2008) Direct genetic analysis of single disseminated cancer cells for prediction of outcome and therapy selection in esophageal cancer. *Cancer Cell*, **13** (5), 441–453.
43 Weckermann, D. *et al.* (2009) Perioperative activation of disseminated tumor cells in bone marrow of patients with prostate cancer. *Journal of Clinical Oncology*, **27** (10), 1549–1556.
44 Fuhrmann, C. *et al.* (2008) High-resolution array comparative genomic hybridization of single micrometastatic tumor cells. *Nucleic Acids Research*, **36** (7), e39.
45 Mathiesen, R.R. *et al.* (2012) High-resolution analyses of copy number changes in disseminated tumor cells of patients with breast cancer. *International Journal of Cancer*, **131** (4), E405–E415.
46 Heitzer, E. *et al.* (2013) Complex tumor genomes inferred from single circulating tumor cells by array-CGH and next-generation sequencing. *Cancer Research*, **73** (10), 2965–2975.
47 Czyz, Z.T. *et al.* (2013) Reliable single-cell array CGH for application to clinical samples. *PLoS One*, **9** (1), e85907.
48 Blainey, P.C. (2013) The future is now: single-cell genomics of bacteria and archaea. *FEMS Microbiology Reviews*, **37** (3), 407–427.
49 Shapiro, E., Biezuner, T., and Linnarsson, S. (2013) Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nature Reviews Genetics*, **14** (9), 618–630.
50 Schadt, E.E., Turner, S., and Kasarskis, A. (2010) A window into third-generation sequencing. *Human Molecular Genetics*, **19** (R2), R227–R240.
51 Xu, M., Fujita, D., and Hanagata, N. (2009) Perspectives and challenges of emerging single-molecule DNA sequencing technologies. *Small*, **5** (23), 2638–2649.
52 Teer, J.K. and Mullikin, J.C. (2010) Exome sequencing: the sweet spot before whole genomes. *Human Molecular Genetics*, **19** (R2), R145–R151.
53 Giulino-Roth, L. *et al.* (2012) Targeted genomic sequencing of pediatric Burkitt lymphoma identifies recurrent alterations in antiapoptotic and chromatin-remodeling genes. *Blood*, **120** (26), 5181–5184.
54 Valencia, C.A. *et al.* (2013) Comprehensive mutation analysis for congenital muscular dystrophy: a clinical PCR-based enrichment and next-generation sequencing panel. *PLoS One*, **8** (1), e53083.
55 Chubb, J.R. *et al.* (2006) Transcriptional pulsing of a developmental gene. *Current Biology*, **16** (10), 1018–1025.
56 Golding, I. *et al.* (2005) Real-time kinetics of gene activity in individual bacteria. *Cell*, **123** (6), 1025–1036.
57 Rajan, S. *et al.* (2011) The living microarray: a high-throughput platform for measuring transcription dynamics in single cells. *BMC Genomics*, **12**, 115.
58 Ziauddin, J. and Sabatini, D.M. (2001) Microarrays of cells expressing defined cDNAs. *Nature*, **411** (6833), 107–110.
59 Femino, A.M. *et al.* (1998) Visualization of single RNA transcripts *in situ*. *Science*, **280** (5363), 585–590.
60 Itzkovitz, S. and van Oudenaarden, A. (2011) Validating transcripts with probes and imaging technology. *Nature Methods*, **8** (4 Suppl.), S12–S19.
61 Levsky, J.M. *et al.* (2002) Single-cell gene expression profiling. *Science*, **297** (5582), 836–840.
62 Lubeck, E. and Cai, L. (2012) Single-cell systems biology by super-resolution

imaging and combinatorial labeling. *Nature Methods*, **9** (7), 743–748.
63 Raj, A. *et al.* (2008) Imaging individual mRNA molecules using multiple singly labeled probes. *Nature Methods*, **5** (10), 877–879.
64 Larsson, C. *et al.* (2010) *In situ* detection and genotyping of individual mRNA molecules. *Nature Methods*, **7** (5), 395–397.
65 Weibrecht, I. *et al.* (2013) *In situ* detection of individual mRNA molecules and protein complexes or post-translational modifications using padlock probes combined with the *in situ* proximity ligation assay. *Nature Protocols*, **8** (2), 355–372.
66 Bengtsson, M. *et al.* (2005) Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. *Genome Research*, **15** (10), 1388–1392.
67 Taniguchi, K., Kajiyama, T., and Kambara, H. (2009) Quantitative analysis of gene expression in a single cell by qPCR. *Nature Methods*, **6** (7), 503–506.
68 Warren, L. *et al.* (2006) Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR. *Proceedings of the National Academy of Sciences of the United States of America*, **103** (47), 17807–17812.
69 Bontoux, N. *et al.* (2008) Integrating whole transcriptome assays on a lab-on-a-chip for single cell gene profiling. *Lab on a Chip*, **8** (3), 443–450.
70 Esumi, S. *et al.* (2008) Method for single-cell microarray analysis and application to gene-expression profiling of GABAergic neuron progenitors. *Journal of Neuroscience Research*, **60** (4), 439–451.
71 Hartmann, C.H. and Klein, C.A. (2006) Gene expression profiling of single cells on large-scale oligonucleotide arrays. *Nucleic Acids Research*, **34** (21), e143.
72 Kamme, F. *et al.* (2003) Single-cell microarray analysis in hippocampus CA1: demonstration and validation of cellular heterogeneity. *Journal of Neuroscience*, **23** (9), 3607–3615.
73 Kurimoto, K. *et al.* (2006) An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic Acids Research*, **34** (5), e42.
74 Sasagawa, Y. *et al.* (2013) Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. *Genome Biology*, **14** (4), R31.
75 Sul, J.Y. *et al.* (2009) Transcriptome transfer produces a predictable cellular phenotype. *Proceedings of the National Academy of Sciences of the United States of America*, **106** (18), 7624–7629.
76 Tietjen, I. *et al.* (2003) Single-cell transcriptional analysis of neuronal progenitors. *Neuron*, **38** (2), 161–175.
77 Hashimshony, T. *et al.* (2012) CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Reports*, **2** (3), 666–673.
78 Islam, S. *et al.* (2011) Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Research*, **21** (7), 1160–1167.
79 Ramskold, D. *et al.* (2012) Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nature Biotechnology*, **30** (8), 777–782.
80 Tang, F. *et al.* (2009) mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods*, **6** (5), 377–382.
81 Bengtsson, M. *et al.* (2008) Quantification of mRNA in single cells and modelling of RT-qPCR induced noise. *BMC Molecular Biology*, **9**, 63.
82 Reiter, M. *et al.* (2011) Quantification noise in single cell experiments. *Nucleic Acids Research*, **39** (18), e124.
83 Eberwine, J. *et al.* (1992) Analysis of gene expression in single live neurons. *Proceedings of the National Academy of Sciences of the United States of America*, **89** (7), 3010–3014.
84 Matz, M. *et al.* (1999) Amplification of cDNA ends based on template-switching effect and step-out PCR. *Nucleic Acids Research*, **27** (6), 1558–1560.
85 Zhu, Y.Y. *et al.* (2001) Reverse transcriptase template switching: a

SMART approach for full-length cDNA library construction. *Biotechniques*, **30** (4), 892–897.

86 Ko, M.S. *et al.* (1990) Unbiased amplification of a highly complex mixture of DNA fragments by 'lone linker'-tagged PCR. *Nucleic Acids Research*, **18** (14), 4293–4294.

87 Iscove, N.N. *et al.* (2002) Representation is faithfully preserved in global cDNA amplified exponentially from sub-picogram quantities of mRNA. *Nature Biotechnology*, **20** (9), 940–943.

88 Liu, J., Hansen, C., and Quake, S.R. (2003) Solving the "world-to-chip" interface problem with a microfluidic matrix. *Analytical Chemistry*, **75** (18), 4718–4723.

89 Sanchez-Freire, V. *et al.* (2012) Microfluidic single-cell real-time PCR for comparative analysis of gene expression patterns. *Nature Protocols*, **7** (5), 829–838.

90 Spurgeon, S.L., Jones, R.C., and Ramakrishnan, R. (2008) High throughput gene expression measurement with real time PCR in a microfluidic dynamic array. *PLoS One*, **3** (2), e1662.

91 Geiss, G.K. *et al.* (2008) Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nature Biotechnology*, **26** (3), 317–325.

92 Warren, L.A. *et al.* (2007) Transcriptional instability is not a universal attribute of aging. *Aging Cell*, **6** (6), 775–782.

93 Wills, Q.F. *et al.* (2013) Single-cell gene expression analysis reveals genetic associations masked in whole-tissue experiments. *Nature Biotechnology*, **31** (8), 748–752.

94 Schmittgen, T.D. and Livak, K.J. (2008) Analyzing real-time PCR data by the comparative C_T method. *Nature Protocols*, **3** (6), 1101–1108.

95 McDavid, A. *et al.* (2013) Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. *Bioinformatics*, **29** (4), 461–467.

96 Raychaudhuri, S. *et al.* (2001) Basic microarray analysis: grouping and feature reduction. *Trends in Biotechnology*, **19** (5), 189–193.

97 Geigl, J.B. and Speicher, M.R. (2007) Single-cell isolation from cell suspensions and whole genome amplification from single cells to provide templates for CGH analysis. *Nature Protocols*, **2** (12), 3173–3184.

98 Mehes, G. *et al.* (2001) Circulating breast cancer cells are frequently apoptotic. *American Journal of Pathology*, **159** (1), 17–20.

99 Klein, C.A. *et al.* (2002) Combined transcriptome and genome analysis of single micrometastatic cells. *Nature Biotechnology*, **20** (4), 387–392.

100 Schmitt, M.W. *et al.* (2012) Detection of ultra-rare mutations by next-generation sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, **109** (36), 14508–14513.

14
Detecting Dysregulated Processes and Pathways

Daniel Stöckel and Hans-Peter Lenhof

14.1
Introduction

Differential expression analysis has opened novel avenues to study the mechanisms of pathogenic processes, and to gain deeper insights into the initiation and progression of complex diseases such as cancer. While differential expression analysis has the potential to revolutionize the diagnosis, prognosis, and therapy of diseases, the translation of these approaches into routine clinical practice remains a major challenge. Over the past decade, a large number of publications have provided strong evidence that comprehensive expression profiles are excellent biomarkers [1–7] for almost any relevant clinical task; however, almost none of these approaches has made it into clinical application and routine. One reason for this is the noisy, high-dimensional data produced by biotechnological high-throughput methods. This requires that the applied statistical methods are especially robust and disregard much of the intricate interplay of the measured genes in order to battle the "curse of dimensionality" [8]. A more severe problem, however, is that batch effects caused, for example, by different experimental environments may dominate the signal in the data and thus complicate or even prevent the implementation of standardized routines. Additionally, the rapid development of new biotechnological high-throughput methods, such as next-generation sequencing (NGS), leads to a quick switchover of the experimental platforms that all have their individual technical specifics and problems, and hence require novel specialized analysis methods and pipelines.

In an ideal situation, the comparison of the biological processes in, for example, tumor and normal cells would be based on comprehensive time series data sets that provide the complete molecular contents of the different cell types, including the three-dimensional positions of all molecules in the cells (and in the extracellular matrix) for each time point. However, nowadays such a ideal starting point for differential analysis must still be considered as a vision or even a utopian dream. Despite the impressive technological progress that has been

Nucleic Acids as Molecular Diagnostics, First Edition. Edited by Andreas Keller and Eckart Meese.

made in the field of proteomics during the last decade, only a few thousands of the proteins in a sample can be quantified using state-of-the-art experimental setups [9]. Instead of quantifying proteins in a sample directly, mRNA expression can be measured as an approximation. However, this relies on the assumption that the number of mRNA copies in the cell is proportional to the amount of produced proteins, neglecting the fact that due to post-transcriptional regulation, mRNA and protein expression are often only weakly correlated [10].

For more than 10 years, microarrays have been the state-of-the-art for measuring comprehensive expression profiles of genes, although they are now quickly being replaced by NGS technologies, called RNA-Seq [11–13]. Both technologies are relatively cost-efficient, require tolerable experimental effort, and allow for measuring the expression of all genes simultaneously. However, both methods also produce rather noisy data sets that have to be normalized using sophisticated statistical approaches. Additionally, microarrays and RNA-Seq approaches can also be used to measure the expression profiles of non-coding RNAs (ncRNAs) present in the cells [14,15]. The number of ncRNAs is steadily increasing and novel insights into the roles of ncRNAs in normal and pathogenic biological processes underline the relevance and importance of these non-coding RNAs (e.g., [16–18]). An important class of ncRNAs are so-called microRNAs (miRNAs) that play a central role in transcriptional and post-transcriptional regulation of gene expression. Mature miRNAs are short, around 22-nucleotide long RNAs that, together with a number of proteins, build the so-called RNA-induced silencing complex (RISC). This complex is able to recognize and bind to sequences within specific target mRNAs that are at least partially complementary to the miRNA sequence. This reaction usually leads to a negative regulation of the corresponding gene products by either degradation of the mRNA or by translational suppression. However, there are also many instances where the expression of the target mRNA seems to be positively regulated. As a single type of miRNA can target various genes, changes of the expression level of such miRNAs may have a strong influence on human metabolism and may even control entire cellular processes. For humans, at the time of writing (May 2014), roughly 1900 precursor and 2600 mature miRNA sequences can be found in the mirBase repository (release 20) [19], with new miRNAs routinely being discovered. Dysregulation of miRNAs has been shown to be linked to various diseases and pathological mechanisms [20–22]. Accordingly, miRNAs are of great interest for diagnostic and therapeutic purposes.

In this chapter, we review bioinformatics approaches for detecting dysregulated genes, biological categories, pathways, and subnetworks that are based on the differential analyses of mRNA and/or ncRNA expression profiles. We will further discuss how the resulting information on dysregulated processes can be used to improve diagnostics, prognostics, as well as therapy stratification. Due to space constraints, we are not able to present all known methods and algorithms, let alone their underlying mathematical concepts. Instead, we will sketch a selection of popular approaches for selected topics. For complementary introductions

into the fields of probability theory, statistics, and machine learning, we refer the reader elsewhere [23–25].

In Section 14.2, we familiarize the reader with expression data and the methods used to prepare and normalize the data. In Section 14.3, we introduce the biological networks that are used in the presented bioinformatics approaches. We summarize standard approaches for measuring the degree of dysregulation of single genes in Section 14.4. Section 14.5 presents an overview of approaches for Over-Representation Analysis (ORA) and Gene Set Enrichment Analysis (GSEA). Effective and efficient methods for detecting dysregulated pathways and subnetworks in biological networks are described in Section 14.6. As the analysis of miRNAs data requires different tools than mRNA data, we will introduce these concepts separately in Section 14.7. Finally, with Section 14.8 we venture in to mostly unexplored territory and provide an outlook on differential network analysis.

14.2 Measuring and Normalizing Expression Profiles

The bioinformatics approaches presented in this chapter combine and integrate comprehensive expression profiles and biological networks in order to:

- Elucidate the molecular mechanisms of pathogenic processes.
- Identify characteristic expression profiles for certain diseases or even disease subtypes (biomarkers).
- Detect putative target molecules for the therapy of diseases.
- Optimize therapies (therapy stratification).

Understanding these methods, however, requires basic knowledge of the different types of expression data, which will be introduced in this section.

We start with a short summary of expression data focusing on microarray technology. We briefly describe the experimental setup, and list some popular methods used for preparing and normalizing the raw expression data. For NGS technologies and the corresponding normalization approaches, we refer the reader elsewhere (e.g., [26,27]).

14.2.1 Microarray Experiments

Here, we discuss the processing that is necessary to transform the raw readout of a microarray experiment into usable and comparable expression values. As a detailed discussion of all available microarray platforms, let alone RNA-Seq techniques, is beyond the scope of this chapter, we limit ourselves to the widely used Affymetrix GeneChip technology.

The key idea of RNA microarrays is to capture selected transcripts via complementary probes that have been immobilized on a substrate (e.g., a glass slide). To this end, Affymetrix microarrays are designed as a rectangular grid where each cell or *spot* of the two-dimensional grid contains many copies of an oligonucleotide *probe* that is complementary to a part of a specific transcript. The oligonucleotides are constructed using a photolithographic process and have a length of 25 bases. Every transcript of interest is covered by approximately 10 probes. Modern chips with millions of probes or spots on a single array cover the exons of all genes, but may also cover splice junctions and ncRNA transcripts [28].

For analyzing RNA extracted from cells or tissues, the purified RNA is reverse transcribed. The resulting cDNA is then amplified using polymerase chain reaction (PCR). During this step fluorescence markers are incorporated into the cDNA. In the next step, the labeled cDNA is put on the chip and hybridizes with complementary transcripts on the microarray. After several washing steps that remove unspecifically hybridized transcripts, the chip is read using a laser scanner. This results in a set of monochrome images. The intensities of the spots are subsequently transformed into expression measures for the target transcripts.

14.2.2 Normalization

Many factors involved in the process described in the previous section may add noise to the raw microarray data, including unspecific hybridization. To control this error, Affymetrix chips contain perfect match (PM) and mismatch (MM) probes. MM probes target the same sequence as PM probes, but differ in one or a few nucleotides. The simplest way to account for background noise is to compute the quantity PM – MM. This strategy, in a modified version, is used by the MAS 5.0 [29] vendor software. However, it has been shown [30,31] that using MM values can be unreliable as these are often larger than their PM counterparts. Irizarry *et al.* [31,32] propose the RMA (Robust Multiarray Average) procedure, which only relies on the PM values for background correction. To ensure comparability across arrays, RMA applies quantile normalization (QN), which effectively matches the expression value distribution across arrays. QN is a non-parametric method that maps the quantiles of a source distribution (the measured data) to a target distribution. For expression data the target distribution is usually assumed to be a normal or log-normal distribution. Furthermore, probes targeting the same transcripts, so-called *probe sets*, must be summarized in order to obtain a per-transcript expression measure. This is achieved by using a linear model. RMA has established itself as the *de facto* standard for the analysis of Affymetrix microarray data.

RMA uses the QN procedure. An alternative normalization approach is variance stabilizing normalization (VSN) [33]. It is based on the observation that measurements at high intensities exhibit a larger variance than those at low

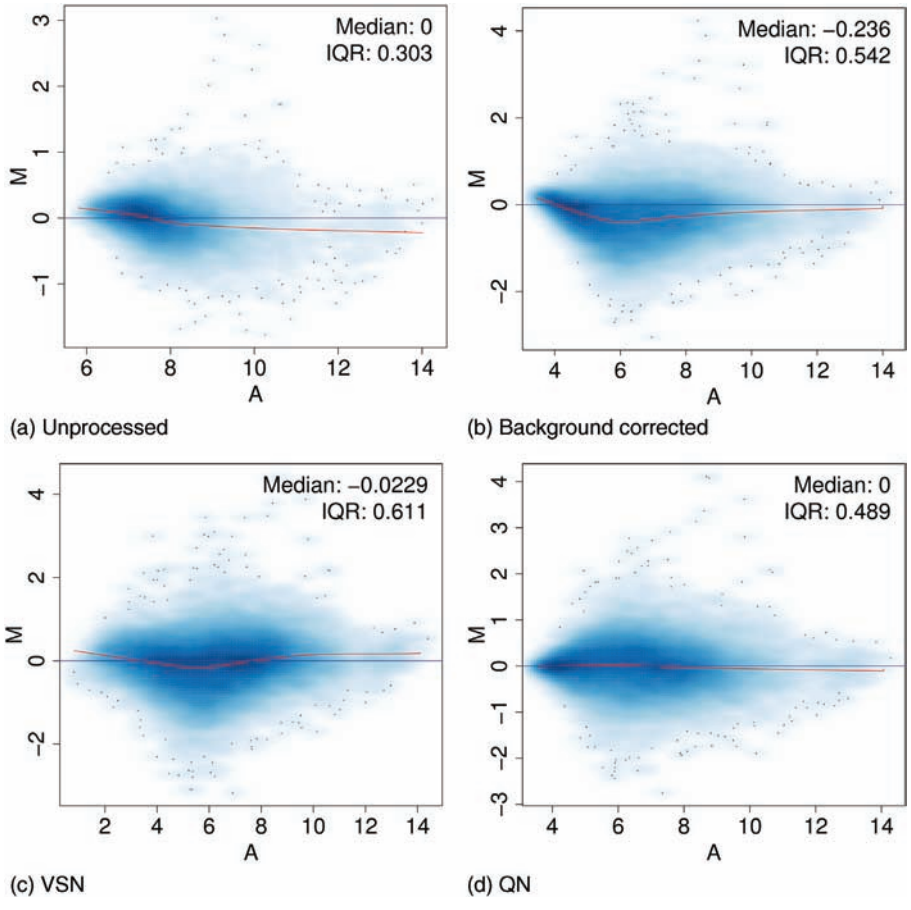


Figure 14.1 MA plots of the sample GSM368775 from the GSE14767 Gene Expression Omnibus (GEO) series. MA plots are commonly used to examine microarray data for unwanted dependence between expression strength and differential expression. The M axis shows the log difference between sample and control data, whereas the A axis shows the mean of the logarithmized sample and control. IQR denotes the interquartile range. In panel (a) no processing has been performed and the variance of the data is dependent on expression intensities. Background correction (b) (here, using the RMA method) can significantly change the shape of the data leading to an overall trend towards under expression. VSN (c) equalizes the variance along the M axis, the overall distribution of the data is hardly changed. QN (d) always generates a well shaped distribution, even in the case of corrupt chips, which can be problematic for further analysis.

intensities. VSN compensates for this effect by considering additive *and* multiplicative error terms.

In general, the choice of the appropriate normalization technique is strongly dependent on the platform used. Figure 14.1 visualizes the impact of the normalization methods on the processed data. For a more detailed introduction to low-level microarray analysis, we refer the reader to Huber *et al.* [34].

14.2.3 Batch Effects

Variations of the expression levels not caused by biological effects but by changes of experimental conditions are called batch effects. The causes of batch effect range from spatial and temporal factors leading to, for example, variations of the temperature or of the atmospheric ozone levels to slight differences in the "batches" of chemical reagents used, and variations in sample preparation and in the experimental protocol. The platform employed (e.g., the model of microarray or sequencer) can also lead to vastly differing results.

Hence, when combining the results from multiple expression experiments, it is necessary to correct for batch effects that arise from different experimental settings. If not accounted for, batch effects can bias the results of an analysis. For example, let us assume that in a large cancer study all healthy tissue has been analyzed on one day and all cancer samples on another day. In this situation, it is likely that features distinguishing healthy and cancer samples are only classifiers for the day the samples were obtained and not meaningful biomarkers. Although this problem can be alleviated by employing a sensible experimental design featuring concurrent measurements of healthy and diseased samples, the fact that the data was generated in two batches can still pose a problem. In practice, the clustering of samples from, for example, different laboratories will likely result in clusters corresponding to the laboratories rather than the actual clinical condition of the sample.

Methods like DWD [35], SVA [36], ComBat [37], or RUV-2 [38] allow us to reduce this type of bias from combined data sets. However, all methods except RUV-2 rely on a good experimental design, where every batch contains at least one sample of every condition. RUV-2 instead requires a user-defined set of genes that is expected to show differential expression and a set of genes that is expected to remain constant. Moreover, it should be noted that a reasonable sample count is needed to allow all methods to operate reliably. Batch correction may also remove real biological variations and correlations. For more details we refer the reader to the survey by Lazar *et al.* [39], which gives an overview over available strategies for the removal of batch effects.

14.3 Biological Networks

Human metabolism is a enormously complex biochemical process involving hundreds of thousands or even millions of different molecules that interact in a highly interwoven way. The endeavor to understand at least tiny parts of the human metabolism required the artificial decomposition of the metabolism into smaller manageable parts that we now call biological processes, categories, or pathways, which have been classified and structured in ontologies and that are collected and stored in a variety of databases. One of the best known projects

here is the Gene Ontology (GO) [40] project that provides a hierarchy of terms that both categorize and annotate known genes and proteins across species. In Section 14.5, we will present methods which operate on categories that can directly be extracted from GO. Other databases like KEGG [41], Reactome [42], or BIND [43] contain pathways – sets of reactions and binding events in which genes, proteins, RNA, and metabolites take part, and that together define a biological function. This leads us to the concept of biological networks. In general, a biological network is a graph $G = (V, E)$, where the set V of nodes represents biological entities such as genes, proteins, or metabolites and the set $E \subseteq V \times V$ of edges their molecular interactions. A protein–protein interaction (PPI) network, for example, is a graph $G = (V, E)$ where the node set V consists of proteins. If two proteins $v_1, v_2 \in V$ are believed to interact in some way, the two nodes are connected by an edge (v_1, v_2). Other types of biological networks are metabolic and signaling or regulatory networks. Metabolic networks encode how metabolites are converted by enzymes, gene regulatory networks describe how genes regulate each other, and signaling networks encode the information flow in the cell. Network databases often incorporate multiple interaction types such that the interactions contained need to be filtered for the information of interest. Another important aspect when working with network data is of course the quality of the data. Projects like KEGG, Reactome, or DIP [41,42,44] carefully examine published studies in order to create a manually curated compilation of reliable interaction data. This *modus operandi* guarantees high-quality databases, but tends to be labor-intensive and may lead to a lack of many true interactions. Text-mining-based approaches that automatically scan a corpus of publications for interactions have also been developed (e.g., [45–47]). While these methods require significantly less manpower, they also lead to more false-positive entries. Prediction methods can be used to *infer* biological networks from data, and are used to build interaction databases like OPHID [48] and PIPs [49]. However, it should be noted that the high dimensionality of the *network inference* task requires a huge amount of data for producing reliable predictions. Even then these networks contain a high amount of errors. Nevertheless, if no network data for a given set of genes or an entire organism is available, these methods can provide valuable insights [50,51]. We will revisit the idea of network inference in the context of cancer therapy (see Section 14.8). Finally, meta databases like STRING [52] or information systems like BN++ [53] or ONDEX [54] integrate existing databases and several data sources into large data warehouses. Besides resulting in a larger dataset, this approach has the advantage that, a confidence value can be assigned to each entry based on the frequency of its occurrence in the source databases.

14.4
Measuring the Degree of Deregulation of Individual Genes

We now assume that we are given the normalized expression data sets from a sample group S (e.g., expression data of cancer tissues) and of a control group C

(e.g., expression data of healthy tissues). The first step of any analysis is the determination of the degree of deregulation for each individual gene and, if possible, the calculation of a significance value or *P*-value for each gene. To this end, various established statistical methods are available for microarray and RNA-Seq data.

14.4.1 Microarray Data

As mentioned in Section 14.2.2, we assume that the expression values for each gene and both groups are normally distributed. However, if the control and sample groups consist of only a few data points, a reliable estimation of the standard deviation is not possible. In this case, where the standard approaches are not applicable, one has to resort to simple measures such as the log-fold quotient:

$$d_i := \log\left(\frac{\widehat{\mu}_i^{\mathrm{s}}}{\widehat{\mu}_i^{\mathrm{c}}}\right)$$

where $\widehat{\mu}_i^{\mathrm{s}}$ and $\widehat{\mu}_i^{\mathrm{c}}$ are the estimated mean expression values of gene i in the sample and the control group. It should be noted that no significance information is available in this case and thus no statistical reliability can be assigned to the obtained scores. When comparing a single sample j (e.g., obtained from one of the cancer patients) against a larger background distribution, the z-score is the measure of choice:

$$z_{ij} := \frac{x_{ij} - \widehat{\mu}_i^{\mathrm{c}}}{\widehat{\sigma}_i^{\mathrm{c}}}$$

where x_{ij} denotes the expression value of gene i in sample j and $\widehat{\sigma}_i^{\mathrm{c}}$ denotes the sample standard deviation of gene i in the control group. In a research scenario where larger cohorts are available, the use of Student's t-statistic is often recommended (for the sake of readability, we drop the subscript i):

$$t := \frac{\widehat{\mu}_{\mathrm{s}} - \widehat{\mu}_{\mathrm{c}}}{\sqrt{\widehat{\sigma}_{\mathrm{c}}^2/|C| + \widehat{\sigma}_{\mathrm{s}}^2/|S|}}$$

In Figure 14.2 we applied the t-statistics to simulated data and computed a *P*-value for the value obtained.

More advanced methods, like the shrinkage t-statistics by Opgen-Rhein and Strimmer [55], improve upon Student's t-statistic. The key idea of the shrinkage t-statistics is not to directly use the usual estimator for the variance:

$$\widehat{\sigma}_i^2 = \frac{1}{m-1}\sum_{j=1}^{m}\left(x_{ij} - \widehat{\mu}_i\right)^2$$

where m is the number of samples, $x_{ij} \in \mathbb{R}$ is the expression value of gene i in sample j, and $\widehat{\mu}_i = \sum_{j=1}^{m} x_{ij}/m$ the respective sample mean. Instead, the estimates

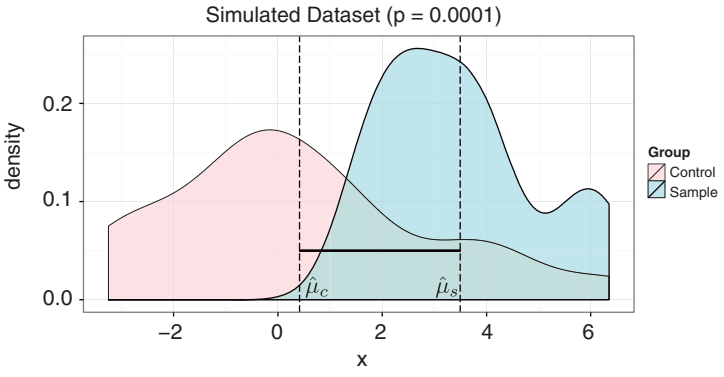


Figure 14.2 The estimated density of two groups drawn from randomly generated, normally distributed random variables ($\mu_c = 1$, $\mu_s = 3$, $\sigma_c = 2$, $\sigma_s = 1.5$). Each group consists of 20 samples. The *t*-test reports a significant shift of the two random groups at confidence level $\alpha = 0.05$. Dashed lines represent the estimated sample means.

for the standard deviation $\widehat{\sigma}_i$ are "shrunk" towards the median standard deviation across all genes $\widehat{\sigma}_{\text{median}}$:

$$\widehat{\sigma}_i^* = \lambda^* \widehat{\sigma}_{\text{median}} + (1 - \lambda^*)\widehat{\sigma}_i$$

The optimal mixing proportion λ^* is given by:

$$\lambda^* = \min\left(1, \frac{\sum_{k=1}^{n} \widehat{\text{Var}}(\widehat{\sigma}_k)}{\sum_{i=1}^{n} (\widehat{\sigma}_i - \widehat{\sigma}_{\text{median}})^2}\right)$$

The term $\widehat{\text{Var}}(\widehat{\sigma}_i)$ denotes an estimate of the variance of the computed standard deviations and can easily be derived from the data. For an overview of other suitable test statistics, see [55], for example.

14.4.2 RNA-Seq Data

RNA-Seq uses short-read sequencing technology for quantifying the expression of a transcript. The reads are aligned to the transcriptome or genome and often the (normalized) read count per gene directly serves as a measure for expression (see Section 14.2). This is a substantial difference to the microarray case, where transcription is measured indirectly via fluorescence intensities.

As we are dealing with count data though, assuming a continuous distribution such as the normal distribution for analysis is not appropriate. Instead, most methods for differential expression estimation either assume an underlying Poisson or negative binomial distribution. Examples for this are the method by Marioni *et al.* [13] and the DEGseq [56] package (Poisson) as well as edgeR [57] and DESeq [58] (Negative Binomial). More recent methods such as

MMSEQ [59], RSEM [60], Cufflinks [61], and BitSeq [62] incorporate both: methods for the computation of *per transcript* expression values and differential expression estimation. Thus, they are capable of estimating expression for haplotypes or splicing isoforms individually.

The detection of differentially expressed genes in RNA-Seq data is a very active research area and thus choosing or recommending specific methods is difficult. However, we would like to note that the set of detected differential genes can vastly differ between two approaches, as outlined by various studies [63–65]. Thus, in order to control the false-positive rate, we recommend being conservative with respect to choosing significance levels.

14.5 Over-Representation Analysis and Gene Set Enrichment Analysis

Using one of the standard methods discussed in the previous section, the degree of dysregulation for every gene can be calculated and stored in a list $L = (d_1, \ldots, d_n)$ sorted in a descending order such that the most upregulated gene g_1 has the value d_1 and the most downregulated gene g_n has the value d_n. A standard problem in computational biology is the following: given the vector L and a biological category or process represented by a set S of genes, we would like to know if the category is dysregulated (i.e., if the process is up- or downregulated).

The most popular approaches for the analysis of gene sets are ORA [66] and GSEA [67]. Both methods test if a predefined category of genes S is enriched with differentially expressed genes. ORA requires the user to select a threshold c for determining the set $L' \subset L$ of differentially expressed genes. Here, gene i is considered differentially expressed if $|d_i| \geq c$ (or $d_i \geq c$ or $d_i \leq -c$, respectively). Assume that t dysregulated genes out of L' are part of S. A P-value can be assigned to S by calculating the probability that t or more genes of L' are members of S using the hypergeometric distribution:

$$p(T \geq t) = 1 - \sum_{i=0}^{t} \frac{\binom{|L'|}{i}\binom{|L| - |L'|}{|S| - i}}{\binom{|L|}{|S|}} \tag{14.1}$$

The choice of a threshold c for determining the set of differentially expressed genes is of course problematic. While setting the value too high may result in significant categories not being detected, setting the value too low may lead to many false-positive discoveries. GSEA overcomes this problem by considering all genes. The fundamental idea is that a category is significantly enriched with upregulated (downregulated) genes if its genes are accumulated at the top (bottom) of the sorted list L. This can be formalized using the Kolmogorov–Smirnov running-sum statistics [68]: when traversing

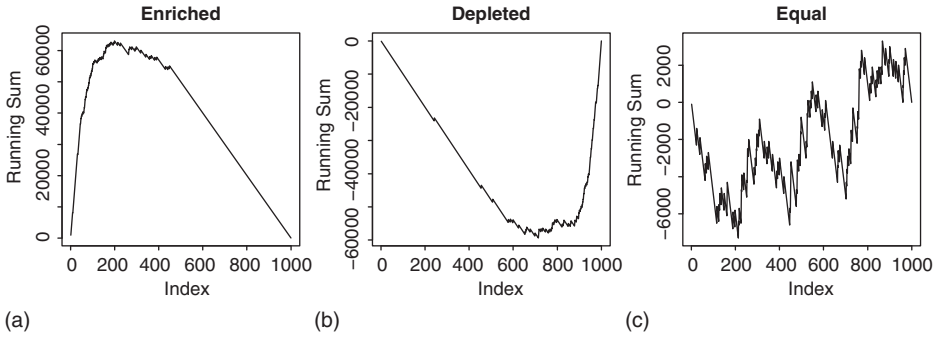


Figure 14.3 Kolmogorov–Smirnov running-sum statistics computed from simulated data. Three situations are depicted that commonly arise: the genes belonging to the query category are enriched at the beginning (a), the end (b), or uniformly distributed across the list (c).

the list from top to bottom, every time a gene from the category is encountered, the so-called running sum $RS(i)$ is increased ($RS(i) = RS(i-1) + w_i^+$) and decreased otherwise ($RS(i) = RS(i-1) - w_i^-$). See Figure 14.3.

The maximal deviation $D_{\max}$ of the running sum from zero is used as a test statistic. Here, $D_{\max}$ is the largest occurring absolute value of the running sum while iterating over the list L:

$$D_{\max} = \max_i \|RS(i)\| \tag{14.2}$$

If the elements of a category are equally distributed throughout the list, the running sum will oscillate around the "zero line" and the resulting $D_{\max}$ value will be small. If, however, there is an enrichment of genes from the considered category at the top (bottom), $D_{\max}$ will take larger values. GSEA approaches differ in the way the magnitudes of the increments and decrements of the running sum are chosen. They can be roughly classified in weighted and unweighted methods (e.g., [67,69]). The significance of the test values $D_{\max}$ are usually computed using permutation tests. In the case of unweighted GSEA, an exact algorithm [69] for the calculation of P-values is available.

Ackermann and Strimmer [70] have conducted an extensive survey of gene set analysis and they have carried out comprehensive tests of the different approaches. Their results indicated, among others, that simple statistics like the sum, mean, or median of the single gene statistics, the max-mean statistics, or the Wilcoxon rank-sum test may show a similar or even better performance than GESA approaches dependent on the test conditions. The authors also describe a sophisticated modular framework for gene set enrichment analysis. Nevertheless GSEA and ORA are standard tools for the analysis of expression data, and are routinely applied in a multitude of scenarios. For reviews on GSEA, see Nam *et al.* [71] or Song *et al.* [72], for example.

14.5.1 Multiple Hypothesis Testing

When discussing gene set enrichment approaches it is important to alert the reader to the issue of multiple hypothesis testing. Statistical tests such as ORA and GSEA require a user-defined threshold α that defines the risk of making a type I error: rejecting the H_0 hypothesis although it was true. This *significance level* separates the data points that the user believes to be significant from those he/she believes to be insignificant. As enrichment methods usually assess multiple categories, it becomes likely that some categories are reported as significant due to pure chance. If we assume that 100 (random) categories will be tested at a significance level of $\alpha = 0.05$, we can expect $\alpha \cdot 100 = 5$ categories for which the H_0 hypothesis is wrongly rejected. In order to counteract this error, *P*-value adjustments such as the Bonferroni [73] and Benjamini–Hochberg [74] procedures *must* be performed. Roback *et al.* [75] give a detailed overview on the multiple hypothesis testing problem and discuss cases in which adjustment is or is not appropriate.

14.5.2 Network-based GSEA Approaches

A common problem of the classical gene set approaches is that, although they consider groups of genes that are usually part of the same biological process, they treat every gene as independent from the others. However, genes that interact in a biological process have strong interdependencies that may bias the results of these classical set-based statistical approaches. Hence, novel GSEA approaches make great efforts to consider and model the interdependencies between the genes involved, which usually requires the availability of an interaction or regulatory network that describes the interactions between the genes or the corresponding proteins.

An example of a gene set approach that uses knowledge about gene or protein interactions is the Gene Graph Enrichment Analysis (GGEA) procedure by Geistlinger *et al.* [76]. GGEA scores pathways according to the consistency of the measured expression patterns and known regulatory interactions. The consistency of the regulations is determined via a Petri Net [77] model and scored for significance using permutation tests. Glaab *et al.* published EnrichNet [78], which incorporates network data in a different fashion. Starting from a set of differentially expressed genes, so-called seed genes, the authors apply a random walk with restarts (RWR) method to calculate distance scores for the remaining nodes that measure the distance to the seed nodes. In order to select significantly enriched pathways the node scores are discretized into *m* equally sized bins B_j. The score of a pathway *P* is then given by:

$$d(P) = \sum_{j=1}^{m} \frac{f_{\mathrm{p}}(j) - f_{\mathrm{a}}(j)}{j \cdot m}$$

where $f_{\mathrm{p}}(j)$ is the percentage of gene scores of target genes that are found in P and are placed in bin j:

$$f_{\mathrm{p}}(j) = 100 \cdot \frac{|B_j \cap P|}{\sum_{j=1}^{m} |B_j \cap P|}$$

The quantity $f_{\mathrm{a}}(j)$ represents the expected percentage of genes in each bin across all pathways and thus encodes the background distribution.

In contrast to models taking only known interactions into account, it is possible to learn putative interactions from the data. An example for such an approach is SAM-GS proposed by Dinu *et al.* [79]. SAM-GS models the expression data as samples from a multivariate normal distribution and applies Dempster's test statistics [80] for scoring the distance between the means of the two groups.

14.6 Detecting Deregulated Networks and Pathways

The tremendous progress in information technology has facilitated the collection of enormous sets of molecular interactions and reactions in single databases. Consequently, modeling and analysis of these interactions using, for example, graph structures and networks is now a common task. Also, modern biotechnological high-throughput methods allow for measuring vast amounts of expression data. This data – combined with molecular networks – opens novel avenues for the analysis of complex biological and pathological processes.

A problem that has received major attention in the last 10 years is the detection of pathways and subnetworks that are dysregulated in pathogenic processes. Here, the input of the problem consists of:

- A PPI network or a regulatory network $G = (V, E)$ where the node set V represents genes or proteins and the edge set E the known interactions between the genes or proteins.
- A list $L = (d_1, \ldots, d_n)$ of real numbers d_i mirroring the degree of deregulation of gene/protein i, represented by the node v_i of the network G. In the following we will refer to d_i as the gene *score*.

The goal is to find the pathways or connected subnetworks in the network G that are maximally dysregulated. Here, the degree of deregulation of a pathway or subnetwork P is usually measured as a function of the degrees of the involved nodes (e.g., the sum of the degrees of the nodes contained in P):

$$d(P) = \sum_{v_i \in P} d_i$$

As we will see later, some approaches map weights mirroring the differential expression to the edges of the graph or they map suitable weights to both nodes and edges.

In 2002, Ideker *et al.* [81] presented the first approach for detecting "active subnetworks, i.e., connected sets of genes with unexpectedly high levels of differential expression." They considered networks that combined PPIs and protein–DNA (regulatory) interactions of yeast, and they formulated the computational challenge as follows: "What are the signaling and regulatory interactions in control of the observed gene expression changes? How is this control exerted?" Ideker *et al.* calculate for each gene i a P-value p_i using the program VERA [82] and convert this P-value into a z-score $z_i = \Phi^{-1}(1 - p_i)$, where Φ^{-1} is the inverse normal cumulative distribution function. Given a connected subnetwork $G' \subseteq G$ of size k, an aggregate z-score $z_{G'}$ for the subnetwork is calculated as:

$$z_{G'} = \frac{1}{\sqrt{k}} \sum_{v_i \in G'} z_i$$

The significance of such a z-score $z_{G'}$ is determined by a calibration against a background distribution of randomly sampled sets of genes of size k whose nodes need not induce a connected subgraph of G. Ideker *et al.* generated such background distributions using a Monte Carlo approach. If μ_k is the mean score of a sufficiently large collection of random gene sets and σ_k its standard deviation, then:

$$s_{G'} = \frac{z_{G'} - \mu_k}{\sigma_k}$$

can be used as a corrected subnet score. Based on this scoring scheme, Ideker *et al.* have developed a heuristic optimization approach for finding the maximal-scoring connected subgraph. Since this problem is NP-hard, they implemented a simulated annealing approach. The implementation was tested on a yeast knock-out study dataset and was able to recover well known regulatory paths from the employed PPI network. This approach is freely available in the jActiveModules Cytoscape [83] plug-in.

In 2005, Rajagopalan and Agarwal [84] proposed a modified scoring function and an effective heuristic based on the above work. These modifications avoid very large, uninterpretable solutions and lead to tremendous improvements of the algorithm's running time.

Ulitsky *et al.* [85] developed DEGAS – an algorithm that searches for minimal connected subnetworks in which the number of dysregulated genes in almost all diseased samples exceeds a given threshold. More precisely, the goal is to detect a minimal connected subnetwork with at least k nodes differentially expressed in all but l analyzed samples. Hence, the approach has two parameters that are easy to interpret: (i) k, the number of differentially expressed genes, and (ii) l the number of allowed outliers. Let $G = (V, E)$ be an undirected graph that represents, for example, a PPI network. Ulitsky *et al.* name a subset $V' \subseteq V$ a

connected (k, l)*-cover* if the subgraph induced by the node set V' is connected and fulfills the property defined above (i.e., the induced subgraph contains at least k nodes differentially expressed in all but l samples). For a given parameter pair (k, l), the goal is to find the smallest connected (k, l)-cover. The authors show that the problem can be formulated as a Connected Set Cover Problem [86] and present some practical heuristics for this NP-hard problem that have provable performance guarantees. Their evaluations on various disease datasets (including Parkinson's disease and tongue cancer) report the consistently higher ability of DEGAS to retrieve known, disease-associated KEGG pathways than their competitors.

Most approaches for detecting deregulated subnetworks search for a maximum weight connected subgraph of a given size k (i.e., the goal is to find the connected subgraph $G' \subseteq G$ of size k where the sum of the weights of the nodes (or edges) is maximal). Dittrich *et al.* [87] provided the first algorithm able to solve the problem of finding a maximum weight connected subgraph to optimality if the considered problem instances are not too large. They transform the problem into an equivalent prize-collection Steiner Tree problem and apply an existing, efficient ILP (integer linear programming) solution [88]. An implementation of the algorithm is available in the BioNet package [89] for the R environment [90]. Another ILP approach has been proposed by Backes *et al.* [91]. In contrast to the BioNet approach, it considers a directed graph representing, for example, a regulatory network. The approach aims at finding maximum weight connected subgraphs G' where all nodes $v' \in G'$ are reachable from a designated root node v_R that also belongs to G'(*cf.* Figure 14.4). This root node could be a molecular key player that might be "responsible" for the observed differential deregulation of the detected subgraph. Moreover, the corresponding gene or protein may represent an ideal target molecule for a therapy that aims at reversing and normalizing the malignant dysregulated processes.

Let $L = (d_1, \ldots, d_n) \in \mathbb{R}^n$ be the vector of gene scores, $X = (x_1, \ldots, x_n) \in \mathbb{B}^n$ a vector with binary variables indicating the choice of subgraph vertices ($x_i = 1$ if v_i is selected and 0 otherwise), and $Y = (y_1, \ldots, y_n) \in \mathbb{B}^n$ a vector with binary variables indicating the choice of the root node ($y_i = 1$ if v_i is the selected root node and 0 otherwise). The complete ILP of Backes *et al.* [91] with the objective function and all constraints is given below:

$$\max_{X,Y \in B^n} \quad \sum_i d_i x_i \quad \text{s.t.} \tag{14.3}$$

$$\sum_i x_i = k \tag{14.4}$$

$$\sum_i y_i = 1 \tag{14.5}$$

$$x_i - y_i - \sum_{j \in \mathrm{In}(\{i\})} x_j \leq 0, \forall_i \tag{14.6}$$

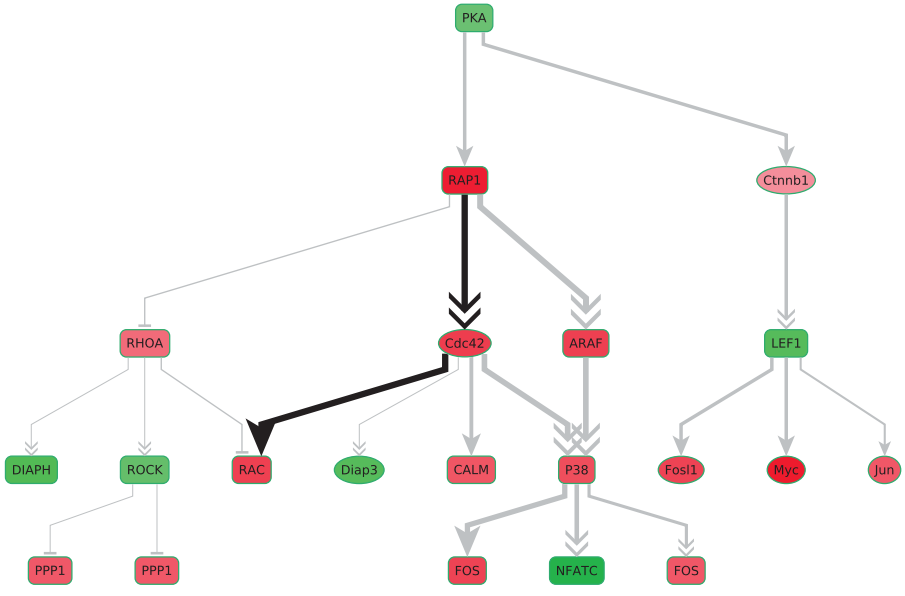


Figure 14.4 The resulting subgraphs computed by the ILP approach of Backes *et al.* [91] applied to the GDS2161 and GDS2162 GEO data sets. The subgraph size ranges from 3 to 20. Red denotes overexpression and green denotes underexpression.

$$\sum_{i \in \mathcal{C}} (x_i - y_i) - \sum_{j \in \mathrm{In}(\mathcal{C})} x_j \leq |\mathcal{C}| - 1, \forall_{\mathcal{C}} \tag{14.7}$$

$$y_i \leq x_i, \forall_i \tag{14.8}$$

Equations (14.4)–(14.5) ensure that exactly k vertices are part of the solution and that exactly one root node is chosen. By the inequalities (Equation 14.6) we ensure for each selected vertex that it is either the root node or that at least one of its parent nodes (parents are given by the set In(.)) has been selected. As, for example, subgraphs consisting of two disconnected cycles fulfill these equations, the inequalities (Equation 14.7) have to be added to disallow these cases explicitly. These constraints ensure that selected cycles $\mathcal{C}$ either contain the root node or that a selected vertex outside of $\mathcal{C}$ with an edge to a node in $\mathcal{C}$ exists. Inequalities (Equation 14.8) ensure that the root node has to be one of the selected nodes. Based on the above ILP formulation, Backes *et al.* [91] have developed and implemented a Branch&Cut algorithm using CPLEX 12 [92]. This Branch&Cut algorithm is available as a stand-alone executable or through a web interface [93].

The HotNet algorithm [94] searches for deregulated subgraphs in a graph derived from the input network. Every edge (s, v) in the so-called *influence graph*, reflects the influence of a vertex s on a vertex v. This influence is computed by simulating how information, which is being pumped into s at a

constant rate, diffuses across the graph. Let $f_s(t) = (f_{sv_1}(t), \ldots, f_{sv_N}(t))$ be the amount of information available in the nodes v_i at time t, and b_s be a vector with 1 on the sth position and 0 otherwise. The diffusion process is given by:

$$\frac{\mathrm{d}f_s(t)}{\mathrm{d}t} = -(L + \gamma I)f_s(t) + b_s u(t) \tag{14.9}$$

where γ is some positive constant and $u(t)$ is the Heaviside step function. The graph Laplacian L is defined as $L := D - A$, where D is the diagonal matrix of vertex degrees and A is the adjacency matrix of the graph. The steady-state solution of Equation (14.9) is only dependent on b_s and given by:

$$f_s^* = (L + \gamma I)^{-1} b_s.$$

Solving this for every vertex, the final adjacency matrix of the influence graph w is then defined as $w(i,j) = \min(|f_{ij}^*|, |f_{ji}^*|)$. The authors propose two heuristics for finding maximal subnetworks in the influence graph. The first heuristics approximates the (connected) maximum coverage problem, whilst the second heuristics uses a simple thresholding strategy. Note that the detected subnetworks need not be connected in the original graph. HotNet was originally designed to work with single nucleotide polymorphism data; however, a generalized version taking arbitrary node scores as input is also available.

Dao *et al.* [95] propose a randomized algorithm that features a polynomial worst-case running time. They use color-coding to obtain a subnetwork that is "optimally discriminative" between two states with high probability. Color-coding is an algorithmic technique that was first applied for the detection of paths and cycles of predefined length k [96]. The basic idea behind color-coding is to randomly assign one of k colors to every vertex. Then, starting from every vertex, colorful paths (paths that contain every color exactly once) are searched. Finding a colorful path is easy, as, in contrast to standard path search algorithms, not the visited vertices, but only the found colors need to be tracked. This random search is repeated until the probability of having missed a path falls below a predefined threshold. Dao *et al.* employ the color-coding scheme not for finding paths, but for detecting the maximum colorful subgraph using a dynamic programming algorithm. The colorful subgraph with the highest score is reported as the final result of the algorithm. Let δ be the probability of not discovering maximum scoring subgraph. The required number of iterations n_i for achieving δ is bound by:

$$n_i \leq \ln(1/\delta)\mathrm{e}^k$$

In contrast to the above algorithms that detects maximally dysregulated subgraphs, the FiDePa algorithm [97] searches for paths of a given length k that maximize the running sum score of the unweighted GSEA procedure (see Equation 14.2). The algorithm assigns a rank to every node according to its differential expression and considers all paths of length k as competing categories. Using

an efficient dynamic programming approach, it calculates the most significant paths of length k. If the considered graph G is a directed acyclic graph, the algorithm guarantees optimality. Otherwise, if the graph contains cycles, they will be disregarded by the dynamic programming approach and hence in this case the algorithm may provide suboptimal solutions. Moreover, multiple suboptimal solutions can be computed if desired. Being based on unweighted GSEA it is possible to compute exact P-values for the discovered paths [69]. Cabusora *et al.* [98] presented a related approach that is also based on the identification of optimal paths that are subsequently unified to subnetworks. Starting from a set of seed nodes, the authors identify the shortest paths between them using Dijkstra's algorithm. Additional paths are added according to the k-shortest paths definition. Excessively long paths are excluded by an additional constraint. The resulting subnetwork is then given by the union of all identified paths and is subsequently scored for significance following the concepts of Ideker *et al.* [81].

14.7 miRNA Expression Data

The development of RNA-Seq and specific miRNA microarrays allows us to measure the expression of all known miRNAs simultaneously. Analyzing and interpreting this data is, however, even more challenging than mRNA expression profiles. First, only few miRNAs are functionally annotated, making it difficult to interpret the effect of the deregulation of most differentially expressed miRNAs. Second, miRNA genes have yet to find their way into regulatory network databases. The main reason for this, besides the relative youth of the field, is how miRNAs bind to their targets. The recognition of target mRNAs via their 3′ untranslated region allows for mismatches and imperfections, usually only relying on perfect complementarity of bases 2–7, the so-called *seed* region [99]. This introduces a significant amount of fuzziness into the target recognition of miRNA, which often manifests as off-target effects due to "random" seed region matching [100]. Hence, the reliable detection of miRNA targets is an important challenge in computational biology. Commonly used methods for this task are based on RNA–RNA interaction prediction [101,102] or machine learning [103]. In this context, paired miRNA–mRNA expression profiling has proved to have great potential for the identification of novel interactions [104–106].

More important for the purpose of diagnostics is the question of functional annotation of miRNAs. For this purpose, target prediction methods can be used. The basic idea here follows the guilt-by-association principle: assigning the function the target is annotated with to the query miRNA solely based on their interaction.

Similarly, it is possible to use (predicted) miRNA targets as a proxy for detecting enriched annotations. By mapping differentially expressed miRNAs into the pathways and categories of their targets, methods like ORA and GSEA (see Section 14.5) can easily be carried over to the miRNA scenario. This method,

besides other resources, is provided in the miRGator database [107]. Laczny *et al.* expand on this idea by requiring that the miRNA and mRNA expression pattern in a given dataset is consistent. Here consistent means that the deregulation of the mRNA and miRNA are dependent as determined using the χ^2-test [108]. The resulting miRTrail procedure is available as a Web server.

14.8 Differential Network Analysis

So far we have covered methods that implicitly assume that the topology of the underlying interaction networks remain unchanged between control and sample. For cancer datasets this is not true. Mutations in cancer cells can cause genes to lose or gain the ability to bind targets. In terms of networks, such mutations result in a rewiring or reprogramming of the cellular interaction network. Traditional expression profile analysis cannot capture these rewiring events, as the expression of a gene alone is not indicative of its interactions with other genes. Also, RNA-Seq profiles only partially solve this problem as the function of a gene may be affected by mutations in non-coding regions that are difficult to interpret.

If, however, for both states – control and diseased – an interaction network were to be determined, we could easily identify novel or missing interactions by comparing the edges of both networks. The analysis not of the expression changes of a single gene, but rather the study of the changing interactions of nodes in a network is called *differential network analysis*. Ideker and Krogan [109] discuss differential network analysis approaches based on specialized interaction screens. While specialized assays are, no doubt, more powerful than generic expression studies, we restrict ourselves to using data from the latter, which are more commonly performed in diagnostic settings. Probably the simplest approach to differential network analysis is differential coexpression analysis as discussed by de la Fuente [110]. For both groups, Pearson's correlation is computed for every pair of genes. These correlations are compared across the groups and significant differences are reported. However, plain correlations can be difficult to interpret in the presence of confounding factors. Let us assume that genes A and B are regulated by C. We would observe a high correlation between A and B, although no interaction between the two exists.

For this reason, recent methods based on Gaussian graphical models [111] and especially the *graphical lasso* algorithm [112] have become popular. Oversimplifying the idea of the graphical lasso is to construct networks based on penalized partial correlations. This means that when computing the correlation between two variables the influences of all other variables are accounted for and small correlation values are likely to be set to zero. Oyen *et al.* [113] present a version of the graphical lasso that allows the estimation of two networks, one for each group, simultaneously. The key idea behind this approach is that, in general, most interactions remain intact and only a few, prominent differences manifest

in the predicted interaction networks. Parikh *et al.* have proposed a similar procedure called TreeGL [114], which is capable of estimating a whole hierarchy of networks.

As only sparse resources concerning miRNA interaction and regulatory networks are available, the idea of network inference is especially useful for the analysis of miRNA datasets. Volinia *et al.* extract Bayesian networks from miRNA expression data for reference and cancer patients, and demonstrate that regulatory reprogramming also affects regulations mediated by miRNA [115].

Finally, we would like to remind the reader that reliable network inference is only possible for small gene sets and large datasets. This of course makes an application in a diagnostic setting, where only one or few samples per patient are available, difficult.

14.9 Conclusion

We have presented methods for detecting dysregulated processes and pathways. Building upon the success of simple statistical methods like ORA and GSEA, a large variety of powerful methods has been developed that are tailored towards different types of data and use cases. In particular, network-based approaches show promise for incorporation into diagnostics, as they give concrete hints towards the cause of a disease. However, despite the availability of many high-quality methods, deployment in a diagnostics workflow is a non-trivial task due to, on the one hand, technical hurdles such as experimental protocols and software usability and, on the other hand, acceptance problems of predictions from a "black box" method. In order to overcome these difficulties and enable the creation of a modern, efficient system for diagnosis, continued coordination and collaboration between medicine and computational biology is mandatory.

References

1 Van De Vijver, M.J., He, Y.D., van't Veer, L.J., Dai, H., Hart, A.A., Voskuil, D.W., Schreiber, G.J., Peterse, J.L., Roberts, C., Marton, M.J. *et al.* (2002) A gene-expression signature as a predictor of survival in breast cancer. *New England Journal of Medicine*, **347** (25), 1999–2009.

2 Spira, A., Beane, J.E., Shah, V., Steiling, K., Liu, G., Schembri, F., Gilman, S., Dumas, Y.M., Calner, P., Sebastiani, P. *et al.* (2007) Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nature Medicine*, **13** (3), 361–366.

3 Dhanasekaran, S.M., Barrette, T.R., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K.J., Rubin, M.A., and Chinnaiyan, A.M. (2001) Delineation of prognostic biomarkers in prostate cancer. *Nature*, **412** (6849), 822–826.

4 Baker, M. (2005) In biomarkers we trust? *Nature Biotechnology*, **23** (3), 297–304.

5 Fielden, M.R., Brennan, R., and Gollub, J. (2007) A gene expression biomarker provides early prediction and mechanistic assessment of hepatic tumor induction by

nongenotoxic chemicals. *Toxicological Sciences*, **99** (1), 90–100.
6 Ge, Y., Zhao, K., Qi, Y., Min, X., Shi, Z., Qi, X., Shan, Y., Cui, L., Zhou, M., Wang, Y. *et al.* (2013) Serum microRNA expression profile as a biomarker for the diagnosis of pertussis. *Molecular Biology Reports*, **40** (2), 1325–1332.
7 Chai, J.C., Park, S., Seo, H., Cho, S.Y., and Lee, Y.S. (2013) Identification of cancer-specific biomarkers by using microarray gene expression profiling. *BioChip Journal*, **7** (1), 57–62.
8 Bellman, R. (2003) *Dynamic Programming*, Dover, New York.
9 Cox, J. and Mann, M. (2011) Quantitative, high-resolution proteomics for data-driven systems biology. *Annual Review of Biochemistry*, **80**, 273–299.
10 Vogel, C. and Marcotte, E.M. (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics*, **13** (4), 227–232.
11 Ozsolak, F. and Milos, P.M. (2010) RNA sequencing: advances, challenges and opportunities. *Nature Reviews Genetics*, **12** (2), 87–98.
12 Kogenaru, S., Yan, Q., Guo, Y., and Wang, N. (2012) RNA-Seq and microarray complement each other in transcriptome profiling. *BMC Genomics*, **13** (1), 629.
13 Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. (2008) RNA-Seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*, **18** (9), 1509–1517.
14 Pritchard, C.C., Cheng, H.H., and Tewari, M. (2012) MicroRNA profiling: approaches and considerations. *Nature Reviews Genetics*, **13** (5), 358–369.
15 Wang, Z., Gerstein, M., and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, **10** (1), 57–63.
16 Kotake, Y., Nakagawa, T., Kitagawa, K., Suzuki, S., Liu, N., Kitagawa, M., and Xiong, Y. (2010) Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of $p15^{INK4B}$ tumor suppressor gene. *Oncogene*, **30** (16), 1956–1962.
17 Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.C., Hung, T., Argani, P., Rinn, J.L. *et al.* (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*, **464** (7291), 1071–1076.
18 He, L., Thomson, J.M., Hemann, M.T., Hernando-Monge, E., Mu, D., Goodson, S., Powers, S., Cordon-Cardo, C., Lowe, S. W., Hannon, G.J. *et al.* (2005) A microRNA polycistron as a potential human oncogene. *Nature*, **435** (7043), 828–833.
19 Kozomara, A. and Griffiths-Jones, S. (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research*, **39** (Suppl. 1), D152–D157.
20 Lee, Y.S. and Dutta, A. (2009) MicroRNAs in cancer. *Annual Review of Pathology*, **4**, 199.
21 Keller, A., Leidinger, P., Bauer, A., ElSharawy, A., Haas, J., Backes, C., Wendschlag, A., Giese, N., Tjaden, C., Ott, K. *et al.* (2011) Toward the blood-borne miRNome of human diseases. *Nature Methods*, **8** (10), 841–843.
22 Hébert, S.S., Horré, K., Nicolaï, L., Papadopoulou, A.S., Mandemakers, W., Silahtaroglu, A.N., Kauppinen, S., Delacourte, A., and De Strooper, B. (2008) Loss of microRNA cluster miR-29a/b-1 in sporadic Alzheimer's disease correlates with increased BACE1/β-secretase expression. *Proceedings of the National Academy of Sciences*, **105** (17), 6415–6420.
23 Grimmett, G. and Stirzaker, D. (2001) *Probability and Random Processes*, Oxford University Press, Oxford.
24 Bishop, C.M. and Nasrabadi, N.M. (2006) *Pattern Recognition and Machine Learning*, 1st edn, Springer, New York.
25 Hastie, T., Tibshirani, R., and Friedman, J.J.H. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edn, vol. **1**, Springer, New York.
26 Dillies, M.A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M.,

Servant, N., Keime, C., Marot, G., Castel, D., Estelle, J. *et al.* (2012) A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in Bioinformatics*, **14** (6), 671–683.

27 Li, H. and Homer, N. (2010) A survey of sequence alignment algorithms for next-generation sequencing. *Briefings in Bioinformatics*, **11** (5), 473–483.

28 Affymetrix, Inc. Gene 1.0 ST Array Design. http://www.affymetrix.com/support/technical/technotes/gene_1_0_st_technote.pdf.

29 Mei, R., Di, X., Ryder, T., Hubbell, E., Dee, S., Webster, T., Harrington, C., Ho, M.H., Baid, J., Smeekens, S. *et al.* (2002) Analysis of high density expression microarrays with signed-rank call algorithms. *Bioinformatics*, **18** (12), 1593–1599.

30 Naef, F., Hacker, C.R., Patil, N., and Magnasco, M. (2002) Empirical characterization of the expression ratio noise structure in high-density oligonucleotide arrays. *Genome Biology*, **3** (4), research0018.

31 Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, **4** (2), 249–264.

32 Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B., and Speed, T.P. (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Research*, **31** (4), e15–e15.

33 Huber, W., Von Heydebreck, A., Sültmann, H., Poustka, A., and Vingron, M. (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, **18** (Suppl. 1), S96–S104.

34 Huber, W., von Heydebreck, A., and Vingron, M. (2007) Low-level analysis of microarray experiments, in *Bioinformatics –From Genomes to Therapies* (ed. T. Lengauer), Wiley-VCH, Weinheim.

35 Benito, M., Parker, J., Du, Q., Wu, J., Xiang, D., Perou, C.M., and Marron, J.S. (2004) Adjustment of systematic microarray data biases. *Bioinformatics*, **20** (1), 105–114.

36 Leek, J.T. and Storey, J.D. (2007) Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genetics*, **3** (9), e161.

37 Johnson, W.E., Li, C., and Rabinovic, A. (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, **8** (1), 118–127.

38 Gagnon-Bartsch, J.A. and Speed, T.P. (2012) Using control genes to correct for unwanted variation in microarray data. *Biostatistics*, **13** (3), 539–552.

39 Lazar, C., Meganck, S., Taminau, J., Steenhoff, D., Coletta, A., Molter, C., Weiss-Sols, D.Y., Duque, R., Bersini, H., and Nowé, A. (2012) Batch effect removal methods for microarray gene expression data integration: a survey. *Briefings in Bioinformatics*, **14** (4), 469–490.

40 Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T. *et al.* (2000) Gene Ontology: tool for the unification of biology. *Nature Genetics*, **25** (1), 25–29.

41 Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research*, **40** (D1), D109–D114.

42 Matthews, L., Gopinath, G., Gillespie, M., Caudy, M., Croft, D., de Bono, B., Garapati, P., Hemish, J., Hermjakob, H., Jassal, B. *et al.* (2009) Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Research*, **37** (Suppl. 1), D619–D622.

43 Bader, G.D., Betel, D., and Hogue, C.W. (2003) BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Research*, **31** (1), 248–250.

44 Salwinski, L., Miller, C.S., Smith, A.J., Pettit, F.K., Bowie, J.U., and Eisenberg, D. (2004) The Database of Interacting Proteins: 2004 update. *Nucleic Acids Research*, **32** (Suppl. 1), D449–D451.

45 Hoffmann, R., Krallinger, M., Andres, E., Tamames, J., Blaschke, C., and Valencia, A. (2005) Text mining for metabolic pathways, signaling cascades, and protein networks. *Science Signaling*, **2005** (283), pe21.

46 Özgür, A., Vu, T., Erkan, G., and Radev, D.R. (2008) Identifying gene-disease associations using centrality on a literature mined gene-interaction network. *Bioinformatics*, **24** (13), i277–i285.

47 Donaldson, I., Martin, J., De Bruijn, B., Wolting, C., Lay, V., Tuekam, B., Zhang, S., Baskin, B., Bader, G.D., Michalickova, K. *et al.* (2003) PreBIND and Textomy – mining the biomedical literature for protein–protein interactions using a support vector machine. *BMC Bioinformatics*, **4** (1), 11.

48 Brown, K.R. and Jurisica, I. (2005) Online Predicted Human Interaction Database. *Bioinformatics*, **21** (9), 2076–2082.

49 McDowall, M.D., Scott, M.S., and Barton, G.J. (2009) PIPS: human Protein–Protein Interaction Prediction database. *Nucleic Acids Research*, **37** (Suppl. 1), D651–D656.

50 Zhang, Q.C., Petrey, D., Deng, L., Qiang, L., Shi, Y., Thu, C.A., Bisikirska, B., Lefebvre, C., Accili, D., Hunter, T. *et al.* (2012) Structure-based prediction of protein–protein interactions on a genome-wide scale. *Nature*, **490** (7421), 556–560.

51 Jansen, R., Yu, H., Greenbaum, D., Kluger, Y., Krogan, N.J., Chung, S., Emili, A., Snyder, M., Greenblatt, J.F., and Gerstein, M. (2003) A Bayesian networks approach for predicting protein–protein interactions from genomic data. *Science*, **302** (5644), 449–453.

52 Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., Doerks, T., Stark, M., Muller, J., Bork, P. *et al.* (2011) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Research*, **39** (Suppl. 1), D561–D568.

53 Küntzer, J., Blum, T., Gerasch, A., Backes, C., Hildebrandt, A., Kaufmann, M., Kohlbacher, O., and Lenhof, H.P. (2006) BN++ – a biological information system. *Journal of Integrative Bioinformatics*, **3** (2), 34.

54 Köhler, J., Baumbach, J., Taubert, J., Specht, M., Skusa, A., Rüegg, A., Rawlings, C., Verrier, P., and Philippi, S. (2006) Graph-based analysis and visualization of experimental results with ONDEX. *Bioinformatics*, **22** (11), 1383–1390.

55 Opgen-Rhein, R., Strimmer, K. *et al.* (2007) Accurate ranking of differentially expressed genes by a distribution-free shrinkage approach. *Statistical Applications in Genetics and Molecular Biology*, **6** (1), 9.

56 Wang, L., Feng, Z., Wang, X., Wang, X., and Zhang, X. (2010) DEGseq: an R package for identifying differentially expressed genes from RNA-Seq data. *Bioinformatics*, **26** (1), 136–138.

57 Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010) edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26** (1), 139–140.

58 Anders, S. and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biology*, **11** (10), R106.

59 Turro, E., Su, S.Y., Gonçalves, Â., Coin, L., Richardson, S., and Lewin, A. (2011) Haplotype and isoform specific expression estimation using multi-mapping RNA-Seq reads. *Genome Biology*, **12** (2), R13.

60 Li, B. and Dewey, C. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, **12** (1), 323.

61 Roberts, A., Trapnell, C., Donaghey, J., Rinn, J.L., Pachter, L. *et al.* (2011) Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biology*, **12** (3), R22.

62 Glaus, P., Honkela, A., and Rattray, M. (2012) Identifying differentially expressed transcripts from RNA-Seq data with biological variation. *Bioinformatics*, **28** (13), 1721–1728.

63 Kvam, V.M., Liu, P., and Si, Y. (2012) A comparison of statistical methods for detecting differentially expressed genes from RNA-Seq data. *American Journal of Botany*, **99** (2), 248–256.

64 Soneson, C. and Delorenzi, M. (2013) A comparison of methods for differential expression analysis of RNA-Seq data. *BMC Bioinformatics*, **14** (1), 91.

65 Wesolowski, S., Birtwistle, M.R., and Rempala, G.A. (2013) A comparison of methods for RNA-Seq differential expression analysis and a new empirical Bayes approach. *Biosensors*, **3** (3), 238–258.

66 Drăghici, S., Khatri, P., Martins, R.P., Ostermeier, G.C., and Krawetz, S.A. (2003) Global functional profiling of gene expression. *Genomics*, **81** (2), 98–104.

67 Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B.L., Gillette, M. A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, **102** (43), 15 545–15 550.

68 Massey, F.J. Jr. (1951) The Kolmogorov–Smirnov test for goodness of fit. *Journal of the American statistical Association*, **46** (253), 68–78.

69 Keller, A., Backes, C., and Lenhof, H.P. (2007) Computation of significance scores of unweighted gene set enrichment analyses. *BMC Bioinformatics*, **8** (1), 290.

70 Ackermann, M. and Strimmer, K. (2009) A general modular framework for gene set enrichment analysis. *BMC Bioinformatics*, **10** (1), 47.

71 Nam, D. and Kim, S.Y. (2008) Gene-set approach for expression pattern analysis. *Briefings in Bioinformatics*, **9** (3), 189–197.

72 Song, S. and Black, M.A. (2008) Microarray-based gene set analysis: a comparison of current methods. *BMC Bioinformatics*, **9** (1), 502.

73 Dunn, O.J. (1961) Multiple comparisons among means. *Journal of the American Statistical Association*, **56** (293), 52–64.

74 Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)*, 289–300.

75 Roback, P.J. and Askins, R.A. (2005) Judicious use of multiple hypothesis tests. *Conservation Biology*, **19** (1), 261–267.

76 Geistlinger, L., Csaba, G., Küffner, R., Mulder, N., and Zimmer, R. (2011) From sets to graphs: towards a realistic enrichment analysis of transcriptomic systems. *Bioinformatics*, **27** (13), i366–i373.

77 Petri, C.A. (1962) Kommunikation mit Automaten, PhD thesis, Institut für Instrumentelle Mathematik, Bonn.

78 Glaab, E., Baudot, A., Krasnogor, N., Schneider, R., and Valencia, A. (2012) EnrichNet: network-based gene set enrichment analysis. *Bioinformatics*, **28** (18), i451–i457.

79 Dinu, I., Potter, J.D., Mueller, T., Liu, Q., Adewale, A.J., Jhangri, G.S., Einecke, G., Famulski, K.S., Halloran, P., and Yasui, Y. (2007) Improving gene set analysis of microarray data by SAM-GS. *BMC Bioinformatics*, **8** (1), 242.

80 Dempster, A.P. (1958) A high dimensional two sample significance test. *Annals of Mathematical Statistics*, **29** (4), 995–1010.

81 Ideker, T., Ozier, O., Schwikowski, B., and Siegel, A.F. (2002) Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics*, **18** (Suppl. 1), S233–240.

82 Ideker, T., Thorsson, V., Siegel, A.F., and Hood, L.E. (2000) Testing for differentially-expressed genes by maximum-likelihood analysis of microarray data. *Journal of Computational Biology*, **7** (6), 805–817.

83 Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, **13** (11), 2498–2504.

84 Rajagopalan, D. and Agarwal, P. (2005) Inferring pathways from gene lists using a

literature-derived network of biological relationships. *Bioinformatics*, **21** (6), 788–793.

85 Ulitsky, I., Krishnamurthy, A., Karp, R.M., and Shamir, R. (2010) DEGAS: *de novo* discovery of dysregulated pathways in human diseases. *PLoS One*, **5** (10), e13 367.

86 Shuai, T.P. and Hu, X.D. (2006) Connected set cover problem and its applications, in *Algorithmic Aspects in Information and Management*, Springer, Berlin, pp. 243–254.

87 Dittrich, M.T., Klau, G.W., Rosenwald, A., Dandekar, T., and Müller, T. (2008) Identifying functional modules in protein–protein interaction networks: an integrated exact approach. *Bioinformatics*, **24** (13), i223–i231.

88 Ljubić, I., Weiskircher, R., Pferschy, U., Klau, G.W., Mutzel, P., and Fischetti, M. (2006) An algorithmic framework for the exact solution of the prize-collecting Steiner Tree problem. *Mathematical Programming*, **105** (2–3), 427–449.

89 Beisser, D., Klau, G.W., Dandekar, T., Müller, T., and Dittrich, M.T. (2010) BioNet: an R-package for the functional analysis of biological networks. *Bioinformatics*, **26** (8), 1129–1130.

90 R Development Core Team (2011) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna.

91 Backes, C., Rurainski, A., Klau, G.W., Müller, O., Stöckel, D., Gerasch, A., Küntzer, J., Maisel, D., Ludwig, N., Hein, M. *et al.* (2012) An integer linear programming approach for finding deregulated subgraphs in regulatory networks. *Nucleic Acids Research*, **40** (6), e43–e43.

92 IBM. CPLEX Optimizer: High-performance mathematical programming solver for linear programming, mixed integer programming, and quadratic programming. http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/index.html.

93 Stöckel, D., Müller, O., Kehl, T., Gerasch, A., Backes, C., Rurainski, A., Keller, A., Kaufmann, M., and Lenhof, H.P. (2013) NetworkTrail – a web service for identifying and visualizing deregulated subnetworks. *Bioinformatics*, **29** (13), 1702–1703.

94 Vandin, F., Upfal, E., and Raphael, B.J. (2011) Algorithms for detecting significantly mutated pathways in cancer. *Journal of Computational Biology*, **18** (3), 507–522.

95 Dao, P., Wang, K., Collins, C., Ester, M., Lapuk, A., and Sahinalp, S.C. (2011) Optimally discriminative subnetwork markers predict response to chemotherapy. *Bioinformatics*, **27** (13), i205–i213.

96 Alon, N., Yuster, R., and Zwick, U. (1994) Color-coding: a new method for finding simple paths, cycles and other small subgraphs within large graphs, in Proceedings of the 26h Annual ACM Symposium on Theory of Computing, ACM, pp. 326–335.

97 Keller, A., Backes, C., Gerasch, A., Kaufmann, M., Kohlbacher, O., Meese, E., and Lenhof, H.P. (2009) A novel algorithm for detecting differentially regulated paths based on gene set enrichment analysis. *Bioinformatics*, **25** (21), 2787–2794.

98 Cabusora, L., Sutton, E., Fulmer, A., and Forst, C.V. (2005) Differential network expression during drug and stress response. *Bioinformatics*, **21** (12), 2898–2905.

99 Maziere, P. and Enright, A.J. (2007) Prediction of microRNA targets. *Drug Discovery Today*, **12** (11), 452–458.

100 Hibio, N., Hino, K., Shimizu, E., Nagata, Y., and Ui-Tei, K. (2012) Stability of miRNA 5′ terminal and seed regions is correlated with experimentally observed miRNA-mediated silencing efficacy. *Scientific Reports*, **2**, 996.

101 Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., Marks, D.S. *et al.* (2004) MicroRNA targets in *Drosophila*. *Genome Biology*, **5** (1), R1.

102 Ragan, C., Zuker, M., and Ragan, M.A. (2011) Quantitative prediction of miRNA–mRNA interaction based on equilibrium concentrations. *PLoS Computational Biology*, **7** (2), e1001090.

103 Wang, X. and El Naqa, I.M. (2008) Prediction of both conserved and

nonconserved microRNA targets in animals. *Bioinformatics*, **24** (3), 325–332.

104 Huang, J.C., Babak, T., Corson, T.W., Chua, G., Khan, S., Gallie, B.L., Hughes, T.R., Blencowe, B.J., Frey, B.J., and Morris, Q.D. (2007) Using expression profiling data to identify human microRNA targets. *Nature Methods*, **4** (12), 1045–1049.

105 Lu, Y., Zhou, Y., Qu, W., Deng, M., and Zhang, C. (2011) A lasso regression model for the construction of microRNA-target regulatory networks. *Bioinformatics*, **27** (17), 2406–2413.

106 Engelmann, J.C. and Spang, R. (2012) A least angle regression model for the prediction of canonical and non-canonical miRNA–mRNA interactions. *PloS One*, **7** (7), e40 634.

107 Cho, S., Jang, I., Jun, Y., Yoon, S., Ko, M., Kwon, Y., Choi, I., Chang, H., Ryu, D., Lee, B. *et al.* (2013) miRGator v3.0: a microRNA portal for deep sequencing, expression profiling and mRNA targeting. *Nucleic Acids Research*, **41** (D1), D252–D257.

108 Laczny, C., Leidinger, P., Haas, J., Ludwig, N., Backes, C., Gerasch, A., Kaufmann, M., Vogel, B., Katus, H.A., Meder, B. *et al.* (2012) miRTrail – a comprehensive webserver for analyzing gene and miRNA patterns to enhance the understanding of regulatory mechanisms in diseases. *BMC Bioinformatics*, **13** (1), 36.

109 Ideker, T. and Krogan, N.J. (2012) *Differential network biology. Molecular systems biology*, **8** (1).

110 de la Fuente, A. (2010) From 'differential expression' to 'differential networking' – identification of dysfunctional regulatory networks in diseases. *Trends in Genetics*, **26** (7), 326–333.

111 Lauritzen, S.L. (1996) *Graphical Models*, Oxford University Press, Oxford.

112 Friedman, J., Hastie, T., and Tibshirani, R. (2008) Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, **9** (3), 432–441.

113 Oyen, D., Niculescu-Mizil, A., Ostroff, R., and Stewart, A. (2013) Controlling the precision-recall tradeoff in differential dependency network analysis. *CoRR*, abs/1307.2611.

114 Parikh, A.P., Wu, W., Curtis, R.E., and Xing, E.P. (2011) TreeGL: reverse engineering tree-evolving gene networks underlying developing biological lineages. *Bioinformatics*, **27** (13), i196–i204.

115 Volinia, S., Galasso, M., Costinean, S., Tagliavini, L., Gamberoni, G., Drusco, A., Marchesini, J., Mascellani, N., Sana, M.E., Jarour, R.A. *et al.* (2010) Reprogramming of miRNA networks in cancer and leukemia. *Genome Research*, **20** (5), 589–599.

15
Companion Diagnostics and Beyond – An Essential Element in the Puzzle of Transforming Healthcare

Jan Kirsten

15.1
Introduction

Nucleic acid-based diagnostics testing is a strongly emerging field and has shown very promising results. With these new inventions, the possibility that therapeutic problems (e.g., low efficacy and side-effects) can be solved by genomics and related biomarker testing is becoming a reality. The use of nucleic acid-based diagnostic tests as guiding elements for a specific drug has become the essential element of companion diagnostics (CDx). An example would be the selection of the best-responding patient population by a genetic marker for a specific therapeutic agent. This scientific revolution has become day-to-day medical reality and a promising solution for financially stretched healthcare systems. As this chapter will show, the benefit of nucleic acid-based CDx tests is twofold – it results in improved medical outcome and at the same time in lower healthcare costs due to personalized targeted treatment.

15.2
The Healthcare Environment

A series of fundamental, interlinked changes dominate the global healthcare agenda. The needs and costs of global healthcare will continue to grow over the next few decades. On the demand side, the population is aging in developed countries, emerging markets are increasingly establishing access to healthcare, and the convergence of lifestyles around the globe will continue to shift disease profiles toward chronic diseases. On the cost side, a substantial share of the growth in spending in developed countries is attributable to incremental technological innovation and health system inefficiencies. In emerging countries, primarily macroeconomic factors (GDP and population increases) are driving spending upwards.

Nucleic Acids as Molecular Diagnostics, First Edition. Edited by Andreas Keller and Eckart Meese.

Governments around the world are trying to counter these developments by reviewing their roles in the healthcare market. Specifically, they are empowering national regulators to verify cost-benefits of new products. Mechanisms are being implemented to review the cost-opportunity of products such as health technology assessments, formulary reviews, budget neutrality thresholds, and so on.

To succeed in this new cost and quality control environment, healthcare companies need to understand how to leverage technological breakthroughs that promise to address previously unmet medical needs while creating value propositions. This strategy should be structured on holistic integrated solutions that improve effective utilization of medical technologies and services by leveraging existing and new diagnostic tests. CDx is an essential element in this context.

15.3 What is Companion Diagnostics?

In general, diagnostics have been suggested as a way to reduce overall healthcare costs (e.g., through earlier detection of disease or risk proposition). The traditional application for *in vitro* diagnostic (IVD) testing has been clinical diagnosis. In fact, it is estimated that such testing comprises 70% or more of the data used to diagnose diseases and medical conditions. However, IVDs are increasingly being utilized for screening, where test results provide physicians with objective feedback that can be incorporated into clinical decision making for the purpose of improving outcomes. In this context, CDx has to be seen has an evolutionary continuance of IVD tests with the goal to ensure the optimal use of a drug as part of therapy.

Today, there is no uniform definition for CDx. A variety of well-known institutions have defined it differently:

> *An IVD companion diagnostic device is an* in vitro *diagnostic device that provides information that is essential for the safe and effective use of a corresponding therapeutic product.* (US Food and Drug Administration (FDA): *In Vitro* Companion Diagnostics Draft Guidance, 2011, http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm262292.htm (accessed August 7, 2013).)

> *Companion diagnostics are tests that provide predictive data as to the suitability of a therapeutic approach and an estimation of the progress of treatment.* (Kalorama Information: Companion Diagnostics Market, 2008)

> *Companion diagnostics are tests that provide information about a patient's genetic and genomic characteristics that is then used to make therapeutic treatment decisions.* (*In Vitro* Diagnostics: The Complete Regulatory Guide, 2010)

> *Companion diagnostics are [diagnostics] geared by definition towards a therapy decisions for a particular drug.* (PWC: Diagnostics 2009: Moving Towards Personalized Medicine, http://www.pwc.com/us/en/healthcare/publications/diagnostics-2009-moving-towards-personalized-medicine.jhtml (accessed August 7, 2013).)

Simplified, CDx can be defined as a diagnostic approved in conjunction with a specific drug for select patients who are eligible for the medication. The drug can only be sold in conjunction with the CDx. In this context, CDx is part of the labeling for the drug.

15.4 What are the Drivers for Companion Diagnostics?

All drivers for the increase of use of CDx have their basis in the challenges Pharma is facing. In summary, there are two key drivers.

The first driver is based on increased hurdles by regulators for new drugs to make it to market. This has resulted in the request to Pharma to show better cost-effectiveness compared with an existing class of drugs while demonstrating equal or fewer side-effects. Such proof has already been linked to pricing in some countries and formulary inclusion by a majority of payers. In this context, CDx is used for selecting patients most eligible for a new drug, showing higher effectiveness in this selected patient population and allowing Pharma to combat these new market proof demands.

The second driver is that Pharma companies are becoming financially strained. Over the next few years, an unprecedented volume of sales will be at risk because of patent expirations (e.g., 2012 blockbusters with around US$67 billion in sales will lose exclusivity) [1].

Moreover, the number of new molecular entities dropped from 49 to 27 between 1997 and 2007. Attrition [1] has been worsening, and the majority of new compounds are not delivering returns that justify research and development (R&D) costs. At the same time, R&D spend is rising. As an example, average cost per product has reached more than US$1.3 billion with a success rate from preclinical at around 8% [1].

CDx allows for more targeted drug development, helping Pharma companies to select early the most promising drug candidates. Overall benefits include:

- Avoiding unnecessary R&D costs (faster and safer trials; higher efficacy of drugs in stratified populations).
- Increased chance of successful market launch and pricing.

It is important to note that with the use of CDx, the potential market for drugs will always be limited to the patient population that, by diagnostic proof, is 100% eligible for the drug. In addition, the Pharma company depends on the diagnostic players with respect to the parallel approval of the CDx with drug. If the CDx

Table 15.1 Overview of the benefits and concerns of Pharma with respect to CDx.

Benefits	Concerns
Targeted R&D projects with higher success rates	Reduced target population resulting in less product sold
Demonstrable higher efficacy	Uncertainty about price compensation for volume reduction
Reduction of side-effects	Dependence on diagnostic player: timely development of test and continuous availability of test
Potentially smaller trial groups	Complexity of parallel approval for drug and test
Resulting faster approval of drug	
Resulting easier market access negotiations	

test faces approval challenges and is delayed, the drug also cannot be launched subject to the direct dependency of the drug on the CDx. This aspect has worried Pharma for a long time. Table 15.1 lists the benefits and concerns of CDx for Pharma companies.

Due to the pressures described above, Pharma companies are increasingly embracing CDx despite worries of lower sales. This is reflected in:

- More than 70% of Pharma companies have dedicated 22 out of 100 full-time equivalents to work on CDx projects [2].
- A few Pharma companies have build their own CDx units in-house (e.g., Novartis).[1)]
- Almost all major Pharma players have engaged in CDx partnerships [3].
- Significant numbers of drugs in development are already accompanied by CDx (e.g., up to 60% for Hoffmann-La Roche) in clinical trials.[2)]

15.5 Companion Diagnostics Market

The CDx market starts off from a small basis. CDx is a small but fast-growing, promising subsegment of molecular diagnostics (MDx). This is because of the dominance of molecular tests found in CDx. However, CDx also includes a smaller degree of immunohistochemistry (IHC) and fluorescence *in situ* hybridization (FISH). Most CDx applications to date have been focused on tandem cancer products (e.g., Herceptin/HercepTest and Gleevec/Ventana Anti-c-KIT) and viruses (e.g., tests and drugs for HIV and hepatitis C). Today, around 1% of

1) Our own research.
2) "More than *60% of our pharmaceutical pipeline projects are coupled with the development of CDx* in order to make treatments more effective. The recent launches (. . .) are examples of the concept of *personalized healthcare becoming reality*" (S. Schwan, CEO Roche, September 5, 2012); Roche Investor Presentation.

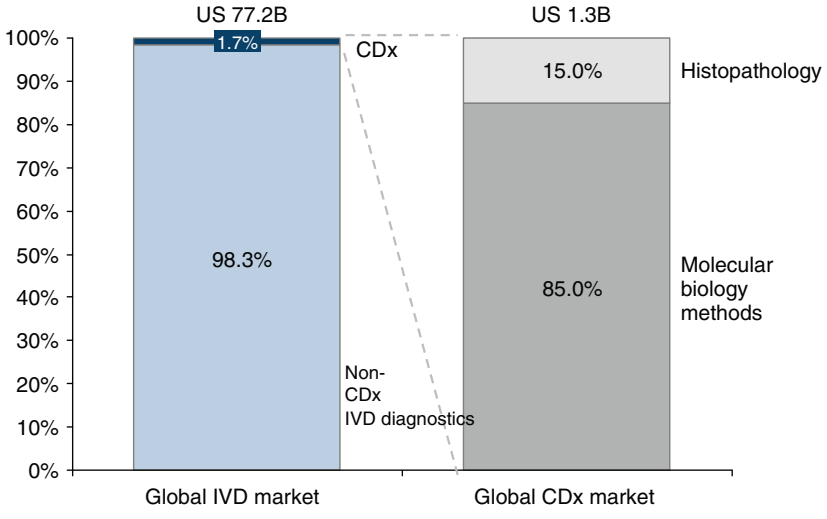


Figure 15.1 IVD and CDx market 2011 (in US $ billion). [1]Including equipment, kits, reagents and lab test services, split regarding technologies (histopathology, molecular biology methods) based on sales of respective kits. (*Sources*: The Market for Companion Diagnostics (Kalorama, 2008); The Worldwide Market for IVD (Kalorama, 2012); *In Vitro* & Companion Diagnostics Industry – Yearbook 2010 (GBI Research, 2011); Medical Imaging Market: X-Ray, Digital X-Ray and CT (Kalorama, 2011); Medical Imaging Market: MRI and Ultrasound (Kalorama, 2012); Roche; BCG Analysis.)

marketed drugs have a CDx and approximately 10% of marketed drugs have pharmacogenomics information as part of their labeling.[3)]

There is significant growth potential reflected in the increasing importance of biomarkers in drug development:

- 35% of late development pipelines (phase IIb–IV) rely on biomarkers [4].
- 58% of preclinical trials rely on biomarker data [5].

Today, the market for CDx is estimated at around $1.3 billion and reflects 1.7% of the total IVD market; more than 85% of all CDx tests are molecular tests or IHC/FISH.[4)] See Figure 15.1.

All market figures are estimated because in many cases financial terms of partnerships are not disclosed. Interestingly, it is known that the fee-for-service deal structure is usually preferred, which shows the dominance of Pharma in the business model today.

3) Our own research and Boston Consulting Group (BCG) analysis (2012).
4) The Market for Companion Diagnostics (Kalorama, 2008); The Worldwide Market for IVD (Kalorama, 2012); *In vitro* & Companion Diagnostics Industry – Yearbook 2010 (GBI Research, 2011); Medical Imaging Market: X-Ray, Digital X-Ray and CT (Kalorama, 2011); Medical Imaging Market: MRI and Ultrasound (Kalorama, 2012); Own research, Boston Consulting Group (BCG) analysis (2012); Roche investor presentation (2011).

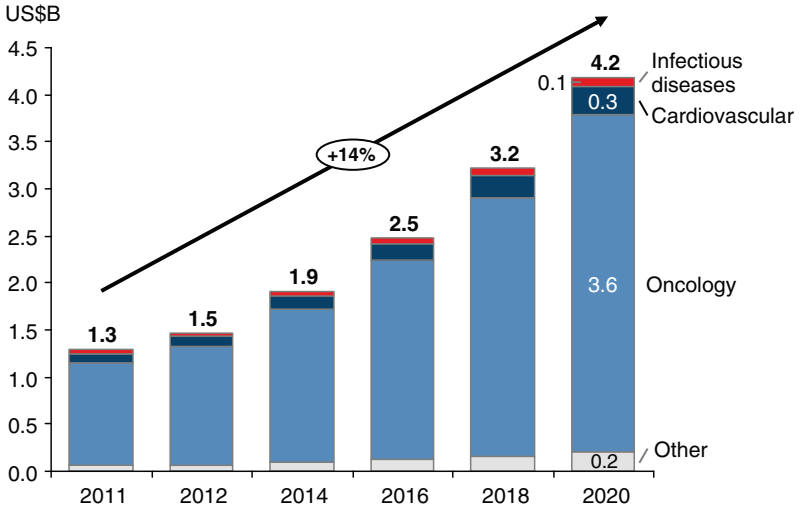


Figure 15.2 CDx market 2011–2020 by therapeutic area (in US$ billion). [5] Driven by pipeline projects in additional therapeutic areas: central nervous system, musculoskeletal, metabolic/hormonal disorders. (*Sources*: See Figure 15.1.)

In 2011, the FDA required CDx testing for around 17 drugs. The most marketed CDx tests are associated with oncology. The secondary focus is on HIV, myelodysplastic syndrome, and cystic fibrosis. See Figure 15.2.

It is expected that the CDx market will grow 14% until 2020, to US$4.2 billion. The number of CDx partnerships between Pharma and diagnostic companies has increased in recent years, reaching 30 in 2011.

The current dominant CDx test providers with the most CDx deals are Roche; Abbott, Siemens, Qiagen, Dako, and Life Technologies. These CDx providers can be separated into two groups: (i) the large IVD players Roche, Abbott, and Siemens Healthcare, and (ii) the specialized MDx players such as Qiagen, Life Technologies, Dako and many other small companies (e.g., Foundation Medicine or MDx Health). After the acquisitions of Clarient and SeqWright, GE is also positioned in the market for CDx with relevant molecular capabilities for oncology.

In addition, Pharma companies like Novartis have started to build up their own CDx capabilities, for internal use only, in order to decrease their dependency on diagnostic companies and to ensure a better alignment of the development of drugs with CDx. Table 15.2 gives a general overview of the different company types competing in the CDx market.

5) The market for companion diagnostics (Kalorama 2008); The worldwide market for IVD (Kalorama 2012); *In vitro* & companion diagnostics industry – yearbook 2010 (GBI research 2011); Medical imaging market: X-Ray, digital x-ray and CT (Kalorama 2011); Medical imaging market: MRI and ultrasound (Kalorama 2012); Own research, Boston Consulting Group (BCG) analysis (2012); Roche investor presentation (2011); FDA webpage: http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/ucm301431.htm (accessed 07. Aug 2013).

Table 15.2 Overview of the types of companyies involved in the CDx market.[a)]

	Pharma companies		Diagnostic companies		
	IVD and Pharma division in the same company	**Diagnostics unit for in-house use in Pharma company**	**Pure molecular diagnostic players**	**Imaging diagnostic player with some MDx**	**Full *in vitro* imaging diagnostic player including IT**
Company example	Roche; Johnson & Johnson	Novartis	Qiagen; Life Technology; MDx Health	GE	Siemens Healthcare
Description	Stand-alone diagnostics business. Several known internal collaborations between Pharma and diagnostics.	Capability to develop own IVD/CDx tests. Diagnostics units focus mainly on development of CDx for own drugs.	Focus on molecular technology/ expertise.	Focus on imaging. First steps in IVD. GE enters into CDx mainly in oncology.	Balanced *in vitro* and imaging diagnostic portfolio. Several CDx partnerships with Pharma in different indications.

a) Source: own research 2012.

15.6 Partnerships and Business Models for Companion Diagnostics

CDx partnerships are, in most cases, non-exclusive. In general, Pharma companies create partnerships with several diagnostic companies that develop and provide the CDx test at a later date. With few exceptions, diagnostic companies have developed and launched a CDx test on their own. Classical partnership examples are listed in Table 15.3.

To better understand the business models, including revenue mechanics of CDx partnerships, it is important to put Pharma challenges in the context of the their value chain – from R&D to Life Cycle Management (LCM).

Figure 15.3 gives an overview of the different starting points for CDx partnerships along the Pharma value chain. Figure 15.3 also shows the current Pharma challenges as drivers for CDx collaborations. The first three models at the top of Figure 15.3 are common today. However, an increasing number of Pharma companies also see the value of diagnostics improving the value of existing drugs (see also Section 15.8).

In the context of the different partnering concepts, the potential revenue models have to be seen as part of the overall business model configuration. Table 15.4 gives an overview of different revenue models, standalones or in combination, depending on the business model envisaged.

The combination of different partnering and revenue models results in three major business models for diagnostic companies. In theory, more models are

Table 15.3 Examples of CDx partnerships.

Approved drug	Pharma company	Diagnostic company	Indication	Description
Gleevec	Novartis	Multiple (e.g., Ventana)	Gastrointestinal stroma tumor	Conventional FDA approval 2001 for chronic myeloid leukemia. 2002, approval of theranostic (c-kit[a)] test) only suitable for 10% of patients.
Herceptin	Roche	Dako	Breast cancer	Herceptin proved ineffective in phase 3 trial in overall population. Received FDA approval for HER26-positive tumors together with companion diagnostic in 1998.
Vectibix	Amgen	Qiagen	Colorectal cancer	Conventional FDA approval in 2006. EU approval requiring KRAS testing in 2007. US label update in 2009: KRAS testing mandatory.

a) Cytokine receptor CD117.

possible, but the three given in Table 15.5 reflect the majority of currently existing business types.

15.7 Regulatory Environment for Companion Diagnostics Tests [6]

Diagnostics have different regulatory requirements and timelines compared with Pharma. This is a key challenge in the parallel development of a CDx and drug, which must be carefully coordinated. In general, the requirements for CDx tests are similar to IVD tests. They need to show:

- Analytical and clinical performance.
- Safety requirements depending on risk classification.

The approval procedure takes an average 12–18 months.

An exception is Laboratory Developed Tests (LTDs), they do not require FDA approval but are governed under oversight mechanisms such as CLIA (Clinical Laboratory Improvements Amendments). However, the FDA mandates that new technologies need FDA approval for use of clinical diagnostics [7]. Furthermore, LDTs have limits with respect to their geographical expansion.

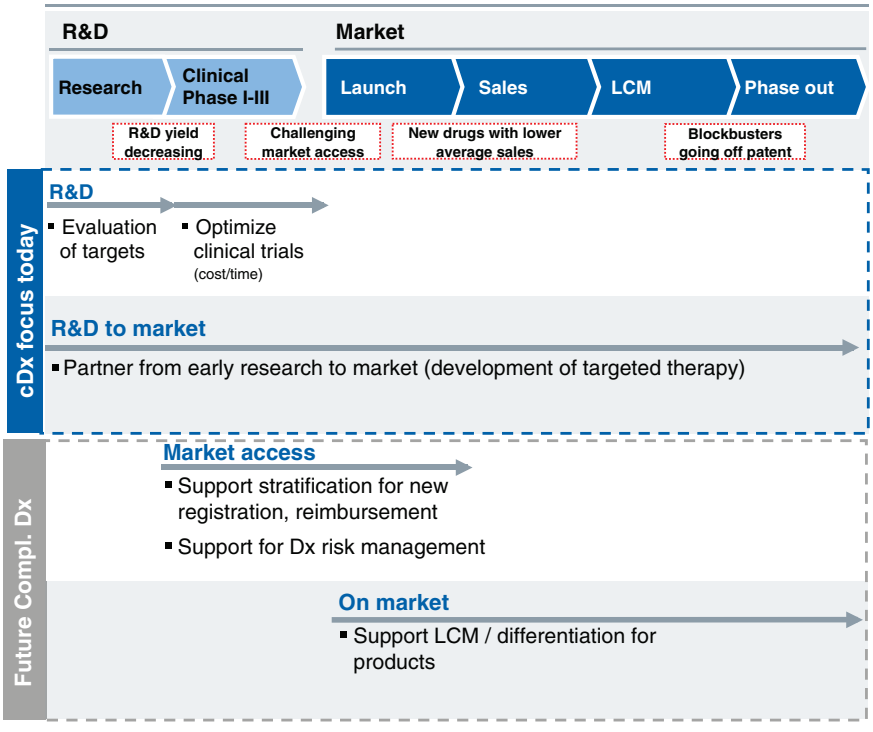


Figure 15.3 Overview of different CDx collaboration models.

Today, the United State is far advanced in terms of CDx awareness and guidance compared with the European Union. The FDA, which is the US regulatory body, released guidance on IVD CDx in July 2011. In this context, a uniform CDx pathway is being developed by the FDA. While the FDA is in charge of the CDx approval process, manufacturers are expected to coordinate with the respective institutions, such as the CDRH (Center for Devices and Radiological Health) and the CDER/CBER (Center for Drug Evaluation and Research/Center for Biologics Evaluation and Research). In the United States, it is also possible to have a joint evaluation of the pharmaceutical and diagnostic test (i.e., an investigation in the same clinical study).

In the European Union, the European Medicines Agency (EMA) is the European regulatory body in charge of the approval of CDx tests. The EU medical device directive for CDx is currently under revision but presently no uniform EU approach exists. Therefore, the regulatory landscape in Europe is very heterogeneous with different national approval procedures. This means that all CDx test are treated as normal IVD tests in Europe requiring a CE mark for European-wide market access under the common EU directive [8]. However, for specific CDx approval, the decisions are made at the national level. Different to the United States, in Europe the drug and test are evaluated separately.

Table 15.4 Overview of major revenue models in CDx.

Revenue model	Description
R&D revenue models	
Upfront payment	Diagnostic company requests upfront payments when entering into a CDx partnership, before R&D starts. This helps to ensure covering a part of the development costs and reduces the risk for the diagnostic company.
Fee-for-service/contract research	Diagnostic company is paid for CDx development services done for Pharma R&D. End-product to be owned by Pharma. Payment is based on per used full-time equivalents or annual lump sum.
R&D cost sharing	Sharing of R&D cost for CDx development between Pharma and diagnostic company. This model is often linked to royalty payments when product is marketed.
Milestones	Diagnostic company development progress is based on reaching defined milestones in development of CDx. Payments only are triggered when agreed upon milestone is reached. Payment for each milestone can differ and is pre-agreed.
Market revenue models	
Royalties	For the developed CDx and related intellectual property the Pharma company grants a percentage of drug sales to the diagnostic company. The Pharma company can also guarantee minimum revenue from CDx and reduce the risk exposure for the diagnostic company.
Reimbursement	The diagnostic company gets reimbursed for the CDx from payors. The reimbursement differs depending on the complexity of the test.

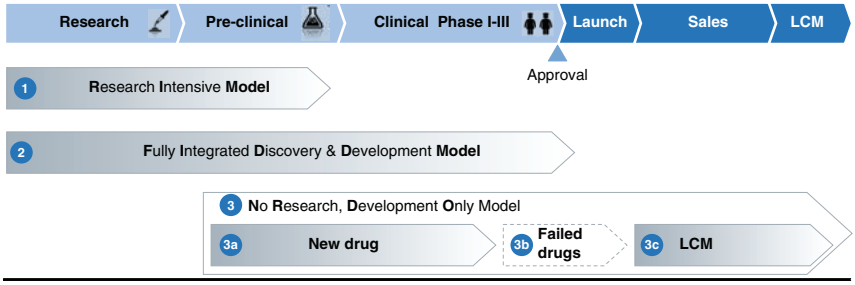
In summary, the regulatory environment continues to be challenging for CDx, especially in Europe. However, both the FDA and EMA are working on better regulations.

15.8 Outlook – Beyond Companion Diagnostics Towards Holistic Solutions

In the future, healthcare systems will favor more technologies and solutions that maximize the output of healthcare supply for their spend. CDx is the first step towards a more efficient and better care provision. The question is what will be the next wave after CDx? Which technologies will generate breakthroughs, and how will holistic approaches to diseases and their management change the patient path and the industry value chain?

On the one hand, only technologies that fulfill future healthcare system-demanded efficiencies will be successful on the market and be real breakthroughs. These include non-diagnostic technologies, such as devices and information technology (IT)/mHealth (mobile health).

Table 15.5 Overview of major business models along the Pharma value chain.



No.	Business model	Description
1	Research-intensive model (RIM)	Diagnostic platform, tool-based, and service model seeking to commercialize services, marker, and technology for Pharma companies. In general three submodels within RIM can be identified: • Platform provider, offering the use of their platform (e.g., Hologic) • Proprietary content provider, offering proprietary asset content (e.g., Qiagen) • Service • Service provider, only offering services (e.g., Clarient) *Target*: contracting with one or more alliance partner(s) who needs support in drug and target research. Revenue model focused mostly on fee-for-service and milestone payments.
2	Fully integrated discovery and development model (FIDDM)	Diagnostic platform companies extend existing capabilities to take innovation further along the product development process, including launch and commercialization of CDx. *Target*: alliance on more favorable terms than RIM, due to longer involvement in value chain. Revenue model can cover any payment from fee-for-service, milestone and upfront payment, royalties on drug, and respective reimbursement for CDx.
3	No research, development only (NRDO)	Focus is on different development models but no research work considered. Three major different business models can be differentiated.
3a		Diagnostic company only supports drug development activities along clinical phases 1–3. Thus, any diagnostic can be leveraged. Revenue models are mostly fee-for-service, milestone, and upfront payment.

(*continued*)

Table 15.5 (*Continued*)

No.	Business model	Description
3b		Diagnostic company supports diagnostic test development to allow for higher efficacy or better risk control after drug failure in approval process. Goal is to re-launch drug together with CDx. Revenue models are mostly milestone and upfront payment. Royalties on drug and respective reimbursement for CDx are possible.
3c		Diagnostic company supports diagnostic test development for existing drug. To be relaunched as new bundle with improved outcome (including clinical trial) to extend life cycle. Revenue model not proven yet. Most likely based on royalties on stabilized or increased drug sales and bundle reimbursement revenues.

One example with high potential is an evolving new group of diagnostic tests – specialized molecular tests – providing higher efficiency for the treatment and system by going beyond simply screening, stratifying patients, or confirming the presence of a disease – called "complementing diagnostics." Complementing diagnostics predict risks, diagnose different disease stages, and allow monitoring of therapeutic response. A complementing diagnostic refers to any diagnostic test that optimizes the treatment as described above, but is not part of a drug label. This means the treatment (e.g., the drug) can be sold without the complementing diagnostic test.

With this paradigm change, the role and importance of point of care (POC) and imaging will change.

Future POC testing will offer rapid turnaround results (usually minutes). In some cases, test results will be obtained without the use of sophisticated analyzers. Typically, speed and convenience is achieved at the expense of quantity of information. As such, POC tests typically provide simple yes/no or positive/negative results. This allows for higher flexibility in the use of "any" molecular IVD tests with respect of the location where the test is done and also with respect of the user (also medical-technical assistants can use future POC systems easily). Major obstacles to fully realize this potential include the fact that the majority of tests are manual and thus prone to operator error, a lack of connectivity with information systems, and narrow menus that require separate devices for different tests. Once these issues are addressed by advances in automation, wireless technology, and miniaturization, and the number and scope of tests increases, the importance and increase utilization of POC testing will accelerate.

The relevance of imaging within healthcare is expected to increase because physiological information that an IVD cannot provide is an important element

for early detection and understanding of certain diseases and their progression, such as cancer, cardiovascular disease, respiratory disease, orthopedic and spinal conditions, and neurological disorders. As such, there is tremendous pressure to develop more effective screening technologies to assist with the prevention and treatment of major diseases. Diagnostic imaging is included in this effort, with the emphasis on improving the accuracy, efficiency, functionality, yield, and speed of the instrumentation, while lowering their cost, increasing accessibility, and reducing radiation exposure.

In this context, the emerging role of devices has to be understood. Devices can be intelligent injection devices, such as RebiSmart, or sensor devices that monitor movements. Such devices support the increasing role of monitoring. For example, patient adherence has been identified as a major lever, with non-compliant patients costing almost twice as much as compliant ones. Therefore, devices and monitoring technologies that increase adherence, either by facilitating delivery or by monitoring and reporting compliance, will gain importance. Therefore, drug–device combinations that increase adherence, and therefore increase the health outcome of the treatment, will increasingly be supportive to healthcare systems.

On the other hand, combination of different diagnostics, devices, IT, and therapeutics (drug and surgery) to "one holistic solution" that improves medical outcomes will emerge. In simple terms, "leveraging the existing and new" by using diagnostics (IVD and/or imaging), IT, and drugs in a more efficient combination along the treatment pathway will become a new trend.

Holistic solutions have to be seen as the final stadium. Such solutions would not only be restricted to certain aspects of care, but would also take into account the entire life-long treatment for any condition of the patient.

Future solutions will most likely consist of five basic elements combined in a holistic solution:

- An optimized care pathway for a specific medical condition involving the most adequate diagnostic procedures.
- Decision support tools (e.g., to help classify patients into risk groups).
- An IT platform to enable management of the solution and exchange between stakeholders
- The therapeutic (drug and/or surgery).
- Program management to provide control, support, quality assurance, and monitoring of outcomes.

In addition to improving quality of care, such holistic solutions aim to decrease costs of care for the healthcare system, not just for the provider. To realize both of these aims, more complex collaborations between diagnostic companies, therapeutic companies, healthcare providers, and payors will be key. This will ensure such holistic solutions can target all aspects of the healthcare delivery system, including inpatient and outpatient care.

With the transformation of healthcare systems, new incentives for healthcare player's new revenue models will emerge:

- Result/outcome-based payments.
- Complex bundle payments for solutions (e.g., diagnostic, drug, service, and IT).

Nevertheless, several challenges are expected to slow the adoption of such solutions. These include:

- The disruptive nature of the approach, which runs contrary to current medical industry business models and clinical practice.
- Patient privacy concerns, especially regarding the potential misuse of patient information (e.g., genetic information).
- The laborious nature of biomarker discovery, even with advances in genomics and proteomics.
- Overall costs of advanced testing, especially as reimbursement generally lags behind technological development.
- Lack of predicate devices for regulatory approval.
- The need for strategic alliances between diagnostics Pharma companies, providers, and payors.

In addition, there is a significant scientific complexity behind holistic solutions as molecular understanding of disease is still insufficient for development in some areas. Furthermore, it is difficult to obtain a clinical evidence base because it often takes thousands of patients to develop, leading to more than a decade of downtime. Additionally, it will be hard to guarantee intellectual property protection for a new biomarker and IT solution.

In light of these challenges, the changing value pools in healthcare have to be evaluated diligently to identify new profit pools companies can tap into. This will help open up opportunities to create synergies and additional benefits through integration of products and technologies.

References

1 Our own research and BCG analysis (2012) – internal project.
2 Tufts Center for the Study of Drug Development, Impact Reports November 2010 FDA, Cutting Edge Information.
3 Tufts Center for the Study of Drug Development, Impact Reports July 2011, FDA, Cutting Edge Information.
4 Tufts Center for the Study of Drug Development, Impact Reports November 2010 and July 2011, FDA, Cutting Edge Information.
5 Tufts Center for the Study of Drug Development, Impact Reports November 2010 and July 2011, FDA, Cutting Edge Information.
6 Miller, I. *et al.* (2012) Market access challenges in the EU for high medical value diagnostic tests. *Personalized Medicine*, **8** (2), 137–148.
7 McKinsey (2013) Personalized Medicine – The Path Forward, http://www.mckinsey.com/~/media/mckinsey/dotcom/client_service/Pharma%20and%20Medical%20Products/PMP%20NEW/PDFs/McKinsey%20on%20Personalized%20Medicine%20March%202013.ashx (accessed August 7, 2013).
8 European Directive on *In Vitro* Diagnostic Medical Devices (98/79/EC), 2013.

16
Ethical, Legal, and Psychosocial Aspects of Molecular Genetic Diagnosis

Wolfram Henn

16.1
General Peculiarities of Genetic Diagnoses

Modern medicine has long become unthinkable without laboratory analyses for disease-related parameters. Consequently, laboratory diagnoses and their results are generally subject to the same ethical and legal categories of clinical indication, patient autonomy, and data protection as all other medical measures and the information related to them.

Over the last 35 years, DNA sequence analysis has rapidly evolved from a highly sophisticated research technology to a cornerstone of everyday laboratory routine, be it in virology, pathology, forensic medicine, or clinical genetics. Laboratory findings about constitutional genetic traits are not fundamentally unlike any other medical information, as highly sensitive data can also be produced in, say, infectiology or psychiatry [1].

However, there are a few respects in which genetic data actually differ from other kinds of medical information, thus requiring adapted (and mostly elevated) standards of pre-test counseling, patient consent, process quality, evaluation, disclosure, and data protection.

- *Dissociation of genetic traits from somatic pathogenesis.* Most laboratory parameters reflect morphologic, metabolic, or functional aspects of an ongoing pathological process, thus being able to detect only diseases that are physically present in the organism. Genetic traits predisposing to disease, however, are usually present already in the fertilized egg, often years or decades before the beginning of the organic malfunction itself. This fact may open windows of time for predictive or prenatal diagnosis [2].
- *Discrepancy between diagnostic and therapeutic options.* In hardly any other field of medicine is there as wide a gap between knowledge of a diagnosis and a promising therapy. In fact, only very few genetic diagnoses immediately lead to a promising treatment, let alone to a cure. This is why, in public understanding, genetic disease is often perceived as a "fateful" disease, in some cultures even as a curse [3].

Nucleic Acids as Molecular Diagnostics, First Edition. Edited by Andreas Keller and Eckart Meese.

- *Trans-individual meaning of genetic diagnoses.* Genetic diagnoses usually have a double meaning: they identify a patient and create persons at risk. A person freshly diagnosed with a genetic disease therefore faces a double challenge: coping with his/her own illness and dealing with the task of informing other family members about their, mostly unanticipated, at-risk status [4].
- *Possible misuse of genetic data.* Any kind of information that may give hints about a person's future health can be of interest for third parties, in particular employers and private insurers. As they might misuse genetic test results for illegitimate purposes (e.g., denial of disability insurance), data protection is of the utmost importance in medical genetics [5].

16.2 Informed Consent and Genetic Counseling

Any kind of diagnostic or therapeutic medical intervention, even simple blood sampling or gathering of medical data in a laboratory, interferes with the investigated person's (who not in all cases is a patient affected with clinical disease) physical, emotional, and social well-being, and thus has to be legally justified. Therefore, genetic testing in a clinical diagnostic context is illegal without both a medical indication and the investigated person's informed consent [6].

The medical indication for a genetic test means a promising, or at least reasonable, methodological approach to a relevant genetic feature, which it is in the patient's interest to identify. The scope of possible indications for genetic testing is somewhat wider than in other fields of medicine, as not only a sick patient's own health may be at stake, but also other important issues like life planning in the wake of future neurodegenerative disease, reproductive choices for carriers of recessive traits, or even other persons' health interests in the context of segregation analyses. A reasonable methodological approach in a clinical diagnostic context comprises technical feasibility (i.e., the choice of techniques validated for the given diagnostic question) plus an acceptable cost/benefit ratio. As an example, application of a large mutation panel or even complete sequencing of the *CFTR* gene is not indicated for siblings of a cystic fibrosis patient with ascertained homozygous mutation ΔF508, but rather mutation-specific polymerase chain reaction (PCR) for that single mutation. Conversely, in a patient with clinical symptoms fitting with cystic fibrosis, but only a proven heterozygous ΔF508, it may well make sense to analyze the *CFTR* gene in further detail in order to identify the assumed second mutation [7].

However clear the medical indication may be, the final decision whether or not to perform a genetic test lies with the person to be investigated. Throughout all developed healthcare systems, informed consent is the legal prerequisite for any legal diagnostic or therapeutic measure.

A valid consent can only be given if the patient is generally able to give consent in terms of age and/or mental ability, and moreover has obtained – or at least had free access to – understandable, comprehensive, and neutral information

about the procedure and its possible consequences. As a rule, as the quality and intensity requirements for patient information become the more demanding, the more severe the possible consequences may be. Although written consent to genetic testing is not mandatory in all countries, it is very advisable for a genetics laboratory to require, along with the sample, a consent form signed by the physician and patient documenting which test should be done and why, and to whom the results should be disclosed.

As to genetic diagnosis, there are several levels of pre-test information.

The most simple way may be printed information leaflets that are handed or sent to the patient without any personal contact with a physician or other expert. However, such paper-only information provides no clue for the laboratory whether the purpose of the test was actually understood; therefore, this approach to consent is legally doubtful and, in most countries, forbidden in the context of clinical genetic diagnosis.

The standard procedure to obtain informed consent in the context of differential diagnosis (i.e., testing for a gene defect underlying a physically present disease) is pre-testing discussion in personal contact between the patient and the physician referring to the test. In a rather simple clinical constellation, like testing for the factor V Leiden mutation in a patient with recurrent thrombosis, a rather short talk may be sufficient. However, more complex issues, such as *BRCA1/2* mutation testing in a patient with bilateral breast cancer, will require at least more time for the talk and more genetics-related experience of the physician. Although not mandatory in cases of differential diagnosis, genetic counseling is very advisable in these more complex issues. Even the shortest pre-test discussion must include the notion that any genetic test may yield results of importance for other family members and the offer of post-test genetic counseling in the case of a pathologic result.

The highest level of pre-test patient education is genetic counseling, comprising a detailed talk not only about the planned laboratory test, but also the individual health history and the family pedigree. In some countries (e.g., Germany), genetic counseling is done only by specialized physicians, whereas in other countries (e.g., the United Kingdom), standard issues of genetic counseling are covered by specialized "genetic nurses" [8].

From the genetic laboratory's perspective, all types of pre-test patient education must end up in a written and thus legally ascertained informed consent. It has become established practice for them to make customized referral forms for referral and consent available on the Internet, and to require them to be sent in along with the sample. The physician's referral and the patient's consent may also be placed on two separate forms. Although there is wide variation as to the length of text and the amount of details on the forms, the following minimum items should be included:

- Patient's name and date of birth.
- Name and address of referring physician.
- Clinical diagnosis/suspicion.

- Diagnostic context (differential diagnosis/predictive/prenatal diagnosis).
- Pre-existing genetic data, especially family-specific mutations for predictive diagnosis.
- Gene(s) to be analyzed.
- Diagnostic approach to the gene(s) (complete analysis or testing for specific mutations).
- Bearer of expenses for the analysis (patient/hospital/insurer).
- Physicians other than the referring one to whom the result should be disclosed.
- Patient's signature confirming understanding of explanations and consent.
- Referring physician's signature.

There may be additional items legally requested in the given country, such as confirmation that genetic counseling has been offered, confirmation that the given consent may be revoked until disclosure of the results, and so on.

The patient's written decision how to proceed with the sample and the data once the analysis is completed is helpful and in some countries mandatory:

- Should the sample be discarded immediately or stored for further diagnostic analyses?
- Does the patient agree to use of sample material for technical validation and quality control purposes?
- Does the patient agree to anonymized/pseudonymized use of sample material for scientific purposes?
- Should the results be stored permanently or discarded after a certain period of time?
- May the results be used as a resource of genetic counseling or testing of blood relatives?

These requirements for consent and documentation are meant for genetic testing within a medical framework. The upcoming direct-to-consumer or "over-the-counter" genetic tests outside medical diagnosis are currently located in a legal and ethical gray area with wide variability between different countries. However, for legal tests of that kind, similar standards of not only laboratory practice but also proband information and consent are desirable, although – particularly as to with regard individual pre-test counseling – hard to achieve [9].

16.2.1 Testing of Persons with Reduced Ability to Consent

A general prerequisite for valid informed consent to genetic testing is not only factual understanding of the purpose of the test, but also an appropriate legal status of the person giving consent. Some persons, however, are factually (e.g., adults with dementia) or formally (e.g., children) unable to give valid consent to a medical procedure.

As it is impossible for a laboratory to check whether the person whose sample has been sent in is factually able to understand the meaning of the test, this responsibility lies with the referring physician. If he/she has signed the referral form, there should be no reason for the laboratory to doubt the validity of the consent.

However, the most important formal criterion for a legally valid consent is the patient's age. Although, in terms of ethics, a person of less than 18 years of age may well be able to understand the meaning of a genetic test and to be factually able to consent, it is important for the laboratory processing a sample to strictly require the signature of a custodian. In most cases this will be a parent (the signature of one parent is sufficient), but it may also be a custodian denominated by a legal authority.

The custodian is responsible and entitled to decide which intervention is in the best interest of the child or ward, provided of course that it is medically indicated. Consequently, he/she is the one with whom the pre-test discussion has to be performed.

However, there are some restrictions with respect to the custodian's right to decide. Genetic testing in a differential diagnostic context can, as a rule, be regarded as being justified, because it serves for therapy planning and may disburden the child or the mentally disabled patient from other, perhaps more invasive, diagnostic measures. For predictive diagnosis on underage persons without immediate therapeutic consequences, however, the future "right not to know" must be taken into consideration, which ends up in the simple rule that no predictive diagnosis should be performed in childhood or adolescence unless its result requires therapeutic or preventive action before adulthood. As an example, predictive testing of a child for an oncogenic *RET* mutation known from a parent makes sense, because a preventive thyroidectomy must be done before age 6 if the mutation is present, whereas predictive testing for a *BRCA1/2* mutation should be postponed until adulthood, as *BRCA*-related breast or ovarian cancer is exceedingly rare before age 25. It may happen that a parent (e.g., a mother with a *BRCA* mutation) urges the physician to test her child predictively at an unduly young age. If so, the genetic counselor as well as the laboratory physician are well advised to refuse the test, as it is not medically indicated although not legally forbidden [10].

Prenatal diagnosis in underage or mentally disabled women is an ethical and legal minefield, as any prenatal test may end up in a decision about a termination. The regulations with respect to that differ strongly between different countries and in some, not even the parents but only a guardianship court may give consent. Whenever a laboratory is confronted with a request of that kind, legal advice is necessary.

If an unannounced sample with an unclear status of consent arrives at the laboratory (e.g., amniotic fluid from an underage woman), the best procedure is to start with the immediately necessary steps of processing, such as establishing the cell culture or DNA extraction, but to postpone the evaluation until legal certainty is achieved.

16.3 Medical Secrecy and Data Protection

All medical data are subject to the patient's informational self-determination, and among these, genetic data are particularly sensitive and many countries have specific laws regulating the use of genetic data in a restrictive manner.

As a general rule, results of genetic tests must be disclosed to no other party than the referring physician and – usually through that physician – the patient or the patient's custodian. Other persons may be granted access to the results only if they have been nominated in advance on the referral form and after the test only with separate written consent of the patient.

In particular, release forms from medical secrecy obligations, as widely used by health or life insurers, should be looked upon with caution even if signed by the patient, as they may be legally invalid for genetic data. If a genetic laboratory receives such a form, it is wise to re-contact the patient and to discuss the issue, and perhaps to ask him/her to personally send the requested copy of the report to the insurer rather than authorizing the laboratory to do so.

Another delicate issue is the use of International Classification of Diseases (ICD) codes, which often are routinely requested by the cost bearers with the bill. If it is unavoidable, the ICD code should be as non-disclosing as possible, such as G60.9 "hereditary and idiopathic neuropathy, unspecified," rather than G60.1 "hereditary motor and sensory neuropathy." If possible, within the national ICD version, subcodes classifying a diagnosis as unascertained should be used.

A problem specific to genetic reports is their possible connection to other family members' findings. In particular, mentioning the name of the index person for a predictive diagnosis on the report about another family member must be avoided. However, as it is important for the laboratory to be able to trace back a report to the index patient, he/she should be mentioned on the report through the laboratory's internal case numbering, but not more.

It is self-evident but, in practice, often demanding to make sure that all printed reports are safeguarded throughout in locked containers and are never accessible for non-authorized persons on the laboratory workers' desks or on their open bookshelves. An often underestimated problem is the technical protection of electronically stored patient data. In the era of computer viruses and data theft, a password-protected registry on a computer with an Internet connection may be not enough and a data safety failure might be a catastrophic event. Therefore, storing all patient data and the backups on separate computers not connected with the Internet should be considered.

Although they cannot be as easily misused as processed data, the patient's blood or DNA samples are undoubtedly sensitive carriers of genetic information that must be protected with the same caution as genetic reports. Unlike in research, pseudonymization or anonymization are not appropriate measures of

data protection in clinical genetic diagnosis, as the samples must remain retraceable for control purposes or for future diagnostic tests on the stored material. In order to make sure that the patient's sample will remain accessible, it is important to ask the patient for consent to permanent storage of the sample already on the referral form. This notion is of particular value with respect to differential diagnosis of patients with life-limiting diseases: a sample from a patient affected from, say, metastasizing breast cancer should not be discarded after a negative *BRCA1* and *BRCA2* test, as it may well happen that the sample will be badly needed after her death for testing of as yet unknown other genes [11].

16.4 Predictive Diagnosis

Diagnosis for a certain disease-predisposing mutation known from the given family is a particularly sensitive issue from both the patient's and the physician's perspective, as it is performed on a clinically healthy person, but has lasting consequences on his/her future life. Proof of an *APC* mutation disposing to familial adenomatous polyposis coli in a young adult, as an example, sets the indication for preventive colectomy, while a negative finding releases him/her from the prospective bowel cancer or preventive surgery as well as from the psychological burden of possibly passing on the deleterious trait to offspring. Even more challenging is predictive diagnosis for neurodegenerative diseases such as Huntington's, as a positive diagnosis does not open any paths for prevention, but may only serve for non-medical considerations of life planning. Conversely, an often underestimated yet common effect of a negative diagnosis is paradoxical emotional distress in spite of the good news.

Therefore, persons at risk for high-penetrance traits are in need of thorough genetic and psychosocial counseling, usually in several sessions, and a reflection period of at least 1 month prior to undergoing predictive diagnosis. This procedure has been outlined in several guidelines and in some countries is even prescribed by law. The result of the test – even if negative – must never be disclosed in written form or on the telephone, but exclusively in a personal conversation with the referring genetic counselor.

This rather complicated way of dealing with patient information after the analysis, together with the patient's right to abstain from information even after completion of the test, calls for cautious handling of the report. The best procedure, nor only for the patient but also for the genetic counselor as the possible bringer of bad news, is to put the written report of the predictive diagnosis into a sealed envelope, which is placed into a larger outer envelope addressed to the counselor with a attached notice: "This envelope contains the genetic testing result of Mrs/Mr. . . ."

At the laboratory level, even the most remote sources of error or misinterpretation must be avoided. The first thing is to offer predictive diagnosis only with respect to high-penetrance traits for which the respective gene mutation confers a high predictive value – this is why "predictive" *APOE4* testing that only modifies a person's future Alzheimer risk is of little practical value. The familial mutation tested for must be ascertained as disease-causing, either based on reliable literature or database information, or on its molecular nature. A familial *BRCA1* mutation, for instance, may serve as a reliable template for predictive testing of relatives if it is noted as recurrent and disease-causing in the BIC (Breast Cancer Information Core) database and/or if its mechanism, such as a stop mutation or a large intragenic deletion, excludes a functional product. All other types of familial mutation, especially those that are not known from the literature but only assigned as likely pathogenic through *in silico* analysis, must be used and interpreted with great caution; in case of relevant doubt they must be regarded as inappropriate for predictive diagnosis.

Technical practice itself should encompass completely separate testing of two independent samples from the at-risk person; some laboratories even insist on two samples taken on different days in order to avoid erroneous labeling of the sample by the referrer.

Predictive diagnosis without a living index patient is another problematic issue, although rather often requested with respect to familial cancer or neurodegenerative disorders. Here, it is a matter of correct pre-test counseling to check how clear the affected patients' clinical diagnoses actually were and which margins of clinical error may reside. As an example, a three-generation family history of Huntington's disease, plus a psychiatrist's report confirming the clinical diagnosis in a deceased parent of the at-risk person make it very probable, but not absolutely sure, that CAG-repeat testing in the Huntingtin gene is the right predictive approach. Still, the rare dominant "Huntington-like diseases" like spinocerebellar ataxia type 17 (*SCA17*) must be taken into consideration in the counseling process as well as in the laboratory report.

Intrafamilial communication of genetic information, in this context of a predictively testable hereditary disease, is solely the task of the patient or predictively identified mutation carrier; however, this is often experienced as a heavy emotional burden. It is the genetic counselor's task to assist the client in that process, but it can be of great help from the laboratory's side if the affected index patient has already agreed to use of his/her test result for genetic counseling and testing of blood relatives. If necessary, this may even be done without mentioning the name of the index patient to them, provided that the blood relationship is certain enough.

16.5 Prenatal Diagnosis

In terms of ethics, prenatal diagnosis affects the possibly conflicting interests of a born and an unborn person – in terms of law the fetus cannot be a bearer of

rights at the same level as the mother. This inherent peculiarity of the feto-maternal unit can cause unsolvable conflicts of interest, which are addressed by laws and guidelines that differ between countries more strongly than in any other field of medicine [12]. Therefore, we will only deal with general considerations here.

The person to give informed consent to prenatal diagnosis is the mother alone and not the father, as already stated in Roman law "*Mater semper certa est, pater semper incertus* [the mother is always certain, the father always uncertain]." Accordingly, she must be comprehensively informed before the prenatal diagnosis not only about the diagnostic procedure and its possible outcome, but also about the ethical and legal aspects of a pregnancy termination in the case of a pathologic result. Due to the complexity of the issue, in some countries personal genetic counseling of the mother is mandatory by law before a prenatal diagnosis.

Beyond the variable legislation, there is widespread consensus that only genetic traits of considerable relevance for the child's health may be tested for prenatally. In particular, prenatal diagnosis for the fetus' gender and prenatal paternity testing are unacceptable and prohibited by law, respectively. As to diseases that do not manifest before adulthood, like Huntington's or familial breast/ovarian cancer, there are contradictory rules and views between different countries; the same holds true for preimplantation diagnosis, although any discussion on this topic is beyond the scope of this chapter. Anyway, as no physician is obliged to perform measures that are contradictory to his/her own ethical convictions, any laboratory has the right not to offer prenatal diagnosis for certain parameters individually considered as unacceptable, regardless of the legal situation. For example, prenatal diagnosis for familial adenomatous polyposis coli is offered by some laboratories, but not by others [13].

Similarly to predictive diagnosis, an incorrect result of prenatal genetic testing can have catastrophic consequences, be it false positive or false negative. Therefore, sources of error specific to prenatal diagnosis must be considered carefully and, as far as possible, excluded. Among these are maternal contamination of the sample (especially in the context of chorionic villus sampling), fetal mosaicism, and sample cross-contamination between non-identical twins. Any remaining uncertainty must be scrutinized in the report, in which the given legal restrictions to prenatal disclosure of gender must also be respected. If appropriate within the given national legal framework, the report must point out mandatory and/or recommended disclosure of the results to the patient through personal genetic counseling.

In the rare but delicate case that during the technical process of prenatal diagnosis – or in any other medically motivated genetic diagnosis – non-paternity is encountered, this fact should be considered by the laboratory physicians as an unwanted "side-effect" of the diagnostic procedure and not be disclosed to the patient, let alone her partner. This reluctance is clearly justified by the fact that the laboratory's task is to solve the medical problem and nothing more.

16.6 Multiparameter Testing

Novel technologies of DNA analysis – array comparative genomic hybridization (CGH) at the chromosomal level and next-generation sequencing (NGS) at the single-gene level – provide not only unprecedented opportunities as to the multitude of genetic target aberrations for a single analytic approach, but also a paradigm shift for pre-test patient information and post-test interpretation of results [14].

Unlike conventional approaches aimed at one single or a few chromosomal loci or genes, both CGH and NGS provide information about a much larger number of potential genetic variants than can be addressed in detail in a pre-test discussion or counseling session with the patient. The usual approach to informed consent through pre-test information about one single, more or less exhaustively explainable locus or gene does not work anymore and the patient's consent to an array CGH or an NGS panel analysis inevitably encompasses sequencing of genes not explicitly mentioned before the test.

This notion increases the probability of unexpected or unclear results, such as incidental proof of deletion of the neurofibromatosis 1 gene on chromosome 17 through array CGH aimed at a child's developmental delay or of a hemophilia A mutation through X exome sequencing in search of X-linked mental retardation.

It is the task of both test design and pre-test information to minimize the probability of such events, according to the "less is more" rule – the best diagnostic strategy is the one that promises a good chance of finding the clinically sought-after diagnosis without generating too much unnecessary genetic data. This will typically result in a stepwise diagnostic approach to a given clinical suspicion. For example, genetic evaluation of a boy with mental retardation should begin with fragile X CGG-repeat analysis addressing the most common reason and only in the case of a inconspicuous result for that single gene may move on to a multigene NGS panel or X exome analysis.

Even the most cautious diagnostic approach will, more often than ever before, generate results of unclear clinical meaning – an issue that must already be addressed in the pre-test discussion with the patient or the custodian. Array CGH has taught us the somewhat humbling lesson that not every initially detected microdeletion or microduplication eventually explains a child's phenotype, but that some of these turn out to be rare familial copy number variants. A premature "Eureka!" that has to be revoked later on is painful for the family and embarrassing for the doctor.

What is more, whole-genome sequencing (WGS) as a routine procedure is within reach, which will further multiply the problems of multiparameter testing outlined above. At the current state of discussion, there are no clear rules of good practice, neither for the indication to perform WGS nor for processing, storage, and disclosure of the data generated through it. Therefore, it can only be recommended to recur to common sense as often a possible, trying to find the balance between disclosing truly necessary information, even if unexpected,

and safeguarding the patient from drowning in a flood of useless data. In other words, the right to know as well as the right not to know must be exercised at their due places [15].

In a future not too far ahead, we will also be confronted with non-invasive complete sequencing of the fetal genome, which will even require protecting an unborn child's right not to know through non-disclosure of genetic information to the prospective parents. Then, the geneticists in the laboratory as well as at the counseling desk will more often than ever find themselves in the role of referee between different persons' conflicting interests of informational self-determination [16].

References

1 Green, M.J. and Botkin, J.R. (2003) "Genetic exceptionalism" in medicine: clarifying the differences between genetic and nongenetic tests. *Annals of Internal Medicine*, **138** (7), 571–575.

2 Motulsky, A.G. (1994) Predictive genetic diagnosis. *American Journal of Human Genetics*, **55** (4), 603–605.

3 Wolff, K., Brun, W., Kvale, G., Ehrencroma, H., Soller, M., and Nordin, K. (2010) How to handle genetic information: a comparison of attitudes among patients and the general population. *Public Health Genomics*, **13** (7–8), 396–405.

4 Black, L., McClellan, K.A., Avard, D., and Knoppers, B.M. (2013) Intrafamilial disclosure of risk for hereditary breast and ovarian cancer: points to consider. *Journal of Community Genetics*, **4** (2), 203–214.

5 Joly, Y., Ngueng Feze, I., and Simard, J. (2013) Genetic discrimination and life insurance: a systematic review of the evidence. *BMC Medicine*, **11**, 25.

6 Burgess, M.M., Laberge, C.M., and Knoppers, B.M. (1998) Bioethics for clinicians. 14. Ethics and genetics in medicine. *CMAJ*, **158** (10), 1309–1313.

7 Dequeker, E., Stuhrmann, M., Morris, M. A., Casals, T. *et al.* (2009) Best practice guidelines for molecular diagnosis of cystic fibrosis and CFTR-related disorders – updated European recommendations. *European Journal of Human Genetics*, **17** (1), 51–65.

8 Rantanen, E., Hietala, M., Kristoffersson, U., Nippert, I. *et al.* (2008) What is ideal genetic counseling? A survey of current international guidelines. *European Journal of Human Genetics*, **16** (4), 445–452.

9 Patrinos, G.P., Baker, D.J., Al-Mulla, F., Vasiliou, V., and Cooper, D.N. (2013) Genetic tests obtainable through pharmacies: the good, the bad, and the ugly. *Human Genomics*, **7** (1), 17.

10 Borry, P., Evers-Kiebooms, G., Cornel, M. C., Clarke, A., and Dierickx, K. (2009) Genetic testing in asymptomatic minors: background considerations towards ESHG Recommendations. *European Journal of Human Genetics*, **17** (6), 711–719.

11 Quillin, J.M., Bodurtha, J.N., Siminoff, L. A., and Smith, T.J. (2011) Physicians' current practices and opportunities for DNA banking of dying patients with cancer. *Journal of Oncology Practice*, **7** (3), 183–187.

12 Soini, S. (2012) Genetic testing legislation in Western Europe – a fluctuating regulatory target. *Journal of Community Genetics*, **3** (2), 143–153.

13 Douma, K.F., Aaronson, N.K., Vasen, H.F., Verhoef, S., Gundy, C.M., and Bleiker, E. M. (2010) Attitudes toward genetic testing in childhood and reproductive decision-making for familial adenomatous polyposis. *European Journal of Human Genetics*, **18** (2), 186–193.

14 Biesecker, L.G., Burke, W., Kohane, I., Plon, S.E., and Zimmern, R. (2012) Next-generation sequencing in the clinic: are we ready? *Nature Reviews Genetics*, **13** (11), 818–824.

15 Henn, W. (2009) Whole-genome sequencing in diagnostic medicine: too much information for doctors and patients? *Transfusion Medicine and Hemostasis*, **36** (4), 31–32.

16 Netzer, C., Schmitz, D., and Henn, W. (2012) To know or not to know the genomic sequence of a fetus. *Nature Reviews Genetics*, **13** (10), 676–677.

Index

Nucleic Acids as Molecular Diagnostics, First Edition. Edited by Andreas Keller and Eckart Meese.

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley's ebook EULA.