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
Contents

List of Contributors  XIII
Preface  XIX

1 Next-Generation Sequencing for Clinical Diagnostics of Cardiomyopathies  1
Jan Haas, Hugo A. Katus, and Benjamin Meder
1.1 Introduction  1
1.2 Cardiomyopathies and Why Genetic Testing is Needed  1
1.3 NGS  2
1.4 NGS for Cardiomyopathies  2
1.5 Sample Preparation  3
1.6 Bioinformatics Analysis Pipeline  4
1.7 Interpretation of Results and Translation into Clinical Practice  4
References  6

2 MicroRNAs as Novel Biomarkers in Cardiovascular Medicine  11
Britta Vogel, Hugo A. Katus, and Benjamin Meder
2.1 Introduction  11
2.2 miRNAs are Associated with Cardiovascular Risk Factors  12
2.3 miRNAs in Coronary Artery Disease  13
2.4 miRNAs in Cardiac Ischemia and Necrosis  15
2.5 miRNAs as Biomarkers of Heart Failure  19
2.6 Future Challenges  20
Acknowledgments  20
References  21

3 MicroRNAs in Primary Brain Tumors: Functional Impact and Potential Use for Diagnostic Purposes  25
Patrick Roth and Michael Weller
3.1 Background  25
3.2 Gliomas  26
3.2.1 miRNA as Biomarkers in Glioma Tissue  28
3.2.2 Circulating miRNA as Biomarkers 29
3.3 Meningiomas 30
3.4 Pituitary Adenomas 31
3.5 Medulloblastomas 31
3.6 Other Brain Tumors 32
3.6.1 Schwannomas 32
3.6.2 PCNSLs 33
3.7 Summary and Outlook 33
References 34

4 Genetic and Epigenetic Alterations in Sporadic Colorectal Cancer: Clinical Implications 39
Pawel Karpinski, Nikolaus Blin, and Maria M. Sasiadek
4.1 Introduction 39
4.2 Chromosomal Instability 40
4.3 Microsatellite Instability 43
4.4 Driver Somatic Mutations in CRC 46
4.4.1 APC 46
4.4.2 TP53 47
4.4.3 KRAS 47
4.4.4 BRAF 47
4.4.5 PIK3CA 48
4.4.6 Other Mutations 48
4.5 Epigenetic Instability in CRC 48
4.6 Hypomethylation 49
4.7 CpG Island Methylator Phenotype 50
4.8 Concluding Remarks 51
References 51

5 Nucleic Acid-Based Markers in Urologic Malignancies 63
Bernd Wullich, Peter J. Goebell, Helge Taubert, and Sven Wach
5.1 Introduction 63
5.2 Bladder Cancer 64
5.2.1 Hereditary Factors for Bladder Cancer 65
5.2.2 Single Nucleotide Polymorphisms 65
5.2.3 RNA Alterations in Bladder Cancer 66
5.2.3.1 FGFR3 Pathway 66
5.2.3.2 p53 Pathway 67
5.2.3.3 Urine-Based Markers 67
5.2.3.4 Serum-Based Markers 68
5.2.4 Sporadic Factors for Bladder Cancer 69
5.2.5 Genetic Changes in Non-Invasive Papillary Urothelial Carcinoma 69
5.2.5.1 FGFR 3 69
5.2.5.2 Changes in the Phosphatidylinositol 3-Kinase Pathway 70
5.2.6 Genetic Changes in Muscle-Invasive Urothelial Carcinoma 72
5.2.6.1 TP53, RB, and Cell Cycle Control Genes 73
5.2.6.2 Other Genomic Alterations 74
5.2.7 Genetic Alterations with Unrecognized Associations to Tumor Stage and Grade 75
5.2.7.1 Alterations of Chromosome 9 75
5.2.7.2 RAS Gene Mutations 76
5.3 Prostate Cancer 77
5.3.1 Hereditary Factors for Prostate Cancer 77
5.3.2 Sporadic Factors for Prostate Cancer 80
5.3.2.1 PSA and Other Protein Markers 80
5.3.2.2 Nucleic Acid Biomarkers 81
5.3.3 Prostate Cancer: Summary 87
5.4 Renal Cell Carcinoma 87
5.4.1 Hereditary Factors for RCC 87
5.4.2 Sporadic Factors for RCC 90
5.4.2.1 The Old 90
5.4.2.2 The New 91
5.5 Summary 92
References 96

6 From the Genetic Make-Up to the Molecular Signature of Non-Coding RNA in Breast Cancer 129
Michael G. Schrauder and Reiner Strick
6.1 Introduction 129
6.2 Molecular Breast Cancer Detection 130
6.2.1 Circulating Free DNA 130
6.2.2 Long Intergenic Non-Coding RNA 132
6.2.2.1 HOTAIR 132
6.2.2.2 H19 133
6.2.2.3 GAS5 134
6.2.2.4 LSNCT5 134
6.2.2.5 LOC554202 134
6.2.2.6 SRA1 134
6.2.2.7 XIST 134
6.2.3 Natural Antisense Transcripts 135
6.2.3.1 HIF-1a-AS 136
6.2.3.2 H19 and H19-AS (91H) 137
6.2.3.3 SLC22A18-AS 137
6.2.3.4 RPS6KA2-AS 137
6.2.3.5 ZFAS1 137
6.2.4 miRNAs 138
6.2.4.1 Tissue-Based miRNA Profiling in Breast Cancer 138
Contents

6.2.4.2 Circulating miRNAs 141

6.3 Molecular Breast Cancer Subtypes and Prognostic/Predictive Molecular Biomarkers 142

References 144

7 Nucleic Acid-Based Diagnostics in Gynecological Malignancies 155

Sebastian F.M. Häusler, Johannes Dietl, and Jörg Wischhusen

7.1 Introduction 155

7.2 Cervix, Vulva, and Vaginal Carcinoma 155

7.2.1 Background 155

7.2.2 Routine Diagnostics for HPV Infection 157

7.2.2.1 Digene Hybrid Capture 2 High-Risk HPV DNA Test (Qiagen) 158

7.2.2.2 Cervista HPV HR (Holologics) 158

7.2.2.3 cobas 4800 System (Roche) 159

7.2.2.4 APTIMA HPV (Gen-Probe) 159

7.2.2.5 Abbot RealTime High Risk HPV Assay (Abbot) 159

7.2.2.6 PapilloCheck Genotyping Assay (Greiner BioOne) 160

7.2.2.7 INNO-LiPA HPV Genotyping Extra (Innogenetics) 160

7.2.2.8 Linear Array (Roche) 160

7.2.2.9 Recommendations for Clinical Use 160

7.2.3 Outlook – DNA Methylation Patterns 161

7.3 Endometrial Carcinoma (Carcinoma Corpus Uteri) 162

7.3.1 Background 162

7.3.2 Routine Diagnostics – Microsatellite Instability 162

7.3.3 Emerging Diagnostics – miRNA Markers 163

7.4 Ovarian Carcinoma 164

7.4.1 Background 164

7.4.2 Routine Diagnostics 165

7.4.3 Emerging Diagnostics/Perspective – miRNA Profiling 166

7.5 Breast Cancer 167

7.5.1 Background 167

7.5.2 Routine Diagnostics 168

7.5.2.1 HER2 Diagnostics 168

7.5.2.2 Gene Expression Profiling 169

7.5.2.3 Hereditary Breast Cancer/BRCA Diagnostics 170

7.5.3 Emerging Diagnostics/Perspectives 173

7.6 Conclusion 175

References 175

8 Nucleic Acids as Molecular Diagnostics in Hematopoietic Malignancies – Implications in Diagnosis, Prognosis, and Therapeutic Management 185

Janine Schwamb and Christian P. Pallasch

8.1 Introduction 185

8.2 Methodological Approaches 186
8.3 Cytogenetic Analysis to Molecular Diagnostics 186
8.4 Minimal Residual Disease 186
8.5 Chronic Myeloid Leukemia 187
8.6 Acute Myeloid Leukemia 189
8.7 Acute Lymphocytic Leukemia 191
8.8 Chronic Lymphocytic Leukemia 192
8.9 Outlook and Perspectives 196
References 196

9 Techniques of Nucleic Acid-Based Diagnosis in the Management of Bacterial and Viral Infectious Diseases 201
Irene Latorre, Verónica Saludes, Juana Díez, and Andreas Meyerhans
9.1 Importance of Nucleic Acid-Based Molecular Assays in Clinical Microbiology 201
9.2 Nucleic Acid Amplification Techniques 202
9.2.1 Target Amplification Techniques 203
9.2.1.1 PCR-Based Techniques 203
9.2.2 Signal Amplification Techniques 204
9.3 Post-Amplification Analyses 205
9.3.1 Sequencing and Pyrosequencing 205
9.3.2 Reverse Hybridization 206
9.3.3 High-Throughput Nucleic Acid-Based Analyses 206
9.3.3.1 DNA Microarrays 206
9.3.3.2 Mass Spectrometry 207
9.3.3.3 NGS 208
9.4 General Overview and Concluding Remarks 209
Acknowledgments 209
References 209

10 MicroRNAs in Human Microbial Infections and Disease Outcomes 217
Verónica Saludes, Irene Latorre, Andreas Meyerhans, and Juana Díez
10.1 Introduction 217
10.2 General Aspects of miRNAs in Infectious Diseases 218
10.2.1 miRNAs in Bacterial Infections 218
10.2.2 miRNAs in Viral Infections 219
10.2.2.1 Cellular miRNAs Control Viral Infections 220
10.2.2.2 Viruses Use miRNAs for Their Own Benefit 221
10.3 miRNAs as Biomarkers and Therapeutic Agents in Tuberculosis and Hepatitis C Infections 222
10.3.1 Tuberculosis: A Major Bacterial Pathogen 222
10.3.1.1 Tuberculosis Diagnosis and the Need for Immunological Biomarkers 222
10.3.1.2 miRNAs Regulation in Response to M. tuberculosis 223
10.3.1.3 Future Perspectives 225
10.3.2 Chronic Hepatitis C: A Major Viral Disease 225
10.3.2.1 Liver Fibrosis Progression and Treatment Outcome 225
10.3.2.2 miRNAs Involved in Liver Fibrogenesis 226
10.3.2.3 Prediction of Treatment Outcome in Chronic HCV-1 Infected Patients 228
10.3.2.4 Future Perspectives 229
10.4 miRNA-Targeting Therapeutics 230
10.5 Concluding Remarks 230
Acknowledgments 231
References 231

11 Towards the Identification of Condition-Specific Microbial Populations from Human Metagenomic Data 241
Cédric C. Laczy and Paul Wilmes
11.1 Introduction 241
11.2 Nucleic Acid-Based Methods in Diagnostic Microbiology 242
11.2.1 Limitations of Culture-Dependent Approaches 242
11.2.2 Culture-Independent Characterization of Microbial Communities 243
11.2.3 Metagenomics 243
11.2.4 Fecal Samples as Proxies to Evaluate Human Microbiome-Related Health Status 244
11.3 Need for Comprehensive Microbiome Characterization in Medical Diagnostics 244
11.4 Challenges for Metagenomics-Based Diagnostics: Read Lengths, Sequencing Library Sizes, and Microbial Community Composition 248
11.5 Deconvolution of Population-Level Genomic Complements from Metagenomic Data 250
11.5.1 Reference-Dependent Metagenomic Data Analysis 251
11.5.1.1 Alignment-Based Approaches 251
11.5.1.2 Sequence Composition-Based Approaches 253
11.5.2 Reference-Independent Metagenomic Data Analysis 254
11.6 Need for Comparative Metagenomic Data Analysis Tools 256
11.6.1 Reference-Based Comparative Tools 257
11.6.2 Reference-Independent Identification of Condition-Specific Microbial Populations from Human Metagenomic Data 257
11.7 Future Perspectives in Microbiome-Enabled Diagnostics 258
Acknowledgments 262
References 262

12 Genome, Exome, and Gene Panel Sequencing in a Clinical Setting 271
Claudia Durand and Saskia Biskup
12.1 Introduction 271
12.1.1 Genetic Inheritance and Sequencing 271
12.1.2 Genetic Testing by DNA Sequencing 272
12.2 Genetic Diagnostics from a Laboratory Perspective – From Sanger to NGS 273
12.2.1 Sanger Sequencing 273
12.2.2 NGS 274
12.2.3 Practical Workflow: From a Patient’s DNA to NGS Sequencing Analysis 276
12.2.3.1 Preparation of gDNA 277
12.2.3.2 Quality Control 277
12.2.3.3 Library Preparation and Evaluation 277
12.2.3.4 Enrichment 277
12.2.3.5 Quality Control 278
12.2.3.6 Sequencing 278
12.3 NGS Diagnostics in a Clinical Setting – Comparison Between Genome, Exome, and Panel Diagnostics 279
12.3.1 Overview 279
12.3.2 Clinical Application of WGS 279
12.3.3 Clinical Application of WES 282
12.3.4 Clinical Application of Diagnostic Panels 284
12.4 Conclusion and Outlook 287
References 289

13 Analysis of Nucleic Acids in Single Cells 291
Stefan Kirsch, Bernhard Pölzer, and Christoph A. Klein
13.1 Introduction 291
13.2 Isolating Single Cells 291
13.3 Looking at the DNA of a Single Cancer Cell 292
13.4 Molecular DNA Analysis in Single Cells 294
13.5 Approaches to Analyze RNA of a Single Cell 296
13.6 Expression Analysis in Single Cells and its Biological Relevance in Cancer 299
13.7 Thoughts on Bioinformatics Approaches 300
13.8 Future Impact of Single-Cell Analysis in Clinical Diagnosis 301
References 303

14 Detecting Dysregulated Processes and Pathways 309
Daniel Stöckel and Hans-Peter Lenhof
14.1 Introduction 309
14.2 Measuring and Normalizing Expression Profiles 311
14.2.1 Microarray Experiments 311
14.2.2 Normalization 312
14.2.3 Batch Effects 314
14.3 Biological Networks 314
14.4 Measuring the Degree of Deregulation of Individual Genes 315
14.4.1 Microarray Data 316
# Contents

14.4.2 RNA-Seq Data 317
14.5 Over-Representation Analysis and Gene Set Enrichment Analysis 318
14.5.1 Multiple Hypothesis Testing 320
14.5.2 Network-based GSEA Approaches 320
14.6 Detecting Deregulated Networks and Pathways 321
14.7 miRNA Expression Data 326
14.8 Differential Network Analysis 327
14.9 Conclusion 328
References 328

15 Companion Diagnostics and Beyond – An Essential Element in the Puzzle of Transforming Healthcare 335
Jan Kirsten
15.1 Introduction 335
15.2 The Healthcare Environment 335
15.3 What is Companion Diagnostics? 336
15.4 What are the Drivers for Companion Diagnostics? 337
15.5 Companion Diagnostics Market 338
15.6 Partnerships and Business Models for Companion Diagnostics 341
15.7 Regulatory Environment for Companion Diagnostics Tests 342
15.8 Outlook – Beyond Companion Diagnostics Towards Holistic Solutions 344
References 348

16 Ethical, Legal, and Psychosocial Aspects of Molecular Genetic Diagnosis 349
Wolfram Henn
16.1 General Peculiarities of Genetic Diagnoses 349
16.2 Informed Consent and Genetic Counseling 350
16.2.1 Testing of Persons with Reduced Ability to Consent 352
16.3 Medical Secrecy and Data Protection 354
16.4 Predictive Diagnosis 355
16.5 Prenatal Diagnosis 356
16.6 Multiparameter Testing 358
References 359

Index 361
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 Diez*
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  
Wrocław 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 Tumorthherapie  
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 Tumorthherapie  
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
List of Contributors

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
List of Contributors

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

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 channelopathies, 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
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 than 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


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.

Biom. Med.

Cancer Research, 69, 6660–6667.

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

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.

Genome Medicine, 5, 28.

Nature Methods, 7, 248–249.


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. 

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.

Cancer Research, 69, 6660–6667.


Bioinformatics, 26, 2069–2070.