Biomarkers in Disease:
Methods, Discoveries and Applications
Series Editor: Victor R. Preedy

Victor R. Preedy
Vinood B. Patel *Editors*

Biomarkers in Cancer

Springer Reference

Biomarkers in Disease: Methods, Discoveries and Applications

Series Editor

Victor R. Preedy
Department of Nutrition and Dietetics
Division of Diabetes and Nutritional Sciences
Faculty of Life Sciences and Medicine
King's College London
London, UK

In the past decade there has been a sea change in the way disease is diagnosed and investigated due to the advent of high throughput technologies, such as microarrays, lab on a chip, proteomics, genomics, lipomics, metabolomics, etc. These advances have enabled the discovery of new and novel markers of disease relating to autoimmune disorders, cancers, endocrine diseases, genetic disorders, sensory damage, intestinal diseases etc. In many instances these developments have gone hand in hand with the discovery of biomarkers elucidated via traditional or conventional methods, such as histopathology or clinical biochemistry. Together with microprocessor-based data analysis, advanced statistics and bioinformatics these markers have been used to identify individuals with active disease or pathology as well as those who are refractory or have distinguishing pathologies. Unfortunately techniques and methods have not been readily transferable to other disease states and sometimes diagnosis still relies on single analytes rather than a cohort of markers. Furthermore, the discovery of many new markers have not been put into clinical practice, partly because of their cost and partly because some scientists are unaware of their existence or the evidence is still at the preclinical stage. In some cases the work needs further scientific scrutiny. There is thus a demand for a comprehensive and focused evidenced-based text and scientific literature that addresses these issues. Hence the formulation of *Biomarkers in Disease: Methods, Discoveries and Applications*. The series covers a wide number of areas including for example, nutrition, cancer, endocrinology, cardiology, addictions, immunology, birth defects, genetics and so on. The chapters are written by national or international experts and specialists.

Series Titles

1. General Methods in Biomarker Research and Their Applications
2. Biomarkers in Cancer
3. Biomarkers in Cardiovascular Disease
4. Biomarkers in Kidney Disease
5. Biomarkers in Bone Disease
6. Biomarkers in Liver Disease

More information about this series at http://www.springer.com/series/13842

Victor R. Preedy • Vinood B. Patel
Editors

Biomarkers in Cancer

With 176 Figures and 88 Tables

Springer Reference

Editors
Victor R. Preedy
Department of Nutrition and Dietetics
Division of Diabetes and Nutritional Sciences
Faculty of Life Sciences and Medicine
King's College London
London, UK

Vinood B. Patel
Department of Biomedical Sciences
Faculty of Science and Technology
University of Westminster
London, UK

ISBN 978-94-007-7680-7 ISBN 978-94-007-7681-4 (eBook)
ISBN 978-94-007-7682-1 (print and electronic bundle)
DOI 10.1007/978-94-007-7681-4

Library of Congress Control Number: 2015941735

Springer Dordrecht Heidelberg New York London

Printed on acid-free paper

Springer Science+Business Media B.V. Dordrecht is part of Springer Science+Business Media (www.springer.com)

Preface

In the present volume, ***Biomarkers in Cancer***, we have over 40 chapters covering a wide range of conditions, body locations, and cancer types. Their allocations to a traditional grouping presents some difficulty as this may mean having only one chapter in a particular section. Instead, we have adopted a pragmatic approach for ease of navigation and so have the following sections:

- General Aspects: Techniques and Overviews
- Bladder, Kidney, Liver, and Lung
- Brain
- Breast and Prostate
- Cervix and Uterus
- Colorectum
- Head and Neck
- Leukemia and Hodgkin Lymphoma
- Further Knowledge

While the Editors recognize the difficulties in assigning particular chapters to particular sections, the book has enormously wide coverage and includes the following areas, analytes, and platforms: omics, circulating tumor cells, oncoproteomics, cardiotoxicity, DNA methylation, kallikreins, MAP17, CA 19-9, PTTG (Securin), small nuclear RNA, centrosome amplification, cytological specimens, microarrays, cell death markers, epigenetics, molecular markers, maspin, LGR5, 2D-DIGE-MS, imaging, TPS, CD133, mitosis targets, HER2, immunohistochemistry, visceral adipocytes, expression profiling, telomerase, carcinoembryonic antigen family cell adhesion molecules, human papillomavirus (HPV), the NeoMark European project, matrix metalloproteinases, tissue microarrays, FGFR4, whole blood transcriptome, nuclear BMI-1, immunophenotyping, and CD163 and TARC. Tissues and conditions include cancers in general, cancers of the bladder, renal cell, liver, lung, brain, breast, prostate, cervix, endometrium, colorectum, head and neck cancers including the oral cavity, salivary gland, oropharynx, nasopharynx, larynx, leukemia, and Hodgkin lymphoma. Finally, the last chapter is devoted to locating resource material for biomarker

discovery and applications. The chapters are written by national or international experts and specialists.

This book is specifically designed for clinical biochemists, oncologists, scientists, epidemiologists, doctors, and nurses, from students to practitioners at the higher level. It is also designed to be suitable for lecturers and teachers in health care and libraries as a reference guide.

April 2015
London

Victor R. Preedy
Vinood B. Patel

Series Preface

In the past decade, there has been a sea change in the way disease is diagnosed and investigated due to the advent of high-throughput technologies and advances in chemistry and physics, leading to the development of microarrays, lab on a chip, proteomics, genomics, lipomics, metabolomics, etc. These advances have enabled the discovery of new and novel markers of disease relating to autoimmune disorders, cancers, endocrine diseases, genetic disorders, sensory damage, intestinal diseases, and many other conditions too numerous to list here. In many instances, these developments have gone hand in hand with the discovery of biomarkers elucidated via traditional or conventional methods, such as histopathology, immunoassays, or clinical biochemistry. Together with microprocessor-based data analysis, advanced statistics, and bioinformatics, these markers have been used to identify individuals with active disease as well as those who are refractory or have distinguishing pathologies.

Unfortunately, techniques and methods have not been readily transferable to other disease states, and sometimes, diagnosis still relies on a single analyte rather than a cohort of markers. Furthermore, the discovery of many new markers has not been put into clinical practice partly because of their cost and partly because some scientists are unaware of their existence or the evidence is still at the preclinical stage. There is thus a demand for a comprehensive and focused evidence-based text and scientific literature that addresses these issues. Hence, the book series ***Biomarkers in Disease: Methods, Discoveries and Applications***. It imparts holistic information on the scientific basis of health and biomarkers and covers the latest knowledge, trends, and treatments. It links conventional approaches with new platforms. The ability to transcend the intellectual divide is aided by the fact that each chapter has

- *Key Facts* (areas of focus explained for the lay person)
- *Definitions of Words and Terms*
- *Potential Applications to Prognosis, Other Diseases, or Conditions*
- *Summary Points*

The material in ***Potential Applications to Prognosis, Other Diseases, or Conditions*** pertains to speculative or proposed areas of research, cross-transference to

other diseases or stages of the disease, translational issues, and other areas of wide applicability.

The series is expected to prove useful for clinicians, scientists, epidemiologists, doctors and nurses, and also academicians and students at an advanced level.

April 2015 Victor R. Preedy
London

Contents

About the Editors

Victor R. Preedy is a senior member of King's College London (Professor of Nutritional Biochemistry) and King's College Hospital (Professor of Clinical Biochemistry, Honorary). He is attached to both the Diabetes and Nutritional Sciences Division and the Department of Nutrition and Dietetics. He is also founding and current Director of the Genomics Centre and a member of the School of Medicine. Professor Preedy graduated in 1974 with an Honors Degree in Biology and Physiology with Pharmacology. He gained his University of London Ph.D. in 1981. In 1992, he received his Membership of the Royal College of Pathologists, and in 1993, he gained his second Doctoral degree for his contribution to the science of protein metabolism in health and disease. Professor Preedy was elected as a Fellow of the Institute of Biology in 1995 and to the Royal College of Pathologists in 2000. Since then, he has been elected as a Fellow to the Royal Society for the Promotion of Health (2004) and the Royal Institute of Public Health and Hygiene (2004). In 2009, Professor Preedy became a Fellow of the Royal Society for Public Health and in 2012 a Fellow of the Royal Society of Chemistry. In his career, Professor Preedy worked at the National Heart Hospital (part of Imperial College London) and the MRC Centre at Northwick Park Hospital. He has collaborated with research groups in Finland, Japan, Australia, USA, and Germany. He is a leading expert on biomedical sciences and has a long-standing interest in analytical methods and their applications to the study of health and disease. He has lectured nationally and internationally. To his credit, Professor Preedy has published over 500 articles, which includes peer-reviewed manuscripts based on original research, reviews, abstracts, and numerous books and volumes.

Vinood B. Patel is currently a Senior Lecturer in Clinical Biochemistry at the University of Westminster and honorary fellow at King's College London. He presently directs studies on metabolic pathways involved in liver disease, particularly related to mitochondrial energy regulation and cell death. Research is being undertaken to study the role of nutrients, antioxidants, phytochemicals, iron, alcohol, and fatty acids in the pathophysiology of liver disease. Other areas of interest are identifying new biomarkers that can be used for diagnosis and prognosis of liver disease, understanding mitochondrial oxidative stress in Alzheimer's disease, and gastrointestinal dysfunction in autism. Dr. Patel graduated from the University of Portsmouth with a degree in Pharmacology and completed his Ph.D. in protein metabolism from King's College London in 1997. His postdoctoral work was carried out at Wake Forest University Baptist Medical School studying structural–functional alterations to mitochondrial ribosomes, where he developed novel techniques to characterize their biophysical properties. Dr. Patel is a nationally and internationally recognized liver researcher and was involved in several NIH-funded biomedical grants related to alcoholic liver disease. He has edited biomedical books in the area of nutrition and health prevention, autism, and biomarkers and has published over 150 articles. In 2014, he was elected as a Fellow to The Royal Society of Chemistry.

Editorial Advisors

Contributors

Panagiotis G. Adamopoulos Department of Biochemistry and Molecular Biology, Faculty of Biology, School of Science, University of Athens, Athens, Greece

Soo Kyung Ahn Division of Breast and Endocrine, Department of Surgery, College of Medicine, University of Ulsan, Asan Medical Center, Sonpa-gu, Seoul, South Korea

Robert W. Allan Department of Pathology, Immunology and Laboratory Medicine, University of Florida, Gainesville, FL, USA

Eugenia Allegra Otolaryngology-Head and Neck Surgery, Department of Medical and Surgical Sciences, University Magna Graecia of Catanzaro, Catanzaro, Italy

Samer Z. Al-Quran Department of Pathology, Immunology and Laboratory Medicine, University of Florida, Gainesville, FL, USA

Arunkumar Anandharaj Department of Otolaryngology, Head and Neck Surgery, LSU Health Shreveport, Shreveport, LA, USA

Diego Ardigò MultiMed s.r.l., Cremona, Italy

Andrew J. Armstrong Duke Cancer Institute, Duke Prostate Center, Divisions of Medical Oncology and Urology, Duke University School of Medicine, Durham, NC, USA

Min-Hyun Baek Department of Obstetrics and Gynecology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, South Korea

Ioanna Balkouranidou Analysis of Circulating Tumor Cells Lab, Laboratory of Analytical Chemistry, Department of Chemistry, University of Athens, Athens, Greece

Hirendra Nath Banerjee Department of Biological Sciences and Pharmaceutical Sciences, Elizabeth City State University/University of North Carolina, Elizabeth City, NC, USA

Ancha Baranova Center for the Study of Chronic Metabolic Diseases, School of System Biology, George Mason University, Fairfax, VA, USA

Research Centre for Medical Genetics, Russian Academy of Medical Sciences, Moscow, Russia

Rodrigo Barderas Biochemistry and Molecular Biology I Department, Facultad de Ciencias Químicas, Universidad Complutense de Madrid, Madrid, Spain

Tanios Bekaii-Saab Division of Medical Oncology, The Ohio State University Wexner Medical Center, James Cancer Hospital, Columbus, OH, USA

Manfred Beleut Qlaym Healthcare AG, Goettingen, Germany

Rossana Berardi Medical Oncology Unit, Università Politecnica delle Marche, Azienda Ospedaliero-Universitaria, Ospedali Riuniti Umberto I – GM Lancisi – G Salesi, Ancona, Italy

Roberta Bertorelle Immunology and Molecular Oncology Unit, Istituto Oncologico Veneto, IOV-IRCCS, Padova, Italy

Jaime Berumen Departamento de Medicina Experimental, Facultad de Medicina, Universidad Nacional Autonoma de Mexico, Mexico city, Mexico

Unidad de Medicina Genomica, Hospital General de Mexico, Mexico city, Mexico

Natalia Buza Department of Pathology, Yale University School of Medicine, New Haven, CT, USA

Daniela Cardinale Unit of Cardioncology, Division of Cardiology, European Institute of Oncology, Milan, Italy

Amancio Carnero Instituto de Biomedicina de Sevilla (IBIS), Hospital Universitario Virgen del Rocio, Consejo Superior de Investigaciones Cientificas, Universidad de Sevilla, Sevilla, Spain

Christopher M. Carter Department of Pathology, Immunology and Laboratory Medicine, University of Florida, Gainesville, FL, USA

J. Ignacio Casal Department of Cellular and Molecular Medicine, Functional Proteomics Laboratory, Centro de Investigaciones Biológicas (CSIC), Madrid, Spain

Stefano Cascinu Medical Oncology Unit, Università Politecnica delle Marche, Azienda Ospedaliero-Universitaria, Ospedali Riuniti Umberto I – GM Lancisi – G Salesi, Ancona, Italy

Jimmy Yu-Wai Chan Department of Surgery, Queen Mary Hospital, The University of Hong Kong, Hong Kong, Hong Kong

Samuel Chao GeneNews Limited, Markham, ON, Canada

Kellie A. Charles School of Medical Sciences (Pharmacology), Sydney Medical School, University of Sydney, St Leonards, NSW, Australia

Wei Chua Department of Medical Oncology, Liverpool Hospital, Liverpool, NSW, Australia

Caroline Chung Department of Radiation Oncology, University of Toronto/ University Health Network-Princess Margaret Cancer Centre, Toronto, ON, Canada

Stephen J. Clarke Kolling Institute of Medical Research, Royal North Shore Hospital, University of Sydney, St Leonards, NSW, Australia

Alessandro Colombo Unit of Cardioncology, Division of Cardiology, European Institute of Oncology, Milan, Italy

Ana Conesa Genomics of Gene Expression Laboratory, Centro de Investigación Príncipe Felipe (CIPF), Valencia, Spain

Chiara Copelli Maxillofacial Surgery, University-Hospital of Parma, Parma, Italy

Claudia Corbo CEINGE-Biotecnologie avanzate, s.c.a r.l, Naples, Italy

Manuel José Rolo da Fonseca Link Consulting – Tecnologias De Informaçao S.A., Lisbon, Portugal

Adriana Madeira Álvares da Silva Conforti Department of Biology, Federal University of Espirito Santo – Center of Agrarian Sciences (CCA/UFES), Alegre, ES, Brazil

Oslei Paes de Almeida Piracicaba Dental School, Department of Oral Diagnosis (Pathology), State University of Campinas, Areão, Piracicaba, São Paulo, Brazil

Marco De Fazio STMicroelectronics, Agrate Brianza (MI), Italy

Anita De Rossi Department of Surgery, Oncology and Gastroenterology, Section of Oncology and Immunology, University of Padova, and Istituto Oncologico Veneto, IOV-IRCCS, Padova, Italy

Istituto Oncologico Veneto, Padova, Italy

Luigi Del Vecchio CEINGE-Biotecnologie avanzate, s.c.a r.l, Naples, Italy

Dipartimento di Medicina Molecolare e Biotecnologie Mediche, Università di Napoli "Federico II", Naples, Italy

Connie I. Diakos Kolling Institute of Medical Research, Royal North Shore Hospital, University of Sydney, St Leonards, NSW, Australia

Roberta Lelis Dutra Department of Pathology, University of Sao Paulo, Laboratório de Citogenômica, LIM 03, São Paulo, SP, Brazil

Oleksandr Ekshyyan Department of Otolaryngology, Head and Neck Surgery, LSU Health Shreveport, Shreveport, LA, USA

Diana P. English Department of Obstetrics, Gynecology and Reproductive Sciences, Yale University School of Medicine, New Haven, CT, USA

Ana María Espinosa Unidad de Medicina Genomica, Hospital General de Mexico, Mexico city, Mexico

Konstantinos Exarchos University of Ioannina, Ioannina, Greece

Felipe Paiva Fonseca Piracicaba Dental School, Department of Oral Diagnosis (Pathology), State University of Campinas, Areão, Piracicaba, São Paulo, Brazil

School of Clinical Dentistry, Department of Oral and Maxillofacial Pathology, The University of Sheffield, Claremont Crescent, Sheffield, UK

James Freeman School of Medical Sciences, Edith Cowan University, Perth, WA, Australia

Shinichi Fukushige Department of Molecular Pathology, Center for Regulatory Epigenome and Diseases, Tohoku University School of Medicine, Sendai, Miyagi, Japan

Maher K. Gandhi University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, QLD, Australia

Wei Gao Department of Surgery, Queen Mary Hospital, The University of Hong Kong, Pokfulam, Hong Kong

Miguel Pérez-Guillermo García Department of Pathology, Hospital General Universitario Santa Lucia (HGUSL), Cartagena, Spain

Marica Gemei CEINGE-Biotecnologie avanzate, s.c.a r.l, Naples, Italy

Violaine Goidts Division of Molecular Genetics (B060), German Cancer Research Center (DKFZ), Heidelberg, Germany

Elin S. Gray School of Medical Sciences, Edith Cowan University, Perth, WA, Australia

Mariano Guardado Unidad de Medicina Genomica, Hospital General de Mexico, Mexico city, Mexico

Stephan A. Hahn Department of Molecular Gastrointestinal Oncology, Ruhr-University Bochum, Molecular GI-Oncology (MGO), Bochum, Germany

Ji-Hyun Han Department of Obstetrics and Gynecology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, South Korea

Liz Y. Han Division of Gynecologic Oncology, Kaiser Permanente Northern California, Oakland, CA, USA

Wonshik Han Division of Breast and Endocrine, Department of Surgery, College of Medicine, University of Ulsan, Asan Medical Center, Sonpa-gu, Seoul, South Korea

Woo-Suk Han Department of Obstetrics and Gynecology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, South Korea

Stefan Holdenrieder Institute of Clinical Chemistry and Clinical Pharmacology, University Hospital Bonn, Bonn, Germany

Akira Horii Department of Molecular Pathology, Tohoku University School of Medicine, Sendai, Miyagi, Japan

Viive M. Howell Kolling Institute of Medical Research, Royal North Shore Hospital, University of Sydney, St Leonards, NSW, Australia

Kimberley Jones University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, QLD, Australia

Dimitrios N. Kanakis Department of Pathology, Medical Faculty, Democritus University of Thrace, Alexandroupolis, Greece

Jisun Kim Division of Breast and Endocrine, Department of Surgery, College of Medicine, University of Ulsan, Asan Medical Center, Sonpa-gu, Seoul, South Korea

Elmar Kirches Department of Neuropathology, Medical Faculty, Otto-von-Guericke University, Magdeburg, Germany

Athina Kladi-Skandali Department of Biochemistry and Molecular Biology, University of Athens, Athens, Greece

A-Ra Ko Department of Obstetrics and Gynecology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, South Korea

Nikolaus Kohles Department of Otorhinolaryngology, Helios-Klinikum Erfurt, Erfurt, Germany

Christos K. Kontos Department of Biochemistry and Molecular Biology, Faculty of Biology, School of Science, University of Athens, Athens, Greece

Kavya Krishna Department of Hematology and Oncology, The Ohio State University Wexner Medical Center, James Cancer Hospital, Columbus, OH, USA

Feryel Ksiâa Department of Pathology, University Hospital Farhat Hached, Sousse, Tunisia

Davide Lanfranco Maxillofacial Surgery, University-Hospital of Parma, Parma, Italy

Michelle Mei Lin Lee GeneNews (Malaysia) Limited, Georgetown, Penang, Malaysia

Jing Li Duke University and Cancer Institute, Durham, NC, USA

Ying Li Department of Pathology, Immunology and Laboratory Medicine, University of Florida, Gainesville, FL, USA

Zeng-Hong Li Department of Surgery, The University of Hong Kong, Pokfulam, Hong Kong

Evi S. Lianidou Analysis of Circulating Tumor Cells Lab, Laboratory of Analytical Chemistry, Department of Chemistry, University of Athens, Athens, Greece

Choong Chin Liew GeneNews Diagnostics, Kuala Lumpur, Malaysia

Chun Ren Lim GeneNews (Malaysia) Limited, Georgetown, Penang, Malaysia

Yvonne G. Lin Division of Gynecologic Oncology, Department of Obstetrics-Gynecology, Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA, USA

Petr Lokhov Department of Proteomics and Mass-Spectrometry, Orekhovich Institute of Biomedical Chemistry, Russian Academy of Medical Sciences, Moscow, Russia

Iúri Drumond Louro Department of Biology, Federal University of Espirito Santo (UFES), Human and Molecular Genetics Center, Vitória, ES, Brazil

Abdelouahid Maghnouj Department of Molecular Gastrointestinal Oncology, Ruhr-University Bochum, Molecular GI-Oncology (MGO), Bochum, Germany

Athina Markou Analysis of Circulating Tumor Cells Lab, Laboratory of Analytical Chemistry, Department of Chemistry, University of Athens, Athens, Greece

Elena Martinelli Maxillofacial Surgery, University-Hospital of Parma, Parma, Italy

Tasuku Matsuoka Department of Surgical Oncology, Osaka City University Graduate School of Medicine, Osaka, Japan

Hideyasu Matsuyama Department of Urology, Graduate School of Medicine, Yamaguchi University, Ube, Japan

Ingrid Medina Unidad de Medicina Genomica, Hospital General de Mexico, Mexico city, Mexico

Vikas Mehta Department of Otolaryngology, Head and Neck Surgery, LSU Health Shreveport, Shreveport, LA, USA

Suzanny Oliveira Mendes Programa de Pós-graduação em Biotecnologia, Núcleo de Genética Humana e Molecular, Federal University of Espirito Santo (UFES), Vitória, ES, Brazil

Kleita Michaelidou Department of Biochemistry and Molecular Biology, University of Athens, Athens, Greece

Hyeong-Gon Moon Division of Breast and Endocrine, Department of Surgery, College of Medicine, University of Ulsan, Asan Medical Center, Sonpa-gu, Seoul, South Korea

Tara Moore-Medlin Department of Otolaryngology, Head and Neck Surgery, LSU Health Shreveport, Shreveport, LA, USA

Francesca Morgese Medical Oncology Unit, Università Politecnica delle Marche, Azienda Ospedaliero-Universitaria, Ospedali Riuniti Umberto I – GM Lancisi – G Salesi, Ancona, Italy

Sergei Moshkovskii Department of Personalized Medicine, Orekhovich Institute of Biomedical Chemistry, Russian Academy of Medical Sciences, Moscow, Russia

Medico-Biological Faculty, Pirogov Russian National Research Medical University, Moscow, Russia

Susumu Nakata Division of Oncological Pathology, Aichi Cancer Center Research Institute, Nagoya, Japan

Joo-Hyun Nam Department of Obstetrics and Gynecology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, South Korea

Cherie-Ann O. Nathan Department of Otolaryngology, Head and Neck Surgery, LSU Health Shreveport, Shreveport, LA, USA

Claudio Navarre Genomics of Gene Expression Laboratory, Centro de Investigación Príncipe Felipe (CIPF), Valencia, Spain

Dong-Young Noh Division of Breast and Endocrine, Department of Surgery, College of Medicine, University of Ulsan, Asan Medical Center, Sonpa-gu, Seoul, South Korea

Azzurra Onofri Medical Oncology Unit, Università Politecnica delle Marche, Azienda Ospedaliero-Universitaria, Ospedali Riuniti Umberto I – GM Lancisi – G Salesi, Ancona, Italy

Jeong-Yeol Park Department of Obstetrics and Gynecology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, South Korea

Vinood B. Patel Department of Biomedical Sciences, Faculty of Science and Technology, University of Westminster, London, UK

Gabriela Tonini Peterle Programa de Pós-graduação em Biotecnologia, Núcleo de Genética Humana e Molecular, Federal University of Espirito Santo (UFES), Vitória, ES, Brazil

Emma Phillips Division of Molecular Genetics (B060), German Cancer Research Center (DKFZ), Heidelberg, Germany

Marco Picone Department of Information Engineering, University of Parma, MultiMed s.r.l., Cremona, Italy

Tito Poli Maxillofacial Surgery, University-Hospital of Parma, Parma, Italy

Victor R. Preedy Department of Nutrition and Dietetics, Division of Diabetes and Nutritional Sciences, Faculty of Life Sciences and Medicine, King's College London, London, UK

Mikhail Pyatnitsky Department of Bioinformatics, Orekhovich Institute of Biomedical Chemistry, Russian Academy of Medical Sciences, Moscow, Russia

Rajkumar Rajendram Department of General Medicine and Intensive Care, John Radcliffe Hospital, Oxford, UK

Diabetes and Nutritional Sciences Research Division, School of Medicine, King's College London, London, UK

Roshanna Rajendram School of Medicine, University of Birmingham, Edgbaston, Birmingham, UK

Dana Marie Roque Division of Gynecologic Oncology, Greenebaum Cancer Center, University of Maryland-Baltimore, Baltimore, MD, USA

Michela Salvatici Division of Laboratory Medicine, European Institute of Oncology, Milan, Italy

Francesco Salvatore CEINGE-Biotecnologie avanzate, s.c.a r.l, Naples, Italy

Dipartimento di Medicina Molecolare e Biotecnologie Mediche, Università di Napoli "Federico II", Naples, Italy

IRCCS-SDN, Naples, Italy

Dario Salvi Life Supporting Technologies, ETSI Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain

Maria Teresa Sandri Division of Laboratory Medicine, European Institute of Oncology, Milan, Italy

Alessandro D. Santin Department of Obstetrics, Gynecology and Reproductive Sciences, Yale University School of Medicine, New Haven, CT, USA

Marcelo dos Santos Department of Medicine, Federal University of Rio Grande do Norte (UFRN) – Campus Caicó, Caicó, RN, Brazil

Shinobu Sato Department of Applied Chemistry, Research Center for Biomicrosensing Technology, Kyushu Institute of Technology, Kitakyushu, Fukuoka, Japan

Agnese Savini Medical Oncology Unit, Università Politecnica delle Marche, Azienda Ospedaliero-Universitaria, Ospedali Riuniti Umberto I – GM Lancisi – G Salesi, Ancona, Italy

Peter Schraml Institute of Surgical Pathology, University Hospital Zurich, Zurich, Switzerland

Andreas Scorilas Department of Biochemistry and Molecular Biology, Faculty of Biology, School of Science, University of Athens, Athens, Greece

Enrico Sesenna Maxillofacial Surgery, University-Hospital of Parma, Parma, Italy

Mari Carmen Turpín Sevilla Department of Pathology (HGUSL) Cartagena, Spain; University Francisco of Vitoria, Madrid, Spain

Louise Seymour University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, QLD, Australia

José García Solano Department of Pathology, Hospital General Universitario Santa Lucia (HGUSL), Cartagena, Spain

Facultad de Ciencias de la Salud, Catholic University of Murcia (UCAM), Murcia, Spain

Ioanna Sourvinou Analysis of Circulating Tumor Cells Lab, Laboratory of Analytical Chemistry, Department of Chemistry, University of Athens, Athens, Greece

Paul M. Speight School of Clinical Dentistry, Department of Oral and Maxillofacial Pathology, The University of Sheffield, Claremont Crescent, Sheffield, UK

Sebastian Steger Fraunhofer Institute for Computer Graphics (IGD) Cognitive Computing and Medical Imaging (A7), Darmstadt, Germany

Shigeori Takenaka Department of Applied Chemistry, Research Center for Biomicrosensing Technology, Kyushu Institute of Technology, Kitakyushu, Fukuoka, Japan

Serena Trapasso Otolaryngology-Head and Neck Surgery, Department of Medical and Surgical Sciences, University Magna Graecia of Catanzaro, Catanzaro, Italy

Mounir Trimeche Department of Pathology, University Hospital Farhat Hached, Sousse, Tunisia

Deepak B. Vangala Department of Medicine and Department of Molecular Gastrointestinal Oncology, Ruhr-University Bochum, Knappschaftskrankenhaus, Bochum, Germany

Pablo Agustin Vargas Piracicaba Dental School, Department of Oral Diagnosis (Pathology), State University of Campinas, Areão, Piracicaba, São Paulo, Brazil

Department of Oral Pathology and Oral Biology, School of Dentistry, Faculty of Health Sciences, University of Pretoria, Pretoria, South Africa

Mukesh Verma Methods and Technologies Branch, Epidemiology and Genetics Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute (NCI), National Institutes of Health (NIH), Rockville, MD, USA

Roi Villar-Vázquez Department of Cellular and Molecular Medicine, Functional Proteomics Laboratory, Centro de Investigaciones Biológicas (CSIC), Madrid, Spain

Thian-Sze Wong Department of Surgery, Queen Mary Hospital, The University of Hong Kong, Pokfulam, Hong Kong

Masakazu Yashiro Department of Surgical Oncology, Osaka City University Graduate School of Medicine, Osaka, Japan

Adel Zaatar Gleneagles Medical Centre, Penang, Malaysia

Pablo Conesa Zamora Department of Pathology, Hospital General Universitario Santa Lucia (HGUSL), Cartagena, Spain

Facultad de Ciencias de la Salud, Catholic University of Murcia (UCAM), Murcia, Spain

Mel Ziman School of Medical Sciences, Edith Cowan University, Perth, WA, Australia

Hannah Zoellner Department of Molecular Gastrointestinal Oncology, Ruhr-University Bochum, Molecular GI-Oncology (MGO), Bochum, Germany

Part I

General Aspects: Techniques and Overviews

OMICS for Tumor Biomarker Research

1

Sergei Moshkovskii, Mikhail Pyatnitsky, Petr Lokhov, and Ancha Baranova

Contents

S. Moshkovskii (✉)
Department of Personalized Medicine, Orekhovich Institute of Biomedical Chemistry, Russian Academy of Medical Sciences, Moscow, Russia

Medico-Biological Faculty, Pirogov Russian National Research Medical University, Moscow, Russia
e-mail: smosh@mail.ru; sergei.moshkovskii@ibmc.msk.ru

M. Pyatnitsky
Department of Bioinformatics, Orekhovich Institute of Biomedical Chemistry, Russian Academy of Medical Sciences, Moscow, Russia
e-mail: mpyat@mail.ru

P. Lokhov
Department of Proteomics and Mass-Spectrometry, Orekhovich Institute of Biomedical Chemistry, Russian Academy of Medical Sciences, Moscow, Russia
e-mail: lokhovpg@rambler.ru

A. Baranova
Center for the Study of Chronic Metabolic Diseases, School of System Biology, George Mason University, Fairfax, VA, USA

Research Centre for Medical Genetics, Russian Academy of Medical Sciences, Moscow, Russia
e-mail: abaranov@gmu.edu; aancha@gmail.com

V.R. Preedy, V.B. Patel (eds.), *Biomarkers in Cancer, Biomarkers in Disease: Methods, Discoveries and Applications*,
DOI 10.1007/978-94-007-7681-4_14

Abstract

OMICS generally refers to a study of some gene expression products, either direct, such as RNA and proteins, or indirect, such as metabolites, and is usually based on genome information. Main sections of OMICS sciences include transcriptomics, proteomics, and metabolomics, powerful research instruments capable of high-throughput detection of biomolecules differentially expressed between tumor and non-tumor samples, including excised tissues or biopsies, blood plasma, saliva, and urine. Consequently, thousands of species of RNAs, proteins, and metabolites were suggested as candidate tumor biomarkers alone or as constituents of multiplex signatures. Despite many difficulties encountered by OMICS panels with an intended use in population screening programs, some of the multiplex panels already have found their applications in the field of theranostics. If the patient is already diagnosed with a certain cancer, RNA or protein biomarker signatures may help to select a specific therapy or to predict the probability of a relapse. A number of clinically relevant, validated, and approved signatures of RNA and protein analytes successfully emerged from OMICS pipelines. It is important to remember that an implementation of these clinical tests took the safety of reliable laboratory techniques, such as polymerase chain reaction and immunoassay.

List of Abbreviations

AUC	Area Under the Curve
DNA	Deoxyribonucleic Acid
ENCODE	Encyclopedia of DNA Elements
ESI	Electrospray Ionization
FDA	US Food and Drug Administration
HPLC	High-Performance Liquid Chromatography
IVDMIA	In Vitro Diagnostic Multivariate Index Assay
LC-MS/MS	Liquid Chromatography-Tandem Mass Spectrometry
LDT	Laboratory-Developed Tests
LOOCV	Leave-One-Out Cross Validation
MALDI-TOF	Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry
MRM	Multiple Reaction Monitoring
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
m/z	Molecular Mass/Charge Ratio
mRNA	Matrix Ribonucleic Acid
miRNA	Micro-Ribonucleic Acid
NMR	Nuclear Magnetic Resonance
PCR	Polymerase Chain Reaction
PPV	Positive Predictive Value
qRT-PCR	Quantitative Real-Time Polymerase Chain Reaction
RNA	Ribonucleic Acid

RNAseq	High-Throughput Sequencing of Ribonucleic Acid
ROC	Receiver Operator Characteristics
SELDI	Surface-Enhanced Laser Desorption/Ionization
SRM	Selected Reaction Monitoring

Key Facts of OMICS Sciences

OMICS sciences are focused on the inventory of multiple molecular species in living organisms or their parts.

Each OMICS discipline is designated using the name of molecular species studied plus the "-omics" suffix, such as proteomics for proteins, lipidomics for lipids, glycomics for glycans, etc.

OMICS sciences for medicine have become possible after human genome sequencing.

Some OMICS sciences, such as transcriptomics and proteomics, are directly based on genome sequence, whereas others are indirectly related to genome, e.g., metabolomics.

OMICS sciences are used to compare levels of multiple molecular species between diseased and healthy control tissues or cells to discover differential molecules, i.e., biomarkers.

Today, transcriptomics, proteomics, and metabolomics are most widely used for biomarker discovery.

Transcriptomics uses nucleic acid microarrays and high-throughput nucleic acid sequencing to catalog RNA molecules.

Proteomics uses mass spectrometry and affinity reagents, such as antibodies, on protein microarrays, for protein inventory.

Metabolomics detects low-molecular metabolites by mass spectrometry or nucleic magnetic resonance spectroscopy.

Definitions of Words and Terms

Transcriptome Transcriptome is a whole of genome transcripts, i.e., RNAs, which are contained in a cell, tissue, or organism. The high-throughput detection and quantitation of multiple RNAs based on genome sequence information is transcriptomics.

Proteome Proteome is a whole of proteins, which are contained in a cell, tissue, or organism. High-throughput detection and quantitation of multiple proteins based on gene sequence information is proteomics.

Metabolome Metabolome is a whole of low-molecular substances, i.e., metabolites, which are contained in a cell, tissue, or organism. High-throughput detection

and quantitation of multiple metabolites based on their physical and chemical properties is metabolomics.

Liquid Chromatography-Tandem Mass-Spectrometry (LC-MS/MS) LC-MS/MS is an analytical method, where molecules of interest are separated by high-performance liquid chromatography which is coupled with mass spectrometer with electrospray ionization. Such mass spectrometer performs tandem mass spectrometry. It measures the molecular mass/charge ratio (m/z) of original ions which may be further fragmented in mass spectrometer and analyzed for more structural information. LC-MS/MS is widely used in shotgun and targeted proteomics and metabolomics. Detectors used for tandem mass spectrometry include, but not limited to, ion trap, quadrupole time of flight (Q-TOF), triple quadrupole (QQQ), Orbitrap, etc.

Electrospray Ionization (ESI) ESI is a method for ionization of molecules based on the application of electrical field to the molecules which are sprayed in small drops of liquid. It is a mild method of ionization which preserves the structure of biomolecules for detection by mass spectrometry. ESI is widely used in proteomics.

Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS) MALDI-TOF MS is a specific method of mass spectrometry. It uses ionization of solid substances co-crystallized with a laser-sensitive "matrix" compound. After laser shot, a molecule analyzed is ionized and desorbed from the crystal state and then analyzed by time-of-flight detector. MALDI-TOF MS is a fast and easy method to analyze proteins and peptides.

Multiple Reaction Monitoring (MRM) Multiple reaction monitoring or selected reaction monitoring (SRM) is a method of targeted tandem mass spectrometry. In the MRM mode, a mass spectrometer can detect only molecules with expected narrow range of m/z and its expected fragments. ESI and a triple quadrupole detector are widely used for MRM. This method was originally used for quantitation of small molecules, such as drugs and chemicals. Since 2005, it was suggested for use to quantify peptides in proteomics. An isotope synthetic standard is required for proper measurement quality.

Introduction: A Role of OMICS in Cancer Biomarker Research

OMICS is a recently generated word that was coined by molecular biologists that figured how to analyze molecular entities in a high-throughput manner. In the late 1990s, the genomics became the very first "OMICS" field assessed in this manner. Respectively, the term "genomics" was derived from "genome," yet another recent addition to the dictionaries (Winkler 1920).

Each "OMICS" discipline is focused on its own "-ome." After the genome, a proteome concept was suggested (Wilkins et al. 1996) to represent a sum of proteins

within the cell, tissue, or organism of interest. Simultaneously, many other classes of biomolecules got their own "-omes" and "OMICS" – transcriptomics for RNAs, metabolomics for metabolites, and lipidomics for lipids. In addition to "OMICS"-based inventories of various molecular entities, some "OMICS" disciplines are focused on the inventories of events, such as interactomics that systematically analyze interactions between various macromolecules (Cesareni et al. 2005).

It is important to note that the genomics stands apart from other OMICS disciplines as it serves as a background for others. In this context, transcriptomics and proteomics are usually referred to as postgenome technologies. Indeed, deciphering the genome of given species makes it possible for other OMICS sciences to emerge. In this chapter, we will not focus on cancer genomics due to the heaps of information already available. Here, OMICS will generally refer to a study of some gene expression products, either direct, such as RNA and proteins, or indirect, such as metabolites.

A majority of sporadic cancers are due to random somatic mutagenesis by way of environmental exposures and endogenous stress that lead to epigenetic deregulation of expression patterns within the cell (Amin et al. 2009). That is why high-throughput versions of mRNA expression analysis are widely used as means for cancer biomarker discovery. An accumulation of somatic or germ line mutations in chromosomal DNA is recognized as major reason for the proliferative features of cancer phenotype. These crucial genomic events are usually designated as driver mutations (Bignell et al. 2010) that may be caused by viral genome insertions, radiation, chemical mutagens, and other environmental carcinogens. How many driver mutations are exactly necessary and/or sufficient to make a viable cancer cell is a topic of hot discussion. As evident from the studies of transgenic and knockout mice, in most cases these numbers are minimal. However, naturally developed tumors are represented by a mix of competing clones varying in their malignant potential and genomic structure. Hence, the DNA extracted from a piece of tumor tissue may harbor thousands of mutations. Indeed, recent efforts in tumor exome sequencing confirmed these findings (Cancer Genome Atlas Research Network 2012; Stephens et al. 2009) and created solid grounds for generation of cancer-type specific genome atlases that provide a knowledge base for modern biomarker discovery.

The primary difficulty with the biomarker-guided detection of the tumors in general population is due to the multiclonal composition of individual tumors and the resultant variation in the levels and the spectrum of biological molecules expressed by tumors that originated within the same tissue. That is why, with a few notable exceptions, single molecular biomarkers, such as mRNA, miRNA, protein, or metabolite, are rarely successful as population screening tools. Hence, a molecular signature concept was developed for both diagnostic and theranostic applications (Subramanian and Simon 2010; Zimmer et al. 2006). These molecular signatures are often derived from OMICS data. In case of tumor detection, the prevalence of the somatic component makes the OMICS-based approaches especially suitable (Fig. 1). The cancers located within the same organ display substantially different proteome profiles that accurately reflect morphological subtype of the tumor (Kobel et al. 2008). Profiling-based studies may focus on any type of molecular biomarkers and may include somatic cancer mutations within coding

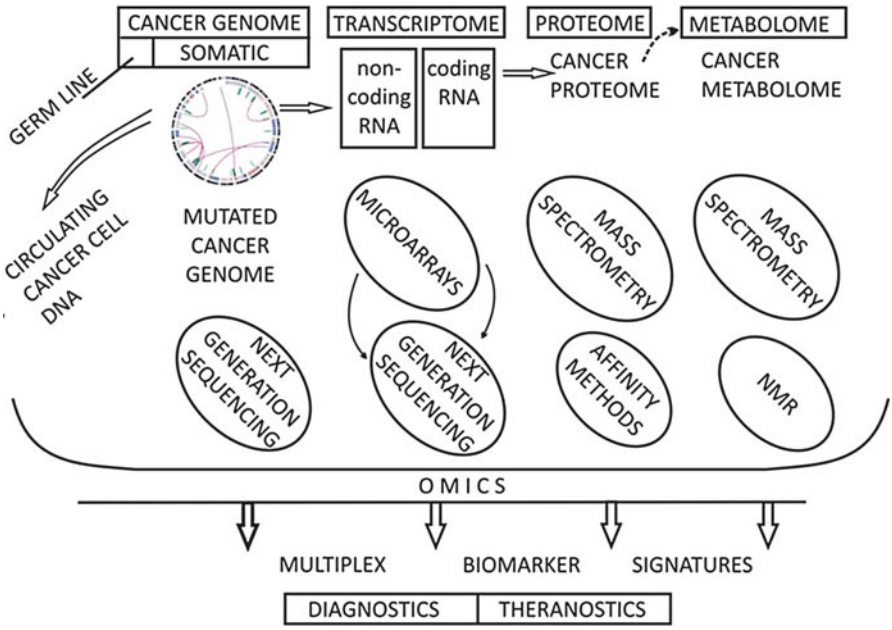

Fig. 1 Outline of main OMICS technologies used for tumor biomarker discovery: from cancer genome to diagnostic and theranostic signature

transcripts (Hawkins et al. 2010), miRNAs (Ferracin et al. 2010), proteins (Kim et al. 2009), or metabolites (Aboud and Weiss 2013).

Measured together, molecular variables of the signature ensure higher diagnostic accuracy or more efficient risk prediction as compared to single biomarkers (Yurkovetsky et al. 2010). The techniques for discerning diagnostic molecular signatures depend on the type of the molecules to be detected. For protein signatures, immunoassays (Edgell et al. 2010) or mass spectrometry (Rodriguez et al. 2010) may be used, while for the detection of mRNA levels, one may use a qRT-PCR, microarrays, or RNAseq. However, the use of "mixed" panels, for example, ones that include both mRNA and protein biomarkers, is limited by substantial increase in costs of performing an assay. The conversion of a series of experimentally quantified values into clinically relevant test is a long journey. The multiplexed signatures require application of the complex statistical techniques, including pattern recognition approaches (Alonzo and Pepe 2007; Hamacher et al. 2009).

An Analysis Pathway for Multiplex Biomarker Panels

A typical clinical classification and risk prediction framework is implemented as follows. Experimental data are collected for samples representing two classes, for example, healthy individuals versus patients with a disease, or two cohorts of

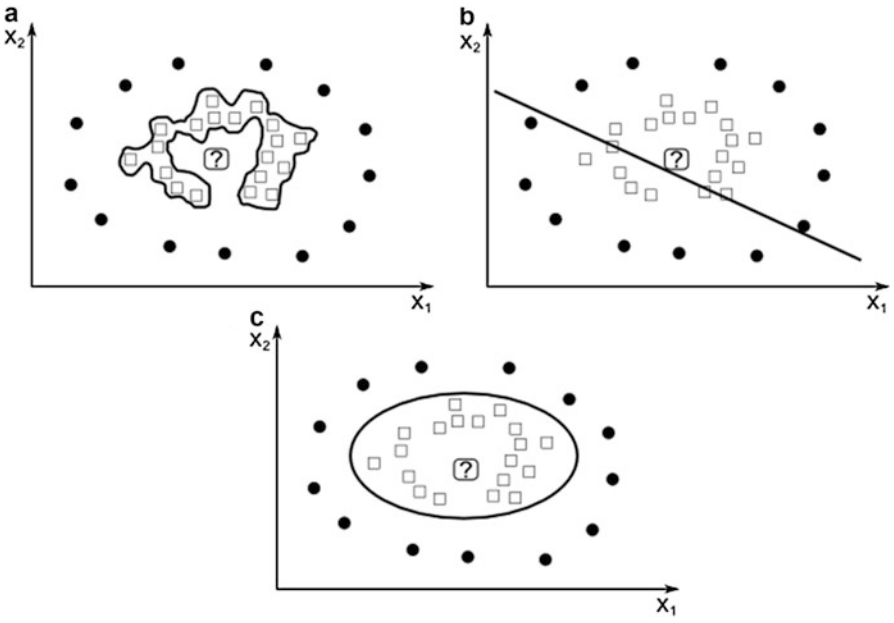

Fig. 2 Illustration of model overfitting and generalization. (**a**) Model with many adjustable parameters is capable of error-free classification of training set but performs poorly on novel data. (**b**) Too simple model does not allow discrimination of training data. (**c**) The model with optimum balance between overfitting and generalization

patients, one with poor and one with good prognosis as measured post hoc, by disease outcomes. Variables may represent gene expression, peak intensities from proteomic or metabolomic spectra, results of genotyping, blood biochemistry values, or demographics. These variables require intensive preprocessing, including imputation of missing values, normalization, smoothing out of the noise, and removing of outliers. A distinctive feature of OMICS data is that the number of samples (N) is orders of magnitude smaller than the available number of variables (P). In this context, a variable is often called a feature and refers to a specific gene or protein among the many thousands of molecules assayed in parallel. This "small n, large p" issue greatly complicates the statistical analysis of OMICS experiments and has many consequences. When the volume of the data grows exponentially, the information needed to describe the feature space with the same sampling density also should grow exponentially (so-called curse of dimensionality). In practical sense, this situation described as sparseness of the data may substantially complicate the building of classification models; good separation of the classes may be achieved even for sets of classifiers chosen randomly (Venet et al. 2011) (Fig. 2).

The obvious solution to "dimensionality curse" is to remove a majority of variables, i.e., perform feature selection (Saeys et al. 2007). Indeed, elimination of the features that carry no association with the disease or its outcome aids the discrimination of the sample classes and increases the classification power.

Additionally, it makes sense to eliminate variables that are highly correlated with each other, for example, co-regulated genes, or mass spectrometry peaks that represent modified forms of the same protein (Pyatnitskiy et al. 2010). Leaving only one "best-in-class" feature for every co-correlated group also aids in the interpretation of the results, as these features are more likely to represent highly interacting genes that correspond to the hubs in regulatory networks. Another possible way to attack dimensionality is identify and interpret a pattern within the data. For that, a number of visualization approaches were developed, for example, principal components analysis, clustering, or multidimensional scaling. It is important to note that the visualization of the data is not a required component of multiplexed biomarker test; rather, it serves as a foothold that helps a researcher to gain confidence with massive dataset.

The Holy Grail of biomarker discovery is the building of a decent classification model. This process includes two major steps: the learning step that extracts information contained in training dataset and model evaluation using validation dataset. The ultimate goal is to develop a model that withstands validation step, i.e., shows required accuracy when tested on independent set of samples set and, thus, deemed suitable for real-world applications. To develop the model that is more likely to pass validation, the model's parameters are fine-tuned to optimize its performance on the training set. An accurate guess whether the model will survive validation is impossible. Perfect classification of training samples does not guarantee good performance in independently collected samples; when good-looking model does perform, we call this model "overfitted" to the training set. When high-dimensional data are inputted in models with many adjustable parameters, the "overfitting" is commonplace. On the other hand, an oversimplified model may not be able to discriminate the samples at all. Thus, the development of the biomarker panel is always a trade-off between overfitting and over simplification.

A great number of classification algorithms have been utilized in the analysis of OMICS data. The most frequently used classifiers include support vector machines (SVM), artificial neural networks, decision trees and random forests, logistic regression, and many others. Comparative studies of various approaches to improve discrimination power of the multidimensional models in terms of prediction accuracy, specificity, and sensitivity had been performed and published earlier. The results of comparisons contradict to each other – in certain cases more sophisticated algorithms like support vector machines outperform others, while in some cases relatively simple techniques show comparable classification accuracy.

It is also important to note that the importance of standard performance metrics is often overestimated (such as area under ROC curve or AUC). As we already mentioned above, overfitted models are unlikely to survive validation in independent cohorts, while underperforming models that rely on solid biological knowledge may actually improve their sensitivity and specificity when validated in larger cohorts. For example, the models generated by neural networks often show superior performance, but, in essence, they remain "black boxes" unavailable for meaningful interpretation. On the other side, the decision trees or the logistic regression

models allow easy derivation of relationships between feature values and prediction outcomes, thus contributing to the understanding of the molecular mechanisms underlying a disease or condition. Furthermore, one should keep in mind that the model cannot perform better than the benchmark comparison test. In practical terms, that means that we have to be absolutely sure in the absolute accuracy of the clinician-assisted diagnosis in order to use it as sensitivity and specificity measure for novel test, which creates a self-perpetuating problem.

From statistics viewpoint, the most important limitation of OMICS-based biomarker discovery is a relatively small number of available samples that impose difficulties in assessing the performance of the model. In ideal world, the training and the validation sets of sample must not overlap. Additionally, both of these sets should be as large as possible. In practice, the size of the sets is limited by the availability of the samples, the factor especially important for relatively rare cancers, and the cost of OMICS profiling per sample. The standard way to solve this problem is to use cross validation, a partitioning of the whole dataset into two parts, where one part is used for model training and another is used to test the trained model. For example, leave-one-out cross validation (LOOCV) involves using a single sample from the original set for the validation of the model obtained by using the remaining samples that comprise the training set. This procedure is repeated until each sample is utilized for the validation. The prediction errors obtained at each run are averaged to estimate the final prediction error of the classification model. Other more powerful methods of cross validation are also available.

Both feature selection and cross validation are vital for building the proper model. The selection of features occurs at each step of the cross validation. Since training set is resampled at each step of cross validation, iterations of this process would yield different feature sets. However, in many studies, the feature selection is performed using the whole dataset, upstream of the cross-validation cycle. This simplified procedure may lead to serious underestimation of the prediction error (Ambroise and McLachlan 2002). Additionally, there are ways to learn from cross-validation procedure that should not be discarded. For example, some samples may be misclassified more often than others; they might be outliers, or genuinely misclassified samples, or other interesting cases that do not fit the typical two-bin output of the model. Studying specific properties of these samples may give additional clues on how to improve the classification model.

The main expected outcome of the OMICS data analysis is the development of multivariate biomarker panels that can be integrated in clinical practice for screening, diagnostic, and prognostic purposes. Hence, final validation of multivariate biomarker panel must be performed on samples that were not previously used during classifier learning or cross validation. In fact, to field test the model, these samples shall be collected independently, in some other medical center, and tested in a different lab. However, the large proportion of multivariate biomarkers developed from OMICS data have not been confirmed in independently collected sets of samples (Gerszten et al. 2008; Sung et al. 2012).

Transcriptome Profiling Approaches and Multiplexed Panels Based on mRNA Levels

Transcriptomics was the first non-DNA-based OMICS. In more than two decades of its development, a plethora of transcriptomics studies were done with a purpose of cancer biomarker discovery. In the early 1990s, when the microarray technique evolved from Southern blotting, first attempts of high-throughput expression profiling were done on colon carcinoma samples (Augenlicht et al. 1991). Since that, much technical advancements were made, but key limitation of expression profiling remained firm – the gene expression profiling methods have to deal with either cell and tissue material. This condition restricts application of expression profiling to biofluids except the assaying of circulating cancer cells. Thus, analysis of expression by microarrays has been mostly used to predict cancer outcome from biopsy tissue specimens (Pusztai et al. 2003) or to post hoc analysis of archived tissue blocks (Waldron et al. 2012). Typically, data from microarrays containing thousands of nucleic acid probes were used to select mRNA candidates with an advent of bioinformatics tools and gimmicks, and the reduced set of candidates was further tested on larger specimen cohort using quantitative real-time PCR or a different platform which was more easy to use than the genome-wide microarray. However, in most cases, the number of profiled specimens was not more than 100, and the resultant predictive models rarely survived validation (Ntzani and Ioannidis 2003).

Nowadays, the transcriptomics drastically changed its technological approach toward RNAseq, the "next-generation" sequencing-based estimation of transcript levels (Ding et al. 2010). In a nutshell, the RNAseq approach provides a global survey of transcriptome activity through en masse generation of short sequence reads from random locations along each of profiled RNAs followed by their mapping to appropriate reference genome. A number of reads that map to a particular gene are proportional to the level of its mRNA level. The RNAseq technique requires lesser amount of RNA than typical microarray (Mutz et al. 2013). However, similar to other high-throughput OMICS analyses, RNAseq results should be validated by qRT-PCR.

Another novel hot field within transcriptomics is an analysis of noncoding RNAs. The results of ENCODE project removed all the doubts about the widespread expression of eukaryotic genomes, with current estimates that more than 62 % of human genome participates in transcription events (Bernstein et al. 2012), in drastic contrast to mere 2 % occupied by protein-coding RNAs. So far, in human genome, ENCODE annotated more that 8,800 small RNAs and 9,600 long noncoding RNA, most of which do not have any attributed function but capable of RNA interference. Abundance of these RNAs, especially small RNAs, stimulated attempts to their potential utility as biomarkers. Recent studies showed that small RNAs, especially microRNAs (miRNAs), remain stable in circulation (Weiland et al. 2012). Vast majority of these circulating miRNA molecules originate from the blood and endothelial cells; however, some tissue-specific miRNAs, for example, from the liver and gut, are represented as well, indicating a broad source of tissue contribution to the total circulating miRNAs (Williams et al. 2013). Hence, miRNA

profile "fingerprints" were suggested as possible biomarkers of developing tumors. A flood of papers and patents about miRNA in cancer increases exponentially and cannot be reviewed in this chapter.

Herein, we will omit gene expression studies as such and further discuss only examples of successful translation of mRNA-based multi-analyte profiles to clinical practice. To evaluate these tests, the US Food and Drug Administration (FDA) designated a novel group identifier IVDMIA (in vitro diagnostic multivariate index assay). In 2004, before IVDMIA, in frame of broader category of laboratory-developed tests (LDTs), the RNA-based prognostic test Oncotype Dx for breast cancer was approved. Later, this test was reclassified as IVDMIA. Other tests already approved through IVDMIA procedure include MammaPrint (2007), Tissue of Origin (2008), as well as Oncotype Dx for colon cancer (2010) and prostate cancer (2013).

Both Oncotype Dx for breast cancer and the MammaPrint determine the risk of breast cancer metastasis, i.e., its relapse after surgery. Based on the assay result, as well as on other clinical features, a medical practitioner would whether assign to the patient a course of adjuvant chemotherapy or not. Both assays quantify the levels of multiple mRNAs in a biopsy sample of excised primary tumor.

MammaPrint assay (Agendia Inc., Netherlands) includes quantitative measurements of expression for 70 genes in mRNA samples extracted from frozen breast carcinoma biopsies. The test intends to estimate the probability of metastatic progression of previously nonmetastatic breast cancer less than 5 cm in size, in women younger than 61. Technically, the assay is based on custom oligonucleotide hybridization microchip (Agilent Technologies) that includes the probes to the set of mRNAs that corresponds to the signature discovered in the microarray study of 112 relatively young breast cancer patients with known outcomes (Van't Veer et al. 2002). The test assigns each patient to the high-risk or low-risk groups. In the first group, chemotherapy would be of benefit. However, some technology problems were identified after the test approval, including 15 % risk overstatement. As a result, the test was retracted from the US markets but remains available in Europe.

Oncotype Dx breast cancer assay (Genomic Health, USA) intends to identify patients with previously diagnosed estrogen receptor (ER)-positive breast carcinoma who should receive adjuvant chemotherapy on top of conventional treatment with tamoxifen. Substantial technical advantage of this assay is its applicability to paraffinized blocks of fixed tissue. Oncotype Dx breast cancer assay is based on 21-gene signature (Dowsett et al. 2010), 16 of which being cancer biomarkers and 5 serve as reference transcripts. This test is based on quantitative real-time PCR. Expression levels for each of these mRNAs are inputted into proprietary algorithm which calculates so-called Recurrence Score (RS), a predictor of chemotherapy benefit that reflects the probabilistic estimate of possible cancer recurrence in 10 years after diagnosis. Recurrence Score is a number between 0 and 100. This value itself does not provide clinically useful information but denotes a risk category for the given patient. RS values less than 11 correspond to a low risk, while RS values between 11 and 25 and more than 25 are recognized as

intermediate and high risk, respectively. Along with other biomarkers, the Recurrence Score guides the decision whether the adjuvant chemotherapy should be administered. Interestingly, low versus high RS patients are significantly more likely to follow the chemotherapy-related recommendation of the test, suggesting a tendency toward less aggressive treatment in high RS that decreases utility of the test (Carlson and Roth 2013).

Both MammaPrint and Oncotype Dx breast cancer assays test for the relative risk of breast carcinoma relapse. Surprisingly, there is only one gene that is common for both signatures, 70-plex in MammaPrint and 21-plex for Oncotype Dx for breast cancer. In one study, both tests were compared side by side using a cohort of 295 patients (Fan et al. 2006). In these settings, the concordance of tests results was at about 80 %. However, 15 of 33 patients classified as intermediate risk by Oncotype Dx were assigned to high-risk group by 70-plex assay. Hence, the intermediate-risk group was the most vulnerable to misclassification. Importantly, OMICS-based molecular tests are especially in demand for this group, as either high- or low-risk patients may be as well ascertained by conventional clinical approach.

A certain contributor to the relative success of the Oncotype Dx is its technological solution to substitute gene expression microarrays by the multiplexed qRT-PCR. Similar platforms are used in Oncotype Dx tests for the colon and prostate cancer. Twelve-plex Oncotype Dx colon cancer assay aims to predict the recurrence for stage II or III of the disease (Venook et al. 2013), while the 17-plex Oncotype Dx prostate cancer assay is developed to provide an opportunity for low-risk patients to avoid invasive treatments such as radical prostatectomy or radiation (Cooperberg et al. 2013). Importantly, only 7 out of 12 colon cancer and 12 out of 17 prostatic carcinoma genes that comprise Oncotype Dx classifiers are target genes; the rest of them are normalization references.

We do not intend to discuss here clinical aspects of the Oncotype Dx test performance, because they are widely described elsewhere (Azim et al. 2013). However, for the biomarker development standpoint, it is important to note that three sets of genes used as signatures for three different cancers do overlap. There are two common transcripts between the colon and the breast tests, one for putative cell cycle protein MKI67 and cell cycle-related transcription factor MYBL2. Additionally, glutathione S-transferase Mu 1 (GSTM1) transcript of breast cancer signature could be paired up with its close homologue and chromosomal locus neighbor GSTM2 in prostate cancer signature. Finally, the colon cancer and the prostate cancer signatures also share one gene, X-chromosome-encoded biglycan (BGN). Moreover, protein-coding genes that comprise Oncotype Dx signatures are closely tied to each other by their functions assessed as protein interaction map (Fig. 3). Note that the hubs of this network are represented by other well-known cancer-associated proteins, such as ERBB2 (HER2) receptor protein kinase that serves as a target for breast cancer drug rituximab and BCL2 proto-oncogene. Oncotype Dx gene sets are also enriched in cluster of neighboring genes co-localized within the same chromosome segments (Table 1). More than one third of the segments share two signature hits. This fact may indicate that the

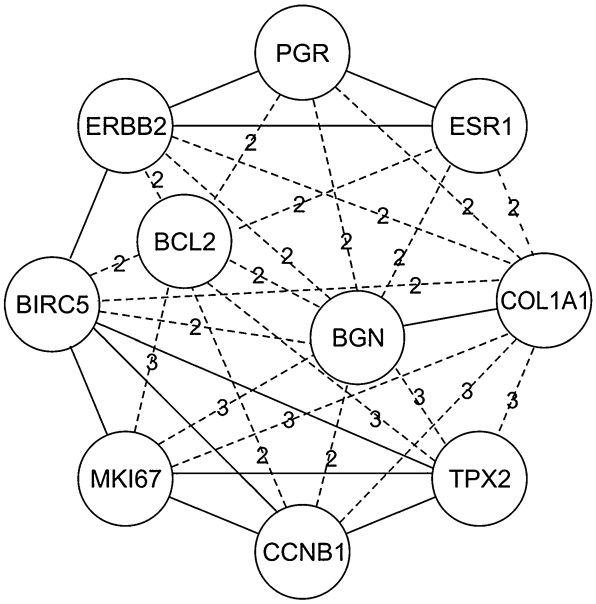

Fig. 3 Functional interactions of Oncomine Dx signature genes. Interaction map of proteins encoded by genes contained in Oncomine Dx breast, colon, and prostate cancer signatures (gene names are shown). The knowledge about protein-protein interactions for bait proteins was received from STRING database (Szklarczyk et al. 2011) with interaction score >0.7 being considered reliable for interaction. Top 10 hub interactors are shown in the network using CytoHubba plug-in (Lin et al. 2008) for Cytoscape software (Shannon et al. 2003). Top hub proteins were specified by Maximal Clique Centrality method. Predicted direct interactions are shown by *solid line*, and indirect connections are shown by *dotted line* with a number of intermediators indicated. Note the well-known cancer-associated proteins, such as ERBB2 (HER2) receptor protein kinase, a target of rituximab drug, and BCL2 proto-oncogene among network hubs

recurrence of human tumors may be associated with specific epigenetic events that require further functional dissection.

The Affymetrix-based transcriptome profiling microchips yielded a practical embodiment in a clinic that is known as a Tissue of Origin test (Pathwork Diagnostics, USA). This test is intended for the determination of tissue type for unassigned metastatic tumors with no lost clear differentiation signs. Tissue of Origin evaluates expression levels for 1,668 genes; these numbers are plugged in the algorithm that assigns the origin of metastatic tumor to one of 15 tissue types. The accuracy of the Tissue of Origin is estimated to be in range of 72–88.5 % for gastric cancer and up to 96.5 % for breast adenocarcinoma. Some modified versions of Tissue of Origin tests are available, such as a 316-plex test for ovary and uterine cancer classification (Lal et al. 2012).

In summary, the main output of gene expression-based signatures is in the field of cancer theranostics, i.e., the personalized management of cancer, rather in its diagnostics.

Table 1 Protein-coding human genes contained in Oncomine Dx breast, colon, and prostate cancer signatures used to determine the recurrence score of these cancers (Cooperberg et al. 2013; Dowsett et al. 2010; Venook et al. 2013) and the genome location of the genes. Gene names are used according to the NextProt knowledge base (www.nextprot.org). Chromosome locations are filled and shown by italic where more than one signature gene is situated. Genes which are contained in more than one cancer signature are highlighted in the same manner

	Type of cancer		
Genome location	Breast (16 genes)	Colon (7 genes)	Prostate (12 genes)
1p13	*GSTM1*	–	*GSTM2*
2p23-24	–	FAP	SRD5A2
5q13	CCNB1	–	–
6q25	ESR1	–	–
7p14	–	INHBA	SFRP4
9p13	BAG1	–	TPM2
9q22	CTSL2	–	–
9q32-33	–	–	ORM1, GSN
10q21	–	–	FAM13C
10q26	*MKI67*	*MKI67*	–
11p15	SCUBE2	–	–
11q22	PGR, MMP11	–	–
15q22	–	DENND4A (C-MYC)	–
17p13	CD68	–	–
17q12	GRB1, ERBB2 (HER2)	–	–
17q21	–	–	COL1A1
17q25	BIRC5	–	–
18q21	BCL2		
19p13	–	GADD45B	KLK2
20q11	–	–	TPX2
20q13	*MYBL2*, AURKA	*MYBL2*	–
Xq28	–	*BGN*	*BGN*

Proteomics: From MALDI-TOF Through Shotgun Techniques to Targeted Approach

The proteomics is technology-driven field that aims at high-throughput inventory of individual proteomes. The basics of protein separation, such as two-dimensional gel electrophoresis, which could visualize whole proteomes or at least their subsets were developed in the 1970s (O'Farrell 1975). At that time, proteome studies were limited by difficulties of identification of proteins within the bands or spots on electrophoretic gels. Availability of genome sequence made possible an identification of proteins by mass spectrometry that deduces the identities of proteins and peptide fragments from mass-to-charge ratios (m/z) being compared to all possible m/z predicted by comparison to the genome parsed into open reading frames (Aebersold 2003).

This process allows one to calculate a probability that theoretically predicted peptide of protein actually exists in the given sample. It is important to note that mass spectrometry provides probabilistic identification of the given protein or peptide but does not sequence this protein de novo. An introduction of the protein sequencing into the routine of the lab would be welcomed, but these techniques are not gaining any traction yet (Hughes et al. 2010).

In mass spectrometry, the accuracy of m/z and, correspondingly, molecular weight quantification is inversely related to the size of analyzed molecule. That is why many proteomics studies start with whole proteome digestion by trypsin and the analysis of resultant mixture. This approach is known as *bottom-up* or *shotgun* proteomics (Washburn et al. 2001). Its main advantage is in ease of separation and identification of short peptides, and its main drawback is the significant loss of information due to the destruction of protein integrity. Conventional pipelines of bottom-up LC-MS/MS proteomics workflow include (i) digestion by isolated protein fraction by trypsin, (ii) separation of resultant peptides by nanoflow HPLC, (iii) electrospray ionization (ESI) of peptides in the flux from HPLC column and tandem mass spectrometry (MS/MS) of peptides and their fragments generated in the mass spectrometer, and (iv) probabilistic identification of peptides from tandem mass spectra by various search algorithms based on genome sequence (Chalkley 2010). Modern bottom-up proteomics pipelines may identify from 1,000 to 10,000 proteins in one sample depending on the workflow and the specimen nature (Zubarev 2013).

Alternative proteomics approach is a "top down," where proteins are analyzed by mass spectrometry in their intact form. Due to their large molecular weights, native protein identification remains far from being routine. The pioneering modifications to tandem mass spectrometry recently demonstrated its power to correctly discern hundreds of proteins in one sample (Tran et al. 2011). However, the complicated, time-consuming procedure of intact proteomics cannot be yet adapted for biomarker discovery. Luckily, in the early 2000s, a relatively simple top-down approach was developed specifically for that purpose, a time-of-flight mass spectrometry with matrix-assisted laser desorption ionization (MALDI-TOF) that could be used for direct profiling of biological fluids, for example, plasma or urine. In this technique, the sample is subjected to fast separation or desalting and applied to the metallic chip (Karpova et al. 2010). In this approach, small proteins and peptides are registered in its intact form but remain unidentified unless downstream experiments are performed.

General experiment design for the MALDI-TOF profiling of the body fluid samples includes the following steps: (i) diseased and control sample preparation with fast separation, (ii) mass-spectra acquisition, (iii) mass-spectra processing to ascertain the intensities for each m/z peak as set of features present in each sample, and (iv) selection of m/z peaks capable of discriminating samples collected from patients with tumors from matched controls and development of multi-peak diagnostic model. Many early papers reported high levels of diagnostic accuracy of MALDI-TOF spectra, some in range of above 90 % (Petricoin et al. 2002). Unfortunately, the MALDI profiles strongly depend on the choice of suitable solvents and