A BEGINNER'S GUIDE TO MICROARRAYS

A BEGINNER'S GUIDE TO MICROARRAYS

edited by

Eric M. Blalock
University of Kentucky Medical Center, U.S.A.

SPRINGER SCIENCE+BUSINESS MEDIA, LLC

Library of Congress Cataloging-in-Publication Data

A beginner's guide to microarrays/edited by Eric M. Blalock
p. ; cm.
Includes bibliographical references and index.
ISBN 978-1-4613-4684-5 ISBN 978-1-4419-8760-0 (eBook)
DOI 10.1007/978-1-4419-8760-0
1. DNA microarrays. I. Blalock, Eric M., 1965-
[DNLM: 1. Oligonucleotide Array Sequence Analysis--methods. 2. Data Interpretation, Statistical. 3. Gene Expression Profiling--methods. 4. Research Design. QZ 52 B417 2003]
QP624.5.D726B456 2003
572.8'636--dc21 2003051408

Originally published by Kluwer Academic Publishers in 2003
Softcover reprint of the hardcover 1st edition 2003

Printed on acid-free paper.

ABOUT THE AUTHORS

CHAPTER 1

Brian Ward is a Principle Investigator at the Life S cience and High Technology Center in St. Louis. He earned his Ph.D. in Chemistry from Michigan State University (1984) studying structure-function relationships of metalloporphyrins. Prior to joining Sigma-Aldrich (1987), he did postdoctoral work studying drug-DNA interactions at Syracuse University (1986).

Kathryn Aboytes is a senior research scientist at the Sigma-Aldrich Life Science and High Technology Center in St. Louis, Missouri. She received the B.A. degree (1989) in microbiology and the M.S. degree (1992) in veterinary microbiology from the University of Missouri, Columbia.

Jason Humphreys is a research scientist at the Sigma-Aldrich Life Science and High Technology Center in St. Louis, Missouri. He received the B.S. degree (1997) in biology from the University of Evansville and the M.S. degree (2000) in biochemistry from the University of Illinois, Champaign-Urbana.

Sonya Reis is an associate research scientist for Sigma-Aldrich at the Life Science and High Technology Center in St. Louis, Missouri. Sonya received her B.S. in biology - medical sciences in 1999 from Southern Illinois University - Edwardsville.

CHAPTER 2

Levente Bodrossy is a research scientist at the Department of Biotechnology, Seibersdorf research, Austria. He is a microbiologist with a primary interest in applying microarrays and related technologies to the fast, parallel detection and identification of microbes. He recieved the M.S. degree (1994) in biology and Ph.D. degree (1997) in biophysics at the University of Szeged, Hungary. Prior to moving to Seibersdorf research he was Senior Lecturer at the Department of Biotechnology, University of Szeged, Hungary. He spent 2 years as visiting scientist at the University of Warwick, UK.

CHAPTER 3

Prior to TeleChem / arrayit.com **Todd Martinsky** served as director of education and consulting at the Codd and Date Consulting Group. Todd worked directly with Dr. E.F. Codd, the father of Relational Database Management Systems and inventor of the Relational Model. Since founding TeleChem, Todd has been instrumental in developing the ArrayIt Brand product line, which includes sample preparation, surface chemistry, hybridization and environmental control products. Todd has led the company to play a significant role in the microarray industry. Along with his daily technical and business direction of the ArrayIt

Brand Product line, Todd has established successful alliances with corporate partners in manufacturing, reagent, equipment and distribution. Todd is responsible for an educational outreach program that ensures that the broadly patented ArrayIt Brand Stealth Micro Spotting Device is applied in the field with optimal scientific and technological results. The Stealth Micro Spotting Device dominates as the most widely used microarray technology in the world.

CHAPTER 4

Robert Searles is the Manager of the Oregon Health and Science University Spotted Microarray Core in Portland, Oregon. He has an undergraduate degree in Biochemistry and Biophysics from Oregon State University and a doctorate from UCLA's Department of Biological Chemistry. He did postdoctoral work in molecular neuroscience. Following his post-doctoral work, he was recruited by the Division of Pathobiology at the Oregon National Primate Research Center to work in viral genomics. He subsequently worked as bioinformaticist for the OHSU West Campus and then established the OHSU cDNA microarray core. He has academic appointments as an Affiliate Assistant Scientist at the OHSU Vaccine and Gene Therapy Institute and as a Staff Scientist at the Oregon National Primate Research Center.

CHAPTER 5

Maureen Sartor is a bioinformatics research associate in the Center for Environmental Genetics at the University of Cincinnati. She received her B.S. degree (1998) in mathematics and her master's degree (2000) in Biomathematics (division of statistics department) from NC State University. After doing consulting work on the analysis of microarrays for Glaxo Wellcome, she now is responsible for statistical analysis of microarrays in her center and advises investigators on experimental design.

Mario Medvedovic, Ph.D. is a Research Assistant Professor in Center for Genome information at University of Cincinnati Medical Center. His research is focused on the use of statistical models in analysis of functional genomics and proteomics data.

Bruce J. Aronow, PhD is an Associate Professor and Director of the HHMI Genome Informatics Core for the University of Cincinnati College of Medicine. He is a member of the Divisions of Pediatric Informatics and Molecular Developmental Biology, the Department of Pediatrics in the College of Medicine, and the Department of Biomedical Engineering at the University of Cincinnati. His research efforts are devoted to the application of comparative and functional genomics methods to an understanding of biological processes and systems.

CHAPTER 6

Eric Blalock is an assistant professor in the Department of Molecular and Biomedical Pharmacology at the University of Kentucky Medical Center. He is a published author in the field of neuroscience, studying the effects of calcium regulation, aging, and epilepsy. For the past several years he has been working closely with the University of Kentucky Department of Statistics and Medical Center Microarray Core facility on the experimental design and statistical analysis of microarray data.

CHAPTER 7

Xuejun Peng is a research assistant in the Medical Center Microarray Core Facility and a PhD candidate in the Department of Statistics at the University of Kentucky. He earned his Medical Diploma (US MD equivalent) in 1994 and obtained his residence training in internal medicine and clinical genetics from 1994 to 1997 from the Hunan Medical University in Changsha, P.R. China. He received an M.S. degree in Statistics in 2000 and has been a Biostatistics consultant and analyst for the past several years.

Arnold J. Stromberg is an associate professor in the Department of Statistics at the University of Kentucky. He is Director of Data Analysis for UK Microarray Core Facility. In addition tohis research interests in bioinformatics, he publishes on robust statistics and control chart methods.

CHAPTER 8

Willy Valdivia-Granda's research involves the development of computational techniques to analyze and model complex biological systems. At North Dakota State University, he is involved in the genomic characterization of plant responses to biotic and abiotic stresses. Willy is also the founder and CEO of Orion Integrated Biosciences, a corporation using micro and nanoarrays for the genomic characterization of organisms of medical, industrial and military relevance. Using DNA microarrays and computational tools, Willy is characterizing the infection mechanisms of malaria (*Plasmodium falciparum*). He is also actively involved in the development of advanced collaborative environments for education. He is the founder of the Virtual Conference on Genomics and Bioinformatics, and Chair and scientific advisor for the 2 002 CHI Microarray Data Analysis International Conferences. In 2002, the Virtual Conference was broadcast without registration fees to more than 2000 researchers in 41 countries, and 33 states within the US, simultaneously using the Access Grid technology and live video streaming.

Contents

Preface

When our laboratory first began using microarrays (~1999) there were precious few books on the subject. In fact, most of our information came from vendors and word of mouth among colleagues. Microarrays have become an ever more integrated component of basic research, and from 1998 to 2001 researchers quadrupled each previous year's publication total. With 2002, a mere doubling of the previous year's scholarly publications indicates that the initial, exponential growth phase of microarray technology may finally be over. Furthermore, basic and clinical research journals now set the bar higher for publication of microarray studies- the Nature family of journals requires MIAME compliance (covered in Chapter 8) prior to publication, and very few top tier journals still accept microarray studies with no replication (Chapters 5, 6 and 7). Thus, microarrays may be moving from a 'hot' buzzword technology into the useful tool that its originators always intended.

In most facilities around the U.S. and in Europe, centralized cores provide a bridge between microarrays and the researchers interested in using them. Within these core systems, certain things have become apparent. First, as a core evolves, its members must explain the technology, its potentials and pitfalls, to the scientists from a broad variety of backgrounds. Second, several very good books covering in-depth aspects of microarray technology and data analysis exist. However, getting user-level information from such thorough treatises sometimes requires more time and effort than the investigator is prepared to invest. Third, there is no primer of microarray technology that covers all of the steps in microarray design and manufacture, as well as experimental design and data analysis.

Our book offers a broad, 'friendly' coverage of many of the most important aspects of microarray technology. We based our coverage on the questions asked of us by new microarray users in universities, laboratories, and microarray list servers. For instance, slide coating is a very important initial step in the preparation of spotted arrays, and has not been addressed thoroughly in any other microarray book. In Chapter 1, authors Kathryn Aboytes, Jason Humphries, Sonya Reis, and Brian Ward remove the black box from this process, and explain how different glass treatments interact with genetic material to form spots. Further, in Chapter 2, Bodrossy Levente provides a detailed 'how-to' guide for oligonucleotide probe design, as well as a powerful yet underused application of microarray technology- the typing and quantification of bacterial contaminants (this often overlooked application may have the most immediate impact on human health and safety of any of the proposed uses of microarrays). Robert Searles dedicates Chapter 3 to setting up and running a microarray core facility, complete with horror stories about here-today, gone-tomorrow vendors, and tried-and-true practices that can get a new facility up and running quickly. In Chapter 4, Todd Martinsky covers the care and use of robotic arrayers and print heads,

and discusses the solutions that work best in spotting. He also includes examples of real world problems and fixes that work. In Chapter 5, Maureen Sartor, Mario Medvedovic, and Bruce Aronow cover data normalization, and clearly describe the ways in which technical error can result in misleading data, as well as how to check and control for its presence. In Chapter 6, I cover approaches for establishing differential gene expression and make recommendations as to the kind of experimental designs that will lead to useful data sets. In addition, I provide detailed, step-by-step procedures for analysis using Excel, and a section, unique to this book, on an automated procedure for identifying not only patterns of expression, but also the likelihood that those patterns would have arisen by chance (based on *post-hoc* statistical analysis). In Chapter 7, Xuejun Peng and Arnold Stromberg address the statistical impact of experimental design, as well as issues of power estimation ('how many chips do I need'), and multiple testing error/ correction. In addition, and unique to this book, proposals for experimental designs that maintain statistical power and reduce experiment cost are discussed. Finally, in Chapter 8, Willy Valdivia Granda gives a thorough overview of the available clustering methodologies along with detailed descriptions of their uses and the software available for these procedures. Further, he provides detailed information about different microarray database structures.

ACKNOWLEDGEMENTS

I thank the chapter authors for their patience, hard work, and dedication to this project, as well as Gretchen Stromberg for her excellent help in document preparation.