

Biostatistical Methods

METHODS IN MOLECULAR BIOLOGY™

John M. Walker, Series Editor

204. **Molecular Cytogenetics: Methods and Protocols**, edited by Yao-Shan Fan, 2002
203. **In Situ Detection of DNA Damage: Methods and Protocols**, edited by Vladimir V. Didenko, 2002
202. **Thyroid Hormone Receptors: Methods and Protocols**, edited by Aria Baniahmad, 2002
201. **Combinatorial Library Methods and Protocols**, edited by Lisa B. English, 2002
200. **DNA Methylation Protocols**, edited by Ken I. Mills and Bernie H. Ramsahoye, 2002
199. **Liposome Methods and Protocols**, edited by Subhash C. Basu and Manju Basu, 2002
198. **Neural Stem Cells: Methods and Protocols**, edited by Tanja Zigova, Juan R. Sanchez-Ramos, and Paul R. Sanberg, 2002
197. **Mitochondrial DNA: Methods and Protocols**, edited by William C. Copeland, 2002
196. **Oxidants and Antioxidants: Ultrastructural and Molecular Biology Protocols**, edited by Donald Armstrong, 2002
195. **Quantitative Trait Loci: Methods and Protocols**, edited by Nicola J. Camp and Angela Cox, 2002
194. **Post-translational Modification Reactions**, edited by Christoph Kannicht, 2002
193. **RT-PCR Protocols**, edited by Joseph O'Connell, 2002
192. **PCR Cloning Protocols, 2nd ed.**, edited by Bing-Yuan Chen and Harry W. Janes, 2002
191. **Telomeres and Telomerase: Methods and Protocols**, edited by John A. Double and Michael J. Thompson, 2002
190. **High Throughput Screening: Methods and Protocols**, edited by William P. Janzen, 2002
189. **GTPase Protocols: The RAS Superfamily**, edited by Edward J. Manser and Thomas Leung, 2002
188. **Epithelial Cell Culture Protocols**, edited by Clare Wise, 2002
187. **PCR Mutation Detection Protocols**, edited by Bimal D. M. Theophilus and Ralph Rapley, 2002
186. **Oxidative Stress and Antioxidant Protocols**, edited by Donald Armstrong, 2002
185. **Embryonic Stem Cells: Methods and Protocols**, edited by Kursad Turksen, 2002
184. **Biostatistical Methods**, edited by Stephen W. Looney, 2002
183. **Green Fluorescent Protein: Applications and Protocols**, edited by Barry W. Hicks, 2002
182. **In Vitro Mutagenesis Protocols, 2nd ed.**, edited by Jeff Braman, 2002
181. **Genomic Imprinting: Methods and Protocols**, edited by Andrew Ward, 2002
180. **Transgenesis Techniques, 2nd ed.: Principles and Protocols**, edited by Alan R. Clarke, 2002
179. **Gene Probes: Principles and Protocols**, edited by Marilena Aquino de Muro and Ralph Rapley, 2002
178. **Antibody Phage Display: Methods and Protocols**, edited by Philippa M. O'Brien and Robert Aitken, 2001
177. **Two-Hybrid Systems: Methods and Protocols**, edited by Paul N. MacDonald, 2001
176. **Steroid Receptor Methods: Protocols and Assays**, edited by Benjamin A. Lieberman, 2001
175. **Genomics Protocols**, edited by Michael P. Starkey and Ramnath Elaswarapu, 2001
174. **Epstein-Barr Virus Protocols**, edited by Joanna B. Wilson and Gerhard H. W. May, 2001
173. **Calcium-Binding Protein Protocols, Volume 2: Methods and Techniques**, edited by Hans J. Vogel, 2001
172. **Calcium-Binding Protein Protocols, Volume 1: Reviews and Case Histories**, edited by Hans J. Vogel, 2001
171. **Proteoglycan Protocols**, edited by Renato V. Iozzo, 2001
170. **DNA Arrays: Methods and Protocols**, edited by Jang B. Rampil, 2001
169. **Neurotrophin Protocols**, edited by Robert A. Rush, 2001
168. **Protein Structure, Stability, and Folding**, edited by Kenneth P. Murphy, 2001
167. **DNA Sequencing Protocols, Second Edition**, edited by Colin A. Graham and Alison J. M. Hill, 2001
166. **Immunotoxin Methods and Protocols**, edited by Walter A. Hall, 2001
165. **SV40 Protocols**, edited by Leda Raptis, 2001
164. **Kinesin Protocols**, edited by Isabelle Vernos, 2001
163. **Capillary Electrophoresis of Nucleic Acids, Volume 2: Practical Applications of Capillary Electrophoresis**, edited by Keith R. Mitchelson and Jing Cheng, 2001
162. **Capillary Electrophoresis of Nucleic Acids, Volume 1: Introduction to the Capillary Electrophoresis of Nucleic Acids**, edited by Keith R. Mitchelson and Jing Cheng, 2001
161. **Cytoskeleton Methods and Protocols**, edited by Ray H. Gavin, 2001
160. **Nuclease Methods and Protocols**, edited by Catherine H. Schein, 2001
159. **Amino Acid Analysis Protocols**, edited by Catherine Cooper, Nicole Packer, and Keith Williams, 2001
158. **Gene Knockout Protocols**, edited by Martin J. Tymms and Ismail Kola, 2001
157. **Mycotoxin Protocols**, edited by Mary W. Trucksess and Albert E. Pohland, 2001
156. **Antigen Processing and Presentation Protocols**, edited by Joyce C. Solheim, 2001
155. **Adipose Tissue Protocols**, edited by Gérard Ailhaud, 2000
154. **Connexin Methods and Protocols**, edited by Roberto Bruzzone and Christian Giaume, 2001
153. **Neuropeptide Y Protocols**, edited by Ambikaipakan Balasubramaniam, 2000
152. **DNA Repair Protocols: Prokaryotic Systems**, edited by Patrick Vaughan, 2000
151. **Matrix Metalloproteinase Protocols**, edited by Ian M. Clark, 2001
150. **Complement Methods and Protocols**, edited by B. Paul Morgan, 2000
149. **The ELISA Guidebook**, edited by John R. Crowther, 2000
148. **DNA-Protein Interactions: Principles and Protocols (2nd ed.)**, edited by Tom Moss, 2001
147. **Affinity Chromatography: Methods and Protocols**, edited by Pascal Bailon, George K. Ehrlich, Wen-Jian Fung, and Wolfgang Berthold, 2000
146. **Mass Spectrometry of Proteins and Peptides**, edited by John R. Chapman, 2000

METHODS IN MOLECULAR BIOLOGY™

Biostatistical Methods

Edited by

Stephen W. Looney

*University of Louisville School of Medicine,
Louisville, Kentucky*

Humana Press



Totowa, New Jersey

© 2002 Humana Press Inc.
999 Riverview Drive, Suite 208
Totowa, New Jersey 07512

humanapress.com

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording, or otherwise without written permission from the Publisher. *Methods in Molecular Biology*TM is a trademark of The Humana Press Inc.

The content and opinions expressed in this book are the sole work of the authors and editors, who have warranted due diligence in the creation and issuance of their work. The publisher, editors, and authors are not responsible for errors or omissions or for any consequences arising from the information or opinions presented in this book and make no warranty, express or implied, with respect to its contents.

This publication is printed on acid-free paper. ∞

ANSI Z39.48-1984 (American Standards Institute) Permanence of Paper for Printed Library Materials.

Production Editor: Diana Mezzina

Cover illustration: Figure 14 from Chapter 4, "Statistical Methods for Proteomics," by Françoise Seillier-Moiseiwitsch, Donald C. Trost, and Julian Moiseiwitsch.

Cover design by Patricia F. Cleary.

For additional copies, pricing for bulk purchases, and/or information about other Humana titles, contact Humana at the above address or at any of the following numbers: Tel.: 973-256-1699; Fax: 973-256-8341; E-mail: humana@humanapr.com; or visit our Website: www.humanapress.com

Photocopy Authorization Policy:

Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by Humana Press Inc., provided that the base fee of US \$10.00 per copy, plus US \$00.25 per page, is paid directly to the Copyright Clearance Center at 222 Rosewood Drive, Danvers, MA 01923. For those organizations that have been granted a photocopy license from the CCC, a separate system of payment has been arranged and is acceptable to Humana Press Inc. The fee code for users of the Transactional Reporting Service is: [0-89603-951-X/02 \$10.00 + \$00.25].

Printed in the United States of America. 10 9 8 7 6 5 4 3 2 1

Library of Congress Cataloging in Publication Data

Biostatistical Methods / edited by Stephen W. Looney.

p. cm. — (Methods in molecular biology ; v. 184)

Includes bibliographical references and index.

ISBN 0-89603-951-X (alk. paper)

1. Biometry. 2. Molecular biology. I. Looney, Stephen W. II. *Methods in molecular biology* (Totowa, N.J.); v. 184.

QH323.5 .B5628 2002
570'.1'5195--dc21

2001026440

To my mother and the memory of my father

Preface

Biostatistical applications in molecular biology have increased tremendously in recent years. For example, a search of the Current Index to Statistics indicates that there were 62 articles published during 1995–1999 that had “marker” in the title of the article or as a keyword. In contrast, there were 29 such articles in 1990–1994, 17 in 1980–1989, and only 5 in 1970–1979. As the number of publications has increased, so has the sophistication of the statistical methods that have been applied in this area of research.

In *Biostatistical Methods*, we have attempted to provide a representative sample of applications of biostatistics to commonly occurring problems in molecular biology, broadly defined. It has been our intent to provide sufficient background information and detail that readers might carry out similar analyses themselves, given sufficient experience in both biostatistics and the basic sciences. Not every chapter could be written at an introductory level, since, by their nature, many statistical methods presented in this book are at a more advanced level and require knowledge and experience beyond an introductory course in statistics. Similarly, the proper application of many of these statistical methods to problems in molecular biology also requires that the statistical analyst have extensive knowledge about the particular area of scientific inquiry. Nevertheless, we feel that these chapters at least provide a good starting point, both for statisticians who want to begin work on problems in molecular biology, and for molecular biologists who want to increase their working knowledge of biostatistics as it relates to their field.

The chapters in this volume cover a wide variety of topics, both in terms of biostatistics and in terms of molecular biology. The first two chapters are very general in nature: In Chapter 1, Emmanuel Lazaridis and Gregory Bloom provide an historical overview of developments in molecular biology, computational biology, and statistical genetics, and describe how biostatistics has contributed to developments in these areas. In Chapter 2, Gregory Bloom and his colleagues describe a new paradigm linking image quantitation and data analysis that should provide valuable insight to anyone working in image-based biological experimentation.

The remaining chapters in *Biostatistical Methods* are arranged in approximately the order in which the corresponding topic or methods of analysis would

be utilized in developing a new marker for exposure to a risk factor or for a disease outcome. The development of such a marker would most likely begin with an examination of the genetic basis for one or more phenotypes. Chapters 3 and 4 deal with two of the most fundamental aspects of research in this area: microarray analysis, which deals with gene expression, and proteomics, which deals with the identification and quantitation of gene products, namely, proteins. Research in either or both of these areas could produce a biomarker candidate that would then be scrutinized for its clinical utility.

Chapters 5 and 6 deal with issues that arise very early in studies attempting to link the results of experimentation in molecular biology with exposure or disease in human populations. In Chapter 5, I discuss many of the issues associated with determining whether a new biomarker will be suitable for studying a particular E-D association. Jane Goldsmith, in Chapter 6, discusses the importance of designing studies with sufficient numbers of subjects in order to attain adequate levels of statistical power.

Chapters 7 and 8 are concerned with genetic effects as they relate to human populations. In Chapter 7, Peter Jones and his colleagues describe statistical models that have proven useful in studying the associations between disease and the inheritance of particular genetic variants. In Chapter 8, Stan Young and his colleagues describe sophisticated statistical methods that can be used to control the overall false-positive rate of the perhaps thousands of statistical tests that might be performed when attempting to link the presence or absence of particular alleles to the occurrence of disease.

Jim Dignam and his colleagues, in Chapter 9, describe the statistical issues that one should consider when evaluating the clinical utility of molecular characteristics of tumors, as they relate to cancer prognosis and treatment efficacy. Finally, in Chapter 10, Greg Rempala and I describe methods that might be used to validate statistical methods that have been developed for analyzing the E-D association in specific situations, such as when the exposure has been characterized poorly.

I would like to express my sincere appreciation to the reviewers of the various chapters in this volume: Rich Evans of Iowa State University, Mario Cleves of the University of Arkansas for Medical Sciences, Stephen George of the Duke University Medical Center, Ralph O'Brien of the Cleveland Clinic Foundation, and Martin Weinrich of the University of Louisville School of Medicine. I am also indebted to John Walker, Series Editor for *Methods in Molecular Biology*, and to Thomas Lanigan, President, Craig Adams, Developmental Editor, Diana Mezzina, Production Editor, and Mary Jo Casey, Manager, Composition Services, Humana Press.

Stephen W. Looney

Contents

Dedication	v
Preface	vii
Contributors	xi
1 Statistical Contributions to Molecular Biology <i>Emmanuel N. Lazaridis and Gregory C. Bloom</i>	1
2 Linking Image Quantitation and Data Analysis <i>Gregory C. Bloom, Peter Gieser, and Emmanuel N. Lazaridis</i>	15
3 Introduction to Microarray Experimentation and Analysis <i>Peter Gieser, Gregory C. Bloom, and Emmanuel N. Lazaridis</i>	29
4 Statistical Methods for Proteomics <i>Françoise Seillier-Moiseiwitsch, Donald C. Trost, and Julian Moiseiwitsch</i>	51
5 Statistical Methods for Assessing Biomarkers <i>Stephen W. Looney</i>	81
6 Power and Sample Size Considerations in Molecular Biology <i>L. Jane Goldsmith</i>	111
7 Models for Determining Genetic Susceptibility and Predicting Outcome <i>Peter W. Jones, Richard C. Strange, Sud Ramachandran, and Anthony Fryer</i>	131
8 Multiple Tests for Genetic Effects in Association Studies <i>Peter H. Westfall, Dmitri V. Zaykin, and S. Stanley Young</i>	143
9 Statistical Considerations in Assessing Molecular Markers for Cancer Prognosis and Treatment Efficacy <i>James Dignam, John Bryant, and Soonmyung Paik</i>	169
10 Power of the Rank Test for Multi-Strata Case-Control Studies with Ordinal Exposure Variables <i>Grzegorz A. Rempala and Stephen W. Looney</i>	191
Index	203

Contributors

- GREGORY C. BLOOM • *H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL*
- JOHN BRYANT • *Department of Biostatistics, University of Pittsburgh and Biostatistical Center, National Surgical Adjuvant Breast and Bowel Project, Pittsburgh, PA*
- JAMES DIGNAM • *Department of Biostatistics, University of Chicago, IL, and Biostatistical Center, National Surgical Adjuvant Breast and Bowel Project, Pittsburgh, PA*
- ANTHONY. FRYER • *Center For Cell and Molecular Medicine, Keele University, North Staffordshire Hospital NHS Trust, Stoke-on-Trent, Staffordshire, UK*
- PETER GIESER • *Net Impact, Inc., Temecula, CA*
- L. JANE GOLDSMITH • *Department of Family and Community Medicine, University of Louisville School of Medicine, Louisville, KY*
- PETER W. JONES • *Department of Mathematics, Keele University, Keele, Staffordshire, UK*
- EMMANUEL N. LAZARIDIS • *H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL*
- STEPHEN W. LOONEY • *Department of Family and Community Medicine, University of Louisville School of Medicine, Louisville, KY*
- JULIAN MOISEIWITSCH • *Department of Endodontics, School of Dentistry, University of Maryland, Baltimore, MD*
- SOONMYUNG PAIK • *Division of Pathology, National Surgical Adjuvant Breast and Bowel Project, Allegheny General Hospital, Pittsburgh, PA*
- SUD RAMACHANDRAN • *Department of Clinical Biochemistry, Good Hope Hospital, Sutton Coldfield, West Midlands, UK*
- GRZEGORZ A. REMPALA • *Department of Mathematics, University of Louisville, Louisville, KY*
- FRANÇOISE SEILLIER-MOISEIWITSCH • *Department of Mathematics and Statistics, University of Maryland, Baltimore Country, Baltimore, MD*
- RICHARD C. STRANGE • *Center for Cell and Molecular Medicine, Keele University, North Staffordshire Hospital NHS Trust, Stoke-on-Trent, Staffordshire, UK*

DONALD C. TROST • *Signal Processing, Clinical Technology, Pfizer Global Research and Development, Groton, CT*

PETER H. WESTFALL • *Area of ISQS, Texas Tech University, Lubbock, TX*

S. STANLEY YOUNG • *GlaxoSmithKline, Research Triangle Park, NC*

DMITRI V. ZAYKIN • *GlaxoSmithKline, Research Triangle Park, NC*