METHODS IN MOLECULAR BIOLOGY

Series Editor
John M. Walker
School of Life and Medical Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

For further volumes:
http://www.springer.com/series/7651

Statistical Human Genetics

Methods and Protocols

Second Edition

Edited by

Robert C. Elston

Case Western Reserve University, Cleveland, Ohio, USA

Humana Press

Editor
Robert C. Elston
Case Western Reserve University
Cleveland, Ohio, USA

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-7273-9 ISBN 978-1-4939-7274-6 (eBook)
DOI 10.1007/978-1-4939-7274-6

Library of Congress Control Number: 2017951222

Printed on acid-free paper

This Humana Press imprint is published by Springer Nature
The registered company is Springer Science+Business Media LLC
The registered company address is: 233 Spring Street, New York, NY 10013, U.S.A.

Preface

The recent advances in genetics, especially in the molecular techniques that have over the last quarter of a century spectacularly reduced the cost of determining genetic markers, open up a field of research that is becoming of increasing help in detecting, preventing, and/or curing many diseases that afflict us. This has brought with it the need for novel methods of statistical analysis and the implementation of these methods in a wide variety of computer programs. The aim in this book is to make these methods and programs more easily accessible to the beginner who has data to analyze, whether a student or a senior investigator. Twenty-eight of the thirty original chapters have been retained (all updated) and, in view of changes in the field, three new chapters have been added. Apart from the first chapter, which defines some of the genetic terms used in the book, each chapter of this book takes up a particular topic and illustrates the use of at least one piece of software that the authors have found helpful for the relevant statistical analysis of their own human genetic data. There is often more than one program that performs a particular type of analysis and, once you have used one program for a particular analysis, you may find you prefer another program—and there is a good chance you will find that the same basic analysis is described in more than one chapter of this book. You may therefore wish to browse over several chapters, in the first place restricting your reading to only the introductory sections, which describe the underlying theory. The chapters are ordered in the approximate logical order in which human genetic studies are often conducted; so, if you are new to research in human genetics, this initial reading could serve as an introduction to the subject. For the most part, the second section of each chapter gives you step-by-step instructions for running programs and interpreting the program outputs, with extra notes in the third section. However, although the aim is very much to offer a "do it yourself" manual, there may well be times when you will need to consult a statistical geneticist, especially for the interpretation of computer output.

Cleveland, Ohio, USA *Robert C. Elston*

Contents

Contributors

Indra Adrianto • *Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA*

Terri H. Beaty • *Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA*

Gurkan Bebek • *Systems Biology and Bioinformatics Graduate Program, Case Western Reserve University School of Medicine, Cleveland, OH, USA; Center for Proteomics and Bioinformatics, Case Western Reserve University School of Medicine, Cleveland, OH, USA; Department of Nutrition, Case Western Reserve University School of Medicine, Cleveland, OH, USA; Department of Electrical Engineering and Computer Science, Case Western Reserve University School of Medicine, Cleveland, OH, USA*

D. Timothy Bishop • *Section of Epidemiology and Biostatistics, Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, UK*

Murielle Bochud • *Institute of Social and Preventive Medicine, Lausanne University Hospital, Lausanne, Switzerland*

Shelley B. Bull • *Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, ON, Canada; Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada*

Rita M. Cantor • *Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA; Center for Neurobehavioral Genetics, Department of Psychiatry, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA*

Wei-Min Chen • *Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA; Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA*

Ren-Hua Chung • *Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Miaoli County, Taiwan*

Fabiola Del Greco M. • *Center for Biomedicine, EURAC Research, Bolzano, Italy*

Apostolos Dimitromanolakis • *Department of Statistical Sciences, Faculty of Arts and Science, Toronto, ON, Canada; Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON, Canada*

Robert C. Elston • *Case Western Reserve University, Cleveland, OH, USA*

Warren Ewens • *Department of Biology, University of Pennsylvania, Philadelphia, PA, USA*

Joseph L. Gastwirth • *Department of Statistics, George Washington University, Washington, DC, USA*

Damian Gola • *Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Universitätsklinikum Schleswig-Holstein - Campus Lübeck, Lübeck, Germany*

Katrina L. Grasby • *Genetic Epidemiolgy, QIMR Berghofer Medical Research Institute, Brisbane, Australia*

Celia M.T. Greenwood • *Lady Davis Research Institute, Jewish General Hospital, Montréal, QC, Canada; Department of Oncology and Department of Epidemiology, Biostatistics & Occupational Health, McGill University, QC, Canada*

Sandeep Grover • *Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Universitätsklinikum Schleswig-Holstein, Lübeck, Germany*

Noémi B. Hall • *Department of Population and Quantitative Health Sciences, Case Western Reserve University School of Medicine, Cleveland, OH, USA*
Daniel D. Kinnamon • *Division of Human Genetics, Department of Internal Medicine, The Ohio State University Wexner Medical Center, Columbus, OH, USA*
Qizhai Li • *Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China*
Xiaoyin Li • *Department of Population and Quantitative Health Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH, USA*
Xin Li • *Department of Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, OH, USA*
Jing Li • *Department of Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, OH, USA*
Jin Ma • *Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, ON, Canada*
Eden R. Martin • *John P. Hussman Institute for Human Genomics, Leonard M. Miller School of Medicine, University of Miami, Miami, FL, USA*
Sarah E. Medland • *Genetic Epidemiolgy, QIMR Berghofer Medical Research Institute, Brisbane, Australia*
Courtney Montgomery • *Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA*
Nathan J. Morris • *Department of Population and Quantitative Health Sciences, Case Western Reserve University School of Medicine, Cleveland, OH, USA*
Miriam A. Mosing • *Department of Neuroscience, Karolinska Institute, Stockholm, Sweden; Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden*
Adam C. Naj • *Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA; Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA; Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA*
Junghyun Namkung • *Molecular Diagnostics Team, IVD Business Unit, SK Telecom, Seoul, South Korea*
Nora L. Nock • *Department of Population and Quantitative Health Sciences, Case Western Reserve University School of Medicine, Cleveland, OH, USA*
Jérémie Nsengimana • *Section of Epidemiology and Biostatistics, Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, UK*
Jeffery O'Connell • *University of Maryland, Baltimore, MD, USA*
Weiwei Ouyang • *Department of Global Biostatistics and Data Science, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA*
Andrew D. Paterson • *Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada; Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada*
Huaizhen Qin • *Department of Global Biostatistics and Data Science, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA; Department of Population and Quantitative Health Sciences, Case Western Reserve University School of Medicine, Cleveland, OH, USA*
Cielito Reyes-Gibby • *Department of Emergency Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA*

Jaya Satagopan • *Memorial Sloan Kettering Cancer Center, New York, NY, USA*
Audrey H. Schnell • *Cardiovascular Imaging Core Laboratory, Harrington Heart & Vascular Institute, University Hospitals Cleveland Medical Center, Cleveland, OH, USA*
Sanjay Shete • *Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA*
Yeunjoo E. Song • *Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA*
Sunah Song • *Case Western Reserve University, Cleveland, OH, USA*
Catherine M. Stein • *Center for Proteomics and Bioinformatics, Case Western Reserve University, Cleveland, OH, USA*
Daniel O. Stram • *Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA*
Shuying Sun • *Texas State University, San Marcos, TX, USA*
Lei Sun • *Department of Statistical Sciences, Faculty of Arts and Science, University of Toronto, Toronto, ON, Canada; Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada*
Xiangqing Sun • *Department of Population and Quantitative Health Sciences, Case Western Reserve University School of Medicine, Cleveland, OH, USA*
Maren Vens • *Institut für Medizinische Biometrie und Epidemiologie, Universitätsklinikum Hamburg-Eppendorf, Hamburg, Germany*
Karin J.H. Verweij • *Genetic Epidemiolgy, QIMR Berghofer Medical Research Institute, Brisbane, Australia; Behavioural Science Institute, Radboud University, HR Nijmegen, The Netherlands*
Jian Wang • *Department of Biostatistics, University of Texas MD Anderson Cancer Center, Houston, TX, USA*
Heming Wang • *Department of Population and Quantitative Health Sciences, Case Western Reserve University School of Medicine, Cleveland, OH, USA; Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA, USA; Division of Sleep Medicine, Harvard Medical School, Boston, MA, USA*
Danica Wiredja • *Systems Biology and Bioinformatics Graduate Program, Case Western Reserve University School of Medicine, Cleveland, OH, USA; Center for Proteomics and Bioinformatics, Case Western Reserve University School of Medicine, Cleveland, OH, USA; Department of Nutrition, Case Western Reserve University School of Medicine, Cleveland, OH, USA*
Sungho Won • *Department of Public Health Science, Graduate School of Public Health, Seoul National University, Seoul, South Korea*
Marvin N. Wright • *Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Universitätsklinikum Schleswig-Holstein - Campus Lübeck, Lübeck, Germany*
Wei Xu • *Department of Biostatistics, Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada; Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada*
Yin Yao • *Unit of Genomic Statistics, Intramural Research Program, National Institute of Mental Health, Bethesda, MD, USA*
Ao Yuan • *Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University, Washington, DC, USA*
Xiaofeng Zhu • *Department of Population and Quantitative Health Sciences, Case Western Reserve University School of Medicine, Cleveland, OH, USA*

Xuan Zhu • *Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA*

Andreas Ziegler • *Institut für Medizinische Biometrie und Statistik & Zentrum für klinische Studien, Universität zu Lübeck, Universitätsklinikum Schleswig-Holstein, Lübeck, Germany*

Brendan P. Zietsch • *School of Psychology, University of Queensland, Brisbane, Australia*