METHODS IN MOLECULAR BIOLOGY

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

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

Statistical Genomics

Methods and Protocols

Edited by

Ewy Mathé

Biomedical Informatics, College of Medicine, Ohio State University, Columbus, OH, USA

Sean Davis

National Institutes of Health, National Cancer Institute, Bethesda, MD, USA

Humana Press

Editors
Ewy Mathé
Biomedical Informatics, College of Medicine
Ohio State University
Columbus, OH, USA

Sean Davis
National Institutes of Health
National Cancer Institute
Bethesda, MD, USA

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-3576-5 ISBN 978-1-4939-3578-9 (eBook)
DOI 10.1007/978-1-4939-3578-9

Library of Congress Control Number: 2016933669

© Springer Science+Business Media New York 2016
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Humana Press imprint is published by Springer Nature
The registered company is Springer Science+Business Media LLC New York

Preface

Statistical Analysis of Genomic Data is, indeed, a very broad topic. We have attempted in this volume to provide chapters with cross-cutting groundwork materials, public data repositories, common applications of statistical analysis in genomics, and some representative toolsets for operating on genomic data. While we cannot be comprehensive in a single volume, we have tried to provide a breadth of both applications and tools. The authors of the individual chapters have largely focused on practical aspects of their topics, as we feel that application is an integral part of learning about statistical analysis of genomic data.

More specifically, the volume is divided into four parts. In the first part, we have included overview material and resources that can be applied across topics later in the book. In the second part, a couple of prominent public repositories for genomic data are covered in some depth. In the third part, several different biological applications of statistical genomics are presented. In the fourth and last part, software tools that can be used to facilitate ad hoc analysis and data integration are highlighted. Finally, we thank the chapter authors for the generosity of their time and insight in preparing their excellent contributions.

Columbus, OH, USA — *Ewy Mathé*
Bethesda, MD, USA — *Sean Davis*

Contents

Part IV Tools

Contributors

Naomi S. Altman • *Department of Statistics and Huck Institutes of Life Sciences, The Pennsylvania State University, PA, USA*

Sonali Arora • *Computational Biology Program, Fred Hutchinson Cancer Research Center, Seattle, WA, USA*

Songjoon Baek • *Laboratory of Receptor Biology and Gene Expression, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA*

Tanya Barrett • *National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA*

Gabriel Becker • *Genentech, South San Francisco, CA, USA*

Matthias Blum • *Equipe Labellisée Ligue Contre le Cancer, Department of Functional Genomics and Cancer, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC)/CNRS/INSERM/Université de Strasbourg, Illkirch Cedex, France*

Matthew Brauer • *Genentech, South San Francisco, CA, USA*

Marc R.J. Carlson • *Computational Biology Program, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; Seattle Children's Research Institute, Seattle, WA, USA*

Mark A. Carty • *Institute for Computational Biomedicine, Weill Cornell Medical College, New York, NY, USA; Memorial Sloan Kettering Cancer Center, New York, NY, USA*

Benilton S. Carvalho • *Brazilian Institute of Neuroscience and Neurotechnology (BRAINN) and Department of Statistics, University of Campinas, Campinas, São Paulo, Brazil*

Lukas Chavez • *German Cancer Research Center, Heidelberg, Germany*

Yunshun Chen • *Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia*

Pierre-Etienne Cholley • *Equipe Labellisée Ligue Contre le Cancer, Department of Functional Genomics and Cancer, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC)/CNRS/INSERM/Université de Strasbourg, Illkirch Cedex, France*

Emily Clough • *National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA*

Aedin Culhane • *Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA, USA; Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA*

Olivier Elemento • *Institute for Computational Biomedicine, Weill Cornell Medical College, New York, NY, USA*

Eibe Frank • *Department of Computer Science, University of Waikato, Hamilton, New Zealand*

Hinrich Gronemeyer • *Equipe Labellisée Ligue Contre le Cancer, Department of Functional Genomics and Cancer, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC)/CNRS/INSERM/Université de Strasbourg, Illkirch Cedex, France*

Florian Hahne • *Novartis Institute for Biomedical Research, Basel, Switzerland*

Nancy F. Hansen • *National Human Genome Research Institute, Rockville, MD, USA*

Loren A. Honaas • *Department of Biology, The Pennsylvania State University, Wenatchee, WA, USA*

Robert Ivanek • *Department of Biomedicine, University of Basel, Basel, Switzerland*
Mark A. Jensen • *Research Administration Directorate, Leidos Biomedical Research Inc., Frederick National Laboratory for Cancer Research, Frederick, MD, USA*
Martin Krzywinski • *Canada's Michael Smith Genome Sciences Centre, Vancouver, BC, Canada*
M. Scott Kuehn • *Opower Inc., San Francisco, CA, USA*
Michael Lawrence • *Genentech, South San Francisco, CA, USA*
Matthias Lienhard • *Max Planck Institute for Molecular Genetics, Berlin, Germany*
Aaron T. L. Lun • *Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia*
Marco Antonio Mendoza-Parra • *Equipe Labellisée Ligue Contre le Cancer, Department of Functional Genomics and Cancer, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC)/CNRS/INSERM/Université de Strasbourg, Illkirch Cedex, France*
Chen Meng • *Chair of Proteomics and Bioanalytics, Technische Universität Mnchen, Freising, Germany*
Martin Morgan • *Computational Biology Program, Fred Hutchinson Cancer Research Center, Seattle, WA, USA*
Shane Neph • *Department of Genome Sciences, Altius Institute for Biomedical Sciences, Seattle, WA, USA*
Valerie Obenchain • *Computational Biology Program, Fred Hutchinson Cancer Research Center, Seattle, WA, USA*
Hervé Pagès • *Computational Biology Program, Fred Hutchinson Cancer Research Center, Seattle, WA, USA*
Jens Reeder • *Genentech, South San Francisco, CA, USA*
Alex P. Reynolds • *Department of Genome Sciences, Altius Institute for Biomedical Sciences, Seattle, WA, USA*
Markus Riester • *Novartis Institutes for BioMedical Research (NIBR), Cambridge, MA, USA*
Mohamed-Ashick M. Saleem • *Equipe Labellisée Ligue Contre le Cancer, Department of Functional Genomics and Cancer, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC)/CNRS/INSERM/Université de Strasbourg, Illkirch Cedex, France*
Tony C. Smith • *Department of Computer Science, University of Waikato, Hamilton, New Zealand*
Gordon K. Smyth • *Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia*
John A. Stamatoyannopoulos • *Altius Institute for Biomedical Sciences, Seattle, WA, USA; Department of Medicine, University of Washington, Seattle, WA, USA; Department of Genome Sciences, University of Washington, Seattle, WA, USA*
Myong-Hee Sung • *Laboratory of Receptor Biology and Gene Expression, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA; Laboratory of Molecular Biology and Immunology, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA*
Levi Waldron • *Department of Epidemiology and Biostatistics, City University of New York, School of Public Health, New York, NY, USA*
Zhining Wang • *Center for Cancer Genomics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA*

Thomas D. Wu • *Genentech, South San Francisco, CA, USA*
Zhijin Wu • *Department of Biostatistics, Brown University, Providence, RI, USA*
Hao Wu • *Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA, USA*
Jean Claude Zenklusen • *Center for Cancer Genomics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA*
Hongen Zhang • *Center for Cancer Research, National Institutes of Health, National Cancer Institute, Bethesda, MD, USA*