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

For over 35 years, biological scientists have come to rely on the research protocols and methodologies in the critically acclaimed *Methods in Molecular Biology* series. The series was the first to introduce the step-by-step protocols approach that has become the standard in all biomedical protocol publishing. Each protocol is provided in readily-reproducible step-by-step fashion, opening with an introductory overview, a list of the materials and reagents needed to complete the experiment, and followed by a detailed procedure that is supported with a helpful notes section offering tips and tricks of the trade as well as troubleshooting advice. These hallmark features were introduced by series editor Dr. John Walker and constitute the key ingredient in each and every volume of the *Methods in Molecular Biology* series. Tested and trusted, comprehensive and reliable, all protocols from the series are indexed in PubMed.

Computational Epigenomics and Epitranscriptomics

Edited by

Pedro H. Oliveira

Genoscope, Évry, France

Humana Press

Editor
Pedro H. Oliveira
Genoscope
Évry, France

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-0716-2961-1 ISBN 978-1-0716-2962-8 (eBook)
https://doi.org/10.1007/978-1-0716-2962-8

This Humana imprint is published by the registered company Springer Science+Business Media, LLC, part of Springer Nature.
The registered company address is: 1 New York Plaza, New York, NY 10004, U.S.A.

Preface

Nucleic acids (DNA and RNA) are key repositories of genetic information, and their primary sequence of four canonical nucleobases (A, C, G, T/U) in genomes and transcriptomes defines the genetic blueprints and cellular identities across all branches of life. Moreover, it is recognized that diversity within an organism is often governed by dynamic chemical modifications of nucleobases, which can operate as a regulatory layer to fine-tune key molecular and cellular processes. Changes in epigenomic and epitranscriptomic landscapes can affect a variety of such processes (e.g., transcription, translation, differentiation, and maintenance of genome integrity) and are often linked to the onset and progression of disease. Our understanding of the biochemistry and biological significance of the more than 45 DNA and 170 RNA chemical modifications reported to date [1, 2] has been largely propelled by high-throughput sequencing technologies and mass-spectrometry-based approaches, coupled with chemical, enzymatic, or antibody-dependent methodologies. In parallel, we have witnessed the development of increasingly robust computational methods and statistical tools tailored to make sense of a growing volume of often heterogeneous and noisy epi-ome data. In this book, the reader is introduced to state-of-the-art computational methods designed to manage, analyze, and generally leverage epigenomic and epitranscriptomic data. Topics include fine-mapping and quantification of modifications, visual analytics, imputation methods, supervised analysis, and integrative approaches for single-cell data. Ultimately this compendium will be of interest to a broad audience including students, biologists, bioinformaticians, and biomedical researchers.

Évry, France — ***Pedro H. Oliveira***

References

1. Sood et al (2019) DNAmod: the DNA modification database. J Cheminfor 11(30): 1–10
2. Boccaletto et al (2018) MODOMICS: a database of RNA modification pathways. 2017 update. Nucleic Acids Res 46(Database issue):D303–D307

Contents

Preface . . . . . *v*
Contributors . . . . . *ix*

1 DNA Methylation Data Analysis Using Msuite 1
Xiaojian Liu, Pengxiang Yuan, and Kun Sun

2 Interactive DNA Methylation Array Analysis with ShinyÉPICo 7
Octavio Morante-Palacios

3 Predicting Chromatin Interactions from DNA Sequence Using DeepC 19
Ron Schwessinger

4 Integrating Single-Cell Methylome and Transcriptome Data with MAPLE 43
Yasin Uzun, Hao Wu, and Kai Tan

5 Quantitative Comparison of Multiple Chromatin Immunoprecipitation-Sequencing (ChIP-seq) Experiments with spikChIP 55
Enrique Blanco, Cecilia Ballaré, Luciano Di Croce, and Sergi Aranda

6 A Guide to MethylationToActivity: A Deep Learning Framework That Reveals Promoter Activity Landscapes from DNA Methylomes in Individual Tumors 73
Karissa Dieseldorff Jones, Daniel Putnam, Justin Williams, and Xiang Chen

7 DNA Modification Patterns Filtering and Analysis Using DNAModAnnot 87
Alexis Hardy, Sandra Duharcourt, and Matthieu Defrance

8 Methylome Imputation by Methylation Patterns 115
Ya-Ting Sabrina Chang, Ming-Ren Yen, and Pao-Yang Chen

9 Sequoia: A Framework for Visual Analysis of RNA Modifications from Direct RNA Sequencing Data 127
Ratanond Koonchanok, Swapna Vidhur Daulatabad, Khairi Reda, and Sarath Chandra Janga

10 Predicting Pseudouridine Sites with Porpoise 139
Xudong Guo, Fuyi Li, and Jiangning Song

11 Pseudouridine Identification and Functional Annotation with PIANO 153
Jiahui Yao, Cuiyueyue Hao, Kunqi Chen, Jia Meng, and Bowen Song

12 Analyzing mRNA Epigenetic Sequencing Data with TRESS 163
Zhenxing Guo, Andrew M. Shafik, Peng Jin, Zhijin Wu, and Hao Wu

13 Nanopore Direct RNA Sequencing Data Processing and Analysis Using MasterOfPores 185
Luca Cozzuto, Anna Delgado-Tejedor, Toni Hermoso Pulido, Eva Maria Novoa, and Julia Ponomarenko

14 Data Analysis Pipeline for Detection and Quantification of Pseudouridine (ψ) in RNA by HydraPsiSeq 207
Florian Pichot, Virginie Marchand, Mark Helm, and Yuri Motorin

15 Analysis of RNA Sequences and Modifications Using NASE 225
Samuel Wein

16 Mapping of RNA Modifications by Direct Nanopore Sequencing and JACUSA2 241
Amina Lemsara, Christoph Dieterich, and Isabel S. Naarmann-de Vries

Index *261*

Contributors

Sergi Aranda • *Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology, Barcelona, Spain*

Cecilia Ballaré • *Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology, Barcelona, Spain*

Enrique Blanco • *Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology, Barcelona, Spain*

Ya-Ting Sabrina Chang • *Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan*

Kunqi Chen • *Key Laboratory of Ministry of Education for Gastrointestinal Cancer, School of Basic Medical Sciences, Fujian Medical University, Fuzhou, Fujian, China*

Pao-Yang Chen • *Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan*

Xiang Chen • *Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, TN, USA*

Luca Cozzuto • *Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain*

Swapna Vidhur Daulatabad • *Department of BioHealth Informatics, School of Informatics and Computing, Indiana University Purdue University, Indianapolis, IN, USA*

Matthieu Defrance • *Université Libre de Bruxelles, Interuniversity Institute of Bioinformatics in Brussels (IB2), Brussels, Belgium*

Anna Delgado-Tejedor • *Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain; Universitat Pompeu Fabra (UPF), Barcelona, Spain*

Luciano Di Croce • *Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology, Barcelona, Spain; Universitat Pompeu Fabra (UPF), Barcelona, Spain; ICREA, Barcelona, Spain*

Karissa Dieseldorff Jones • *Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, TN, USA*

Christoph Dieterich • *Klaus Tschira Institute for Integrative Computational Cardiology, University Heidelberg, Heidelberg, Germany; Department of Internal Medicine III (Cardiology, Angiology, and Pneumology), University Hospital Heidelberg, Heidelberg, Germany; German Centre for Cardiovascular Research (DZHK)-Partner Site HD/MA, Heidelberg, Germany*

Sandra Duharcourt • *Université Paris Cité, CNRS, Institut Jacques Monod, 75013 , Paris, France*

Xudong Guo • *College of Information Engineering, Northwest A&F University, Yangling, China*

Zhenxing Guo • *Department of Biostatistics and Bioinformatics, Emory University Rollins School of Public Health, Atlanta, GA, USA*

Cuiyueyue Hao • *Department of Mathematical Sciences, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu, China*

Alexis Hardy • *Université Libre de Bruxelles, Interuniversity Institute of Bioinformatics in Brussels (IB2), Brussels, Belgium*

Mark Helm • *Institute of Pharmacy of Pharmacy and Biochemistry, Johannes Gutenberg University Mainz, Mainz, Germany*

Toni Hermoso Pulido • *Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain*

Sarath Chandra Janga • *Department of BioHealth Informatics, School of Informatics and Computing, Indiana University Purdue University, Indianapolis, IN, USA; Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA; Centre for Computational Biology and Bioinformatics, Indiana University School of Medicine, 5021 Health Information and Translational Sciences (HITS), Indianapolis, IN, USA*

Peng Jin • *Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA*

Ratanond Koonchanok • *Department of Human-Centered Computing, School of Informatics and Computing, Indiana University Purdue University, Indianapolis, IN, USA*

Amina Lemsara • *Klaus Tschira Institute for Integrative Computational Cardiology, University Heidelberg, Heidelberg, Germany; Department of Internal Medicine III (Cardiology, Angiology, and Pneumology), University Hospital Heidelberg, Heidelberg, Germany*

Fuyi Li • *College of Information Engineering, Northwest A&F University, Yangling, China; Department of Microbiology and Immunology, The Peter Doherty Institute for Infection and Immunity, The University of Melbourne, Melbourne, VIC, Australia*

Xiaojian Liu • *Institute of Cancer Research, Shenzhen Bay Laboratory, Shenzhen, China*

Virginie Marchand • *Université de Lorraine, CNRS, INSERM, UAR2008/US40 IBSLor, EpiRNA-Seq Core facility, F-54000, Nancy, France*

Jia Meng • *Department of Biological Sciences, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu, China; AI University Research Centre, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu, China; Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool, UK*

Octavio Morante-Palacios • *Epigenetics and Immune Disease Group, Josep Carreras Research Institute (IJC), Barcelona, Spain; Germans Trias i Pujol Research Institute (IGTP), Barcelona, Spain*

Yuri Motorin • *Université de Lorraine, CNRS, INSERM, UAR2008/US40 IBSLor, EpiRNA-Seq Core facility, F-54000, Nancy, France; Université de Lorraine, CNRS, UMR7365 IMoPA, F-54000, Nancy, France*

Isabel S. Naarmann-de Vries • *Klaus Tschira Institute for Integrative Computational Cardiology, University Heidelberg, Heidelberg, Germany; Department of Internal Medicine III (Cardiology, Angiology, and Pneumology), University Hospital Heidelberg, Heidelberg, Germany; German Centre for Cardiovascular Research (DZHK)-Partner Site HD/MA, Heidelberg, Germany*

Eva Maria Novoa • *Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain; Universitat Pompeu Fabra (UPF), Barcelona, Spain*

Florian Pichot • *Institute of Pharmacy of Pharmacy and Biochemistry, Johannes Gutenberg University Mainz, Mainz, Germany; Université de Lorraine, CNRS, INSERM, UAR2008/US40 IBSLor, EpiRNA-Seq Core facility, F-54000, Nancy, France*

Julia Ponomarenko • *Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain; Universitat Pompeu Fabra (UPF), Barcelona, Spain*

Daniel Putnam • *Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, TN, USA*

Khairi Reda • *Department of Human-Centered Computing, School of Informatics and Computing, Indiana University Purdue University, Indianapolis, IN, USA*

Ron Schwessinger • *MRC WIMM Centre for Computational Biology, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK*

Andrew M. Shafik • *Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA*

Bowen Song • *Department of Mathematical Sciences, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu, China; Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool, UK*

Jiangning Song • *Biomedicine Discovery Institute, Monash University, Melbourne, VIC, Australia; Monash Data Futures Institute, Monash University, Melbourne, VIC, Australia*

Kun Sun • *Institute of Cancer Research, Shenzhen Bay Laboratory, Shenzhen, China*

Kai Tan • *Center for Childhood Cancer Research, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; Department of Genetics, University of Pennsylvania, Philadelphia, PA, USA; Penn Epigenetics Institute, University of Pennsylvania, Philadelphia, PA, USA; Department of Pediatrics, University of Pennsylvania, Philadelphia, PA, USA*

Yasin Uzun • *Center for Childhood Cancer Research, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA*

Samuel Wein • *Center for Bioinformatics Tübingen, University of Tübingen, Tübingen, Germany*

Justin Williams • *Department of Tumor Cell Biology, St. Jude Children's Research Hospital, Memphis, TN, USA*

Hao Wu • *Department of Genetics, University of Pennsylvania, Philadelphia, PA, USA; Penn Epigenetics Institute, University of Pennsylvania, Philadelphia, PA, USA*

Hao Wu • *Department of Biostatistics and Bioinformatics, Emory University Rollins School of Public Health, Atlanta, GA, USA*

Zhijin Wu • *Department of Biostatistics, Brown University, Providence, RI, USA*

Jiahui Yao • *Department of Biological Sciences, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu, China*

Ming-Ren Yen • *Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan*

Pengxiang Yuan • *Institute of Cancer Research, Shenzhen Bay Laboratory, Shenzhen, China*