

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>

Computational Methods for Single-Cell Data Analysis

Edited by

Guo-Cheng Yuan

Dana–Farber Cancer Institute and Harvard Chan School of Public Health, Boston, MA, USA

 Humana Press

Editor

Guo-Cheng Yuan
Dana–Farber Cancer Institute and Harvard Chan
School of Public Health
Boston, MA, USA

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-9056-6 ISBN 978-1-4939-9057-3 (eBook)
<https://doi.org/10.1007/978-1-4939-9057-3>

Library of Congress Control Number: 2018967307

© Springer Science+Business Media, LLC, part of Springer Nature 2019

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. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Humana Press imprint is published by the registered company Springer Science+Business Media, LLC, part of Springer Nature.

The registered company address is: 233 Spring Street, New York, NY 10013, U.S.A.

Preface

The cell is the fundamental unit of life. The biological functions of an organ or tissue are results of coordinated action of a large number of cells, each having its own properties and dynamic behavior. While it is traditional to classify cells with similar function and morphology as cell types, it is also well-recognized that, even within each cell type, there remain significant differences, or states, among individual cells. Current knowledge of the repertoire of cell types and cell states, as well as their dynamic changes, remains highly incomplete. Systematic, comprehensive characterization of spatial and temporal organization of cellular heterogeneity, along with the mechanisms underlying cell-type/state transition and maintenance, has important implications in development and diseases.

It is not until recently that it has become feasible to systematically investigate cellular heterogeneity at the single-cell resolution, thanks to the rapid development of a number of advanced technologies including sequencing, imaging, and microfluidic devices. Collectively, single-cell technologies have created exciting opportunities to systematically characterize the molecular behavior of individual cells at the omics scale. At the same time, the analysis and integration of single-cell omic data are difficult due to a number of challenges such as sparsity, technical variability, and spatial-temporal complexity.

During the past few years, numerous computational methods and software packages have been developed to overcome these challenges. The aim of this book is to introduce to the community the state of the art of computational approaches in single-cell data analysis. Each chapter presents a computational toolbox that is aimed to overcome a specific challenge in single-cell analysis, such as data normalization, rare cell-type identification, and spatial transcriptomics analysis. Rather than explaining the mathematical details, here the focus is on hands-on implementation of computational methods for analyzing experimental data. Taken together, these chapters cover a wide range of tasks and may serve as a handbook for single-cell data analysis.

Finally, I would like to thank Prof. John M. Walker for his kind invitation and sustained support throughout the preparation of this book. I would also like to express my sincere gratitude to all the contributors for sharing their protocols.

Boston, MA, USA

Guo-Cheng Yuan

Contents

<i>Preface</i>	<i>v</i>
<i>Contributors</i>	<i>ix</i>
1 Quality Control of Single-Cell RNA-seq	1
<i>Peng Jiang</i>	
2 Normalization for Single-Cell RNA-Seq Data Analysis	11
<i>Rhonda Bacher</i>	
3 Analysis of Technical and Biological Variability in Single-Cell RNA Sequencing	25
<i>Beomseok Kim, Eunmin Lee, and Jong Kyoung Kim</i>	
4 Identification of Cell Types from Single-Cell Transcriptomic Data	45
<i>Karthik Shekhar and Vilas Menon</i>	
5 Rare Cell Type Detection	79
<i>Lan Jiang</i>	
6 scMCA: A Tool to Define Mouse Cell Types Based on Single-Cell Digital Expression	91
<i>Huiyu Sun, Yincong Zhou, Lijiang Fei, Haide Chen, and Guoji Guo</i>	
7 Differential Pathway Analysis	97
<i>Jean Fan</i>	
8 Pseudotime Reconstruction Using TSCAN	115
<i>Zhicheng Ji and Hongkai Ji</i>	
9 Estimating Differentiation Potency of Single Cells Using Single-Cell Entropy (SCENT)	125
<i>Weiyang Chen and Andrew E. Teschendorff</i>	
10 Inference of Gene Co-expression Networks from Single-Cell RNA-Sequencing Data	141
<i>Alicia T. Lamere and Jun Li</i>	
11 Single-Cell Allele-Specific Gene Expression Analysis	155
<i>Meichen Dong and Yuchao Jiang</i>	
12 Using BRIE to Detect and Analyze Splicing Isoforms in scRNA-Seq Data	175
<i>Yuanhua Huang and Guido Sanguinetti</i>	
13 Preprocessing and Computational Analysis of Single-Cell Epigenomic Datasets	187
<i>Caleb Laveau, Divy Kangayan, and Martin J. Aryee</i>	

14 Experimental and Computational Approaches for Single-Cell
Enhancer Perturbation Assay 203
Shiqi Xie and Gary C. Hon

15 Antigen Receptor Sequence Reconstruction and Clonality Inference
from scRNA-Seq Data 223
Ida Lindeman and Michael J. T. Stubbington

16 A Hidden Markov Random Field Model for Detecting Domain
Organizations from Spatial Transcriptomic Data 251
Qian Zhu

Index 269

Contributors

- MARTIN J. ARYEE • *Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA; Department of Pathology, Massachusetts General Hospital, Boston, MA, USA; Broad Institute of MIT and Harvard, Cambridge, MA, USA*
- RHONDA BACHER • *Department of Biostatistics, University of Florida, Gainesville, FL, USA*
- HAIDE CHEN • *Center for Stem Cell and Regenerative Medicine, Zhejiang University School of Medicine, Hangzhou, China; Stem Cell Institute, Zhejiang University, Hangzhou, China*
- WEIYAN CHEN • *CAS Key Lab of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, Shanghai Institute of Biological Sciences, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China*
- MEICHEN DONG • *Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA*
- JEAN FAN • *Department of Chemistry and Chemical Biology, Harvard University, Boston, MA, USA*
- LIJIANG FEI • *Center for Stem Cell and Regenerative Medicine, Zhejiang University School of Medicine, Hangzhou, China; Stem Cell Institute, Zhejiang University, Hangzhou, China*
- GUOJI GUO • *Center for Stem Cell and Regenerative Medicine, Zhejiang University School of Medicine, Hangzhou, China; Stem Cell Institute, Zhejiang University, Hangzhou, China*
- GARY C. HON • *Department of Obstetrics and Gynecology, Cecil H. and Ida Green Center for Reproductive Biology Sciences, University of Texas Southwestern Medical Center, Dallas, TX, USA*
- YUANHUA HUANG • *EMBL-European Bioinformatics Institute, Cambridgeshire, UK*
- HONGKAI JI • *Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA*
- ZHICHENG JI • *Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA*
- LAN JIANG • *Howard Hughes Medical Institute, Boston Children's Hospital, Boston, MA, USA; Program in Cellular and Molecular Medicine, Boston Children's Hospital, Boston, MA, USA; Division of Hematology/Oncology, Department of Pediatrics, Boston Children's Hospital, Boston, MA, USA*
- PENG JIANG • *Regenerative Biology Laboratory, Morgridge Institute for Research, Madison, WI, USA*
- YUCHAO JIANG • *Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA; Department of Genetics, School of Medicine, University of North Carolina, Chapel Hill, NC, USA; Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, USA*
- DIVY KANGAYAN • *Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA; Department of Pathology, Massachusetts General Hospital, Boston, MA, USA*
- BEOMSEOK KIM • *Department of New Biology, DGIST, Daegu, Republic of Korea*
- JONG KYOUNG KIM • *Department of New Biology, DGIST, Daegu, Republic of Korea*
- ALICIA T. LAMERE • *Mathematics Department, Bryant University, Smithfield, RI, USA*

- CALEB LAREAU • *Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA; Department of Pathology, Massachusetts General Hospital, Boston, MA, USA*
- EUNMIN LEE • *Department of New Biology, DGIST, Daegu, Republic of Korea*
- JUN LI • *Applied and Computational Mathematics and Statistics Department, University of Notre Dame, Notre Dame, IN, USA*
- IDA LINDEMAN • *Wellcome Sanger Institute, Hinxton, Cambridge, UK; KG Jebsen Coeliac Disease Research Centre and Department of Immunology, University of Oslo, Oslo, Norway*
- VILAS MENON • *Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA; Columbia University Medical Center, New York, NY, USA*
- GUIDO SANGUINETTI • *School of Informatics, University of Edinburgh, Edinburgh, UK*
- KARTHIK SHEKHAR • *Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA*
- MICHAEL J. T. STUBBINGTON • *Wellcome Sanger Institute, Hinxton, Cambridge, UK*
- HUIYU SUN • *Center for Stem Cell and Regenerative Medicine, Zhejiang University School of Medicine, Hangzhou, China; Stem Cell Institute, Zhejiang University, Hangzhou, China*
- ANDREW E. TESCHENDORFF • *CAS Key Lab of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, Shanghai Institute of Biological Sciences, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China; UCL Cancer Institute, University College London, London, UK*
- SHIQI XIE • *Department of Obstetrics and Gynecology, Cecil H. and Ida Green Center for Reproductive Biology Sciences, University of Texas Southwestern Medical Center, Dallas, TX, USA*
- YINCONG ZHOU • *Stem Cell Institute, Zhejiang University, Hangzhou, China; College of Life Sciences, Zhejiang University, Hangzhou, China*
- QIAN ZHU • *Dana-Farber Cancer Institute, Boston, MA, USA*