

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>

Gene Regulatory Networks

Methods and Protocols

Edited by

Guido Sanguinetti

School of Informatics, University of Edinburgh, Edinburgh, UK

Vân Anh Huynh-Thu

Department of Electrical Engineering and Computer Science, University of Liège, Liège, Belgium

Editors

Guido Sanguinetti
School of Informatics
University of Edinburgh
Edinburgh, UK

Vân Anh Huynh-Thu
Department of Electrical Engineering
and Computer Science
University of Liège
Liège, Belgium

ISSN 1064-3745

ISSN 1940-6029 (electronic)

Methods in Molecular Biology

ISBN 978-1-4939-8881-5

ISBN 978-1-4939-8882-2 (eBook)

<https://doi.org/10.1007/978-1-4939-8882-2>

Library of Congress Control Number: 2018962962

© 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

High-throughput technologies have brought about a revolution in molecular biology. Over the last two decades, the research paradigm has moved from a characterization of individual genes and their function increasingly toward a systems-level appreciation of how the complex interactions between multiple genes shape the dynamics and functions of biological systems. At the same time, the computational and statistical challenges posed by the interpretation of such data have motivated an exciting cross-fertilization between the disciplines of biology and the mathematical and computational sciences, leading to the birth of the interdisciplinary field of systems biology.

A crucial computational task in systems biology is the so-called *reverse engineering* task: given observations of multiple biological features (e.g., protein levels) across different time points/conditions, determine computationally the interaction structure (the *network*) that best explains the data. Within the context of modeling gene expression, this is the task of inferring *gene regulatory networks* (GRNs) from data.

GRN inference has been a major challenge in systems biology for nearly two decades, and, while challenges still abound, it is rapidly reaching maturity both in terms of the concepts involved, and in terms of the software tools available. In this book, we aim to take stock of the situation, providing an overview of methods that cover the majority of recent developments, as well as indicating the path forward for future developments.

The book opens with a tutorial overview of the main biological and mathematical concepts and a survey of the current software landscape. This is meant to be an entry level chapter, which the interested, graduate-level practitioner (either computational or biological) can consult as a rough guide to the concepts described more in detail in the more technical chapters in the book. The next two chapters then focus on Bayesian methods to infer networks from time varying data, while Chapters 4 and 5 describe how to attempt to extract causal information (as opposed to purely correlative) from biological data. Chapters 6 and 7 describe network inference techniques in the presence of multiple heterogeneous data sets, while Chapters 8 and 9 focus on nonparametric and hybrid statistical methods for network inference. The following five Chapters 10–14 focus on the idea of inference of different (but related) networks, arising either from intrinsic heterogeneity (such as in single-cell data) or due to multiple conditions being assayed, and further explore concepts of differential networks and network stability. Finally, the last two chapters focus more on a mechanistic view of the biological process, covering methods for exploring networks within large, mechanistic models of biological dynamics.

As most books, this volume presents an incomplete snapshot of an evolving field, and, given the considerable research activity in this area, it is clear that we can look forward to considerable progress within the next decades. Our hope is that this collection will be instrumental in assessing the current state of the art and in focusing research on the common challenges faced by the field.

Edinburgh, UK
Liège, Belgium

Guido Sanguinetti
Vân Anh Huynh-Thu

Contents

<i>Preface</i>	<i>v</i>
<i>Contributors</i>	<i>ix</i>
1 Gene Regulatory Network Inference: An Introductory Survey.....	1
<i>Vân Anh Huynh-Thu and Guido Sanguinetti</i>	
2 Statistical Network Inference for Time-Varying Molecular Data with Dynamic Bayesian Networks.....	25
<i>Frank Dondelinger and Sach Mukherjee</i>	
3 Overview and Evaluation of Recent Methods for Statistical Inference of Gene Regulatory Networks from Time Series Data	49
<i>Marco Grzegorzczuk, Andrej Aderhold and Dirk Husmeier</i>	
4 Whole-Transcriptome Causal Network Inference with Genomic and Transcriptomic Data	95
<i>Lingfei Wang and Tom Michoel</i>	
5 Causal Queries from Observational Data in Biological Systems via Bayesian Networks: An Empirical Study in Small Networks	111
<i>Alex White and Matthieu Vignes</i>	
6 A Multiattribute Gaussian Graphical Model for Inferring Multiscale Regulatory Networks: An Application in Breast Cancer	143
<i>Julien Chiquet, Guillem Rigaill, and Martina Sundqvist</i>	
7 Integrative Approaches for Inference of Genome-Scale Gene Regulatory Networks.....	161
<i>Alireza Fotuhi Siahpirani, Deborah Chasman, and Sushmita Roy</i>	
8 Unsupervised Gene Network Inference with Decision Trees and Random Forests	195
<i>Vân Anh Huynh-Thu and Pierre Geurts</i>	
9 Tree-Based Learning of Regulatory Network Topologies and Dynamics with Jump3	217
<i>Vân Anh Huynh-Thu and Guido Sanguinetti</i>	
10 Network Inference from Single-Cell Transcriptomic Data	235
<i>Helena Todorov, Robrecht Cannoodt, Wouter Saelens, and Yvan Saeys</i>	
11 Inferring Gene Regulatory Networks from Multiple Datasets	251
<i>Christopher A. Penfold, Iulia Gherman, Anastasiya Sybirna, and David L. Wild</i>	
12 Unsupervised GRN Ensemble	283
<i>Pau Bellot, Philippe Salembier, Ngoc C. Pham, and Patrick E. Meyer</i>	
13 Learning Differential Module Networks Across Multiple Experimental Conditions	303
<i>Pau Erola, Eric Bonnet, and Tom Michoel</i>	

14	Stability in GRN Inference	323
	<i>Giuseppe Jurman, Michele Filosi, Roberto Visintainer, Samantha Riccadonna, and Cesare Furlanello</i>	
15	Gene Regulatory Networks: A Primer in Biological Processes and Statistical Modelling	347
	<i>Olivia Angelin-Bonnet, Patrick J. Biggs and Matthieu Vignes</i>	
16	Scalable Inference of Ordinary Differential Equation Models of Biochemical Processes	385
	<i>Fabian Fröhlich, Carolin Loos, and Jan Hasenauer</i>	
	<i>Index</i>	423

Contributors

- ANDREJ ADERHOLD • *School of Mathematics and Statistics, University of Glasgow, Glasgow, UK*
- OLIVIA ANGELIN-BONNET • *Institute of Fundamental Sciences, Palmerston North, New Zealand*
- PAU BELLOT • *Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB Consortium, Bellaterra, Barcelona, Spain*
- PATRICK J. BIGGS • *Institute of Fundamental Sciences, Palmerston North, New Zealand; School of Veterinary Science, Massey University, Palmerston North, New Zealand*
- ERIC BONNET • *Centre Nationale de Recherche en Genomique Humaine, Institut de Biologie Francois Jacob, Direction de la Recherche Fondamentale, CEA, Evry, France*
- ROBRECHT CANNOODT • *Data Mining and Modelling for Biomedicine, VIB Center for Inflammation Research, Ghent, Belgium; Center for Medical Genetics, Ghent University Hospital, Ghent, Belgium*
- DEBORAH CHASMAN • *Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, WI, USA*
- JULIEN CHIQUET • *UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, Paris, France*
- FRANK DONDELINGER • *Lancaster Medical School, Lancaster University, Lancaster, UK*
- PAU EROLA • *Division of Genetics and Genomics, The Roslin Institute, The University of Edinburgh, Midlothian, Scotland, UK*
- MICHELE FILOSI • *CIBIO, University of Trento, Trento, Italy*
- FABIAN FRÖHLICH • *Institute of Computational Biology, Helmholtz Zentrum München, Neuherberg, Germany; Center for Mathematics, Technische Universität München, Garching, Germany*
- CESARE FURLANELLO • *Fondazione Bruno Kessler, Trento, Italy*
- PIERRE GEURTS • *Department of Electrical Engineering and Computer Science, University of Liège, Liège, Belgium*
- IULIA GHERMAN • *Warwick Integrative Synthetic Biology Centre, School of Engineering, University of Warwick, Coventry, UK*
- MARCO GRZEGORCZYK • *Johann Bernoulli Institute, University of Groningen, Groningen, The Netherlands*
- JAN HASENAUER • *Institute of Computational Biology, Helmholtz Zentrum Muenchen, Neuherberg, Germany; Center for Mathematics, Technische Universität München, Garching, Germany*
- DIRK HUSMEIER • *School of Mathematics and Statistics, University of Glasgow, Glasgow, UK*
- VÂN ANH HUYNH-THU • *Department of Electrical Engineering and Computer Science, University of Liège, Liège, Belgium*
- GIUSEPPE JURMAN • *Fondazione Bruno Kessler, Trento, Italy*
- CAROLIN LOOS • *Institute of Computational Biology, Helmholtz Zentrum München, Neuherberg, Germany; Center for Mathematics, Technische Universität München, Garching, Germany*

- PATRICK E. MEYER • *Bioinformatics and Systems Biology (BioSys) Unit, Université de Liège, Liège, Belgium*
- TOM MICHOEL • *Division of Genetics and Genomics, The Roslin Institute, The University of Edinburgh, Midlothian, Scotland, UK; Computational Biology Unit, Department of Informatics, University of Bergen, Bergen, Norway*
- SACH MUKHERJEE • *German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany*
- CHRISTOPHER A. PENFOLD • *Wellcome/CRUK Gurdon Institute, University of Cambridge, Cambridge, UK*
- NGOC C. PHAM • *Bioinformatics and Systems Biology (BioSys) Unit, Université de Liège, Liège, Belgium*
- SAMANTHA RICCADONNA • *Fondazione Edmund Mach, San Michele all'Adige, Italy*
- GUILLEM RIGAILL • *Institute of Plant Sciences Paris-Saclay, UMR 9213/UMR1403, CNRS, INRA, Université Paris -Sud, Université d'Evry, Université Paris-Diderot, Sorbonne Paris-Cité, Paris, France; Laboratoire de Mathématiques et Modélisation d'Evry (LaMME), Université d'Evry, Val d'Essonne, UMR CNRS 8071, ENSIE, USC INRA, Paris, France*
- SUSHMITA ROY • *Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, WI, USA; Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, USA*
- WOUTER SAELENS • *Data Mining and Modelling for Biomedicine, VIB Center for Inflammation Research, Ghent, Belgium; Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium*
- YVAN SAEYS • *Data Mining and Modelling for Biomedicine, VIB Center for Inflammation Research, Ghent, Belgium; Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium*
- PHILIPPE SALEMBIER • *Universitat Politècnica de Catalunya, Barcelona, Spain*
- GUIDO SANGUINETTI • *School of Informatics, University of Edinburgh, Edinburgh, UK*
- ALIREZA FOTUHI SIAHPIRANI • *Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, WI, USA; Department of Computer Sciences, University of Wisconsin-Madison, Madison, WI, USA*
- MARTINA SUNDQVIST • *UMR MIA-Paris, AgroTechParis, INRA, Université Paris-Saclay, Paris, France*
- ANASTASIYA SYBIRNA • *Wellcome/CRUK Gurdon Institute, University of Cambridge, Cambridge, UK; Wellcome/MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK; Physiology, Development and Neuroscience Department, University of Cambridge, Cambridge, UK*
- HELENA TODOROV • *Data Mining and Modelling for Biomedicine, VIB Center for Inflammation Research, Ghent, Belgium; Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium; Centre International de Recherche en Infectiologie, Inserm, U1111, Université Claude Bernard Lyon 1, CNRS, UMR5308, École Normale Supérieure de Lyon, Univ Lyon, Lyon, France*
- MATTHIEU VIGNES • *Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand*
- ROBERTO VISINTAINER • *CIBIO, University of Trento, Trento, Italy*
- LINGFEI WANG • *Division of Genetics and Genomics, The Roslin Institute, The University of Edinburgh, Midlothian, Scotland, UK*

ALEX WHITE • *Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand*
DAVID L. WILD • *Department of Statistics and Systems Biology Centre, University of Warwick, Coventry, UK*