

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>

Bioinformatics in MicroRNA Research

Editors

Jingshan Huang

School of Computing, University of South Alabama, Mobile, AL, USA

Glen M. Borchert

*Department of Pharmacology, University of South Alabama, Mobile, AL, USA;
Department of Biology, University of South Alabama, Mobile, AL, USA*

Dejing Dou

Department of Computer and Information Science, University of Oregon, Eugene, OR, USA

Jun (Luke) Huan

Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS, USA

Wenjun Lan

School of Bio-Engineering, Qilu University of Technology, Jinan, Shandong, China

Ming Tan

Mitchel Cancer Institute, University of South Alabama, Mobile, AL, USA

Bin Wu

Department of Endocrinology, First Affiliated Hospital, Kunming Medical University, Kunming, Yunnan, China

 **Humana Press**

Editors

Jingshan Huang
School of Computing
University of South Alabama
Mobile, AL, USA

Dejing Dou
Department of Computer
and Information Science
University of Oregon
Eugene, OR, USA

Wenjun Lan
School of Bioengineering
Qilu University of Technology
Jinan, Shandong, China

Bin Wu
Department of Endocrinology
First Affiliated Hospital
Kunming Medical University
Kunming, Yunnan, China

Glen M. Borchert
Department of Pharmacology
University of South Alabama
Mobile, AL, USA

Department of Biology
University of South Alabama
Mobile, AL, USA

Jun (Luke) Huan
Department of Electrical Engineering
and Computer Science
University of Kansas
Lawrence, KS, USA

Ming Tan
Mitchel Cancer Institute
University of South Alabama
Mobile, AL, USA

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-7044-5 ISBN 978-1-4939-7046-9 (eBook)
DOI 10.1007/978-1-4939-7046-9

Library of Congress Control Number: 2017937362

© Springer Science+Business Media LLC 2017

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.

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

As a special class of noncoding RNAs, microRNAs (miRNAs or miRs for short) have been reported to perform important roles in various biological and pathological processes by regulating respective target genes. To completely understand and fully delineate miR functions, besides performing biological experiments and querying PubMed and TarBase for biologically validated miR targets, biologists can also query various miR target prediction databases/websites for computationally predicted targets. More often than not, biologists need to extract additional information for each and every miR target, either validated or putative, with regard to its related information such as protein functions and affiliated signaling pathways. In short, biologists are facing significant barriers in fully delineating miR functions and the following effective bio-curation. Therefore, there is an urgent need for a comprehensive book focusing on miR target genes, miR regulation mechanisms, miR functions performed in various human diseases, and miR databases/knowledge bases.

This book is intended to give an in-depth introduction to and discussion of miRs and their targets, miR functions, and computational techniques applied in miR research. The primary audience includes, but is not limited to, computational biologists, computer scientists, bioinformaticians, bench biologists, and clinical investigators. No prior knowledge of computer science, databases, semantic technologies, or molecular biology is assumed. But we do assume that readers have some biology background knowledge at the high-school level.

A brief overview of the book structure is as follows. Chapter 1 introduces the concepts of miRs and long noncoding RNAs (lncRNAs) as well as some recent advances in miR/lncRNA biology. Chapters 2, 3, and 4 discuss protein participants in miR regulation; viral microRNAs, host miRs regulating viruses, and bacterial miR-like RNAs; and biomarkers, diagnostics, and therapeutics aspects of miRs, respectively. Chapter 5 introduces basic concepts of relational databases and biomedical big data. Chapter 6 provides an overview of semantic technologies and bio-ontologies. Chapter 7 discusses genome-wide analysis of miR-regulated transcripts. Chapters 8 and 9 describe in detail computational prediction of miR target genes, regulatory interactions between miRs and their targets, as well as an introduction of various miR target prediction databases and relevant Web resources. Chapter 10 discusses some limitations of existing approaches that aim to improve miR target prediction accuracy. Chapters 11 and 12 introduce genomic regulation of miR expression in disease development and next generation sequencing for miR expression profile. Chapters 13 through 16 discuss advanced topics in computational/bioinformatics approaches in miR research, including the handling of high-dimension data, identification and removal of noisy data, logical reasoning, and machine learning techniques. Finally, Chapters 17–19 introduce some advances of miR research in three human diseases: diabetes, obesity, and thyroid carcinoma.

Mobile, AL, USA

Mobile, AL, USA

Eugene, OR, USA

Lawrence, KS, USA

Jinan, Shandong, China

Mobile, AL, USA

Kunming, Yunnan, China

Jingshan Huang

Glen M. Borchert

Dejing Dou

Jun (Luke) Huan

Wenjun Lan

Ming Tan

Bin Wu

Contents

<i>Preface</i>	<i>v</i>
<i>Contributors</i>	<i>ix</i>
1 MicroRNAs, Long Noncoding RNAs, and Their Functions in Human Disease	1
<i>Min Xue, Ying Zhuo, and Bin Shan</i>	
2 MicroRNA Expression: Protein Participants in MicroRNA Regulation	27
<i>Valeria M. King and Glen M. Borchert</i>	
3 Viral MicroRNAs, Host MicroRNAs Regulating Viruses, and Bacterial MicroRNA-Like RNAs.	39
<i>Sara-Elizabeth Cardin and Glen M. Borchert</i>	
4 MicroRNAs: Biomarkers, Diagnostics, and Therapeutics	57
<i>Weili Huang</i>	
5 Relational Databases and Biomedical Big Data	69
<i>N.H. Nisansa D. de Silva</i>	
6 Semantic Technologies and Bio-Ontologies	83
<i>Fernando Gutierrez</i>	
7 Genome-Wide Analysis of MicroRNA-Regulated Transcripts	93
<i>David Chevalier and Glen M. Borchert</i>	
8 Computational Prediction of MicroRNA Target Genes, Target Prediction Databases, and Web Resources	109
<i>Justin T. Roberts and Glen M. Borchert</i>	
9 Exploring MicroRNA::Target Regulatory Interactions by Computing Technologies	123
<i>Yue Hu, Wenjun Lan, and Daniel Miller</i>	
10 The Limitations of Existing Approaches in Improving MicroRNA Target Prediction Accuracy	133
<i>Rasiyah Loganantharaj and Thomas A. Randall</i>	
11 Genomic Regulation of MicroRNA Expression in Disease Development	159
<i>Feng Liu</i>	
12 Next-Generation Sequencing for MicroRNA Expression Profile	169
<i>Yue Hu, Wenjun Lan, and Daniel Miller</i>	
13 Handling High-Dimension (High-Feature) MicroRNA Data	179
<i>Yue Hu, Wenjun Lan, and Daniel Miller</i>	
14 Effective Removal of Noisy Data Via Batch Effect Processing	187
<i>Ryan G. Benton</i>	
15 Logical Reasoning (Inferencing) on MicroRNA Data	197
<i>Jingsong Wang</i>	

16	Machine Learning Techniques in Exploring MicroRNA Gene Discovery, Targets, and Functions	211
	<i>Sumi Singh, Ryan G. Benton, Anurag Singh, and Anshuman Singh</i>	
17	Involvement of MicroRNAs in Diabetes and Its Complications	225
	<i>Bin Wu and Daniel Miller</i>	
18	MicroRNA Regulatory Networks as Biomarkers in Obesity: The Emerging Role	241
	<i>Lihua Zhang, Daniel Miller, Qiuping Yang, and Bin Wu</i>	
19	Expression of MicroRNAs in Thyroid Carcinoma	261
	<i>Gaohong Zhu, Lijun Xie, and Daniel Miller</i>	
	<i>Index</i>	281

Contributors

- RYAN G. BENTON • *Department of Computer Science, University of South Alabama School of Computing, Mobile, AL, USA*
- GLEN M. BORCHERT • *Department of Pharmacology, University of South Alabama, Mobile, AL, USA; Department of Biology, University of South Alabama, Mobile, AL, USA*
- SARA-ELIZABETH CARDIN • *Department of Biology, University of South Alabama, Mobile, AL, USA*
- DAVID CHEVALIER • *Department of Biology, East Georgia State College, Swainsboro, GA, USA*
- N.H. NISANSA D. DE SILVA • *Department of Computer and Information Science, University of Oregon, Eugene, OR, USA*
- FERNANDO GUTIERREZ • *Department of Computer and Information Science, University of Oregon, Eugene, OR, USA*
- YUE HU • *College of Bioengineering, Qilu University of Technology, Jinan, Shandong, People's Republic of China*
- WEILI HUANG • *Miracle Query, Incorporated, Eugene, OR, USA*
- VALERIA M. KING • *Department of Biology, University of South Alabama, Mobile, AL, USA*
- WENJUN LAN • *School of Bioengineering, Qilu University of Technology, Jinan, Shandong, People's Republic of China*
- FENG LIU • *National Research Center for Translational Medicine (Shanghai), Rui-Jin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China*
- RASIAH LOGANANTHARAJ • *Bioinformatics Research Lab, The Center for Advanced Computer Studies, University of Louisiana, Lafayette, LA, USA*
- DANIEL MILLER • *School of Computing, University of South Alabama, Mobile, AL, USA*
- THOMAS A. RANDALL • *Integrative Bioinformatics, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, Durham, NC, USA*
- JUSTIN T. ROBERTS • *Department of Biology, University of South Alabama, Mobile, AL, USA*
- BIN SHAN • *Elson S. Floyd College of Medicine, Washington State University, Spokane, WA, USA*
- ANSHUMAN SINGH • *School of Computer Science and Mathematics, University of Central Missouri, Warrensburg, MO, USA*
- ANURAG SINGH • *Center for Advanced Computer Studies, University of Louisiana, Lafayette, LA, USA*
- SUMI SINGH • *School of Computer Science and Mathematics, University of Central Missouri, Warrensburg, MO, USA*
- JINGSONG WANG • *Oracle Corporation, Redwood Shores, CA, USA*
- BIN WU • *Department of Endocrinology, First Affiliated Hospital, Kunming Medical University, Kunming, Yunnan, China*

LIJUN XIE • *Department of Nuclear Medicine, First Affiliated Hospital of Kunming Medical University, Kunming, Yunnan, China*

MIN XUE • *Xuzhou College of Medicine, Xuzhou, Jiangsu, China*

QIUPING YANG • *Department of Geriatrics, First Affiliated Hospital of Kunming Medical University, Kunming, Yunnan, China*

LIHUA ZHANG • *Department of Geriatrics, First Affiliated Hospital of Kunming Medical University, Kunming, Yunnan, China*

GAOHONG ZHU • *Department of Nuclear Medicine, First Affiliated Hospital of Kunming Medical University, Kunming, Yunnan, China*

YING ZHUO • *Kadlec Regional Medical Center, Richland, WA, USA*