Methods in Molecular Biology 1912

Springer Protocols

Xin Lai
Shailendra K. Gupta
Julio Vera *Editors*

Computational Biology of Non-Coding RNA

Methods and Protocols

Humana Press

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 Biology of Non-Coding RNA

Methods and Protocols

Edited by

Xin Lai

Department of Dermatology, Erlangen University Hospital, Erlangen, Germany

Shailendra K. Gupta

Department of Systems Biology and Bioinformatics, University of Rostock, Rostock, Germany

Julio Vera

Department of Dermatology, Erlangen University Hospital, Erlangen, Germany

Humana Press

Editors
Xin Lai
Department of Dermatology
Erlangen University Hospital
Erlangen, Germany

Julio Vera
Department of Dermatology
Erlangen University Hospital
Erlangen, Germany

Shailendra K. Gupta
Department of Systems Biology
and Bioinformatics
University of Rostock
Rostock, Germany

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-8981-2 ISBN 978-1-4939-8982-9 (eBook)
https://doi.org/10.1007/978-1-4939-8982-9

Library of Congress Control Number: 2018965598

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

Noncoding RNAs (ncRNAs) have emerged as a major class of regulatory genes for a diverse range of biological functions. ncRNAs exert their effect in the context of complex regulatory networks by targeting protein-coding genes, as well as other ncRNAs. Computational methods, such as high-throughput sequencing data analysis, biochemical network reconstruction and analysis, and kinetic modeling, have been applied to provide the functional characterization and annotation of ncRNAs. The aim of this volume is to review current knowledge, novel methods, and open challenges in understanding the regulatory role of ncRNAs in biomedicine. In particular, this volume offers a collection of state-of-the-art methods including the identification of novel ncRNAs and their targets, functional annotation, and disease association in different biological contexts.

The book is divided into five parts: (1) Overview of Disease-Specific ncRNAs, (2) Computational Methods and Workflows for ncRNA Discovery and Annotation Based on High-Throughput Sequencing Data, (3) Bioinformatics Tools and Databases for ncRNA Analyses, (4) Network-Based Methods for Characterizing ncRNA Function, and (5) Kinetic Modeling of ncRNA-Mediated Gene Regulation. In the first part, Schulte et al. discuss in detail the current knowledge on the role of regulatory microRNAs (miRNAs) and long ncRNA (lncRNA) in inflammatory and infectious diseases (Chapter 1). In Chapter 2, Logotheti, Marquardt, and Pützer discuss the cross-talk between p73 and miRNAs in realizing the cancer metastasis phenotype. This chapter also highlights translational opportunities and research challenges for the p73/miRNA cancer network in the context of cancer metastasis.

The subsequent five chapters describe experimental and computational methods, workflows that are useful for identification and functional annotation of circular RNAs (circRNAs), lncRNAs, and miRNAs. In Chapter 3, Sharma et al. describe software and workflows to identify circRNAs from RNA sequencing (RNA-seq) data as well as a step-by-step procedure to experimentally validate in silico identified circRNAs. Chapter 4 by Mathew et al. presents methods and pipelines for the discovery of novel lncRNAs from RNA-seq data of zebrafish. In Chapter 5, Wolfien et al. review and discuss state-of-the-art experimental and computational workflows for the identification and characterization of ncRNAs, and they emphasize the use of transcriptome-wide association studies, molecular network analyses, and artificial intelligence-guided predictions. They also propose a promising strategy for developing reproducible and sharable computational workflows for the study of ncRNAs using RNA-seq data. In Chapter 6, Nigita et al. review methods and resources for genome-wide detection of RNA editing in ncRNAs and discuss the features of ncRNA editing associated with their function as well as with the onset of human diseases. Finally, Demirci, Yousef, and Allmer present computational methodologies allowing the identification and validation of functional miRNA-mRNA interactions and discuss key points for investigating miRNAs as biomarkers of human diseases (Chapter 7).

The next three chapters present bioinformatics tools and databases for ncRNA analyses. In Chapter 8, Bagnacani et al. present a strategy that makes use of community effort to develop sharable and reusable computational workflows using the Galaxy platform. The concept is exemplified with a database for the identification and characterization of cooperative miRNA regulation of gene expression. In Chapter 9, Monga and Kumar provide a

comprehensive review of in silico miRNA resources and databases. They present tools for miRNA annotation, miRNA-disease association and identification and characterization of miRNA variants. In Chapter 10, Maracaja-Coutinho et al. present an up-to-date review about 229 databases related with ncRNA research with particular emphasis on biomedicine, including ncRNA databases for cancer, cardiovascular, nervous systems, pathogens, and other diseases.

Chapters 11–13 highlight network-based methods for characterizing ncRNA functions. In Chapter 11, Nacher and Akutsu review the current research on network controllability methods. The authors also present a tripartite network model including ncRNA-target and protein-disease layers and analyze the network with controllability methods to identify the subset of critical control ncRNAs associated with human diseases. Piro and Marsico summarize recent developments in lncRNA function prediction and lncRNA-disease associations using methods based on network analysis and competing endogenous RNAs (ceRNAs) function prediction (Chapter 12). Paul presents a mutual information-based Maximum-Relevance Maximum-Significance algorithm to identify miRNA-mRNA regulatory modules from interaction networks in a case study on gynecologic cancer (Chapter 13).

The last four chapters of this book showcase the use of kinetic modeling for providing quantitative understanding about mechanisms of ncRNA-mediated regulation of gene expression. In Chapter 14, Bocci et al. review and discuss kinetic models that investigate the ability of miRNA-mediated feedback loops in provoking bistability in gene expression. They link this phenomenon with the switch of cellular phenotypes and the control of intrinsic and extrinsic signaling noise. In Chapter 15, Martirosyan et al. review kinetic models that describe competition and depletion of shared miRNAs by ceRNAs, showing the importance of ceRNA effect in processing gene expression noise. In Chapter 16, Tian et al. demonstrated the use of kinetic modeling to explain the emerging properties associated with miRNA-mRNA reciprocal regulation and show that such regulation can lead to bistable switches in gene expression directing cell fate decision. In the last chapter, Rabajante and Rosario discuss mathematical modeling frameworks for studying lncRNA regulation of the mammalian cell cycle. They also present a model of ordinary differential equations to simulate and analyze cell cycle regulation in response to DNA damage.

Taken together, all chapters presented in this book provide a state-of-the-art collection of computational methods and approaches that will be of value to researchers interested in the ncRNA field. Computational biology-based methods for the identification and analyses of ncRNAs in the context of biomedicine are rapidly growing with both challenges and opportunities to identify ncRNA biomarkers and to develop ncRNA therapeutics. We believe that this volume is a valuable and useful resource for addressing many of these exciting challenges in ncRNA research and hope it will be of interest to many peers. This book is also dedicated to the loss of Dr. Baltazar D. Aguda who has made numerous and important contribution to improve our understanding of ncRNA-mediated regulation of gene expression using mathematical models.

Erlangen, Germany — *Xin Lai*
Rostock, Germany — *Shailendra K. Gupta*
Erlangen, Germany — *Julio Vera*

Contents

Contributors

Mario Acunzo • *Division of Pulmonary Diseases and Critical Care Medicine, Virginia Commonwealth University, Richmond, VA, USA*

Tatsuya Akutsu • *Bioinformatics Center, Institute for Chemical Research, Kyoto University, Kyoto, Japan*

Jens Allmer • *Applied Bioinformatics, Bioscience, Wageningen University & Research, Wageningen, The Netherlands*

Andrea Bagnacani • *Department of Systems Biology and Bioinformatics, Institute of Computer Science, University of Rostock, Rostock, Germany*

Chiara Enrico Bena • *DISAT, Politecnico di Torino, Turin, Italy; Italian Institute for Genomic Medicine, Turin, Italy*

Wilhelm Bertrams • *Institute for Lung Research, Universities of Giessen and Marburg Lung Center, Member of the German Center for Lung Research (DZL), Philipps-University Marburg, Marburg, Germany*

Federico Bocci • *Center for Theoretical Biological Physics, Rice University, Houston, TX, USA; Department of Chemistry, Rice University, Houston, TX, USA*

Pedro Vinícius Borges • *Department of Computer Science, Bioinformatics Graduation Program (PPGBIOINFO), Federal University of Technology—Paraná, Cornélio Procópio, Brazil*

Carla Bosia • *DISAT, Politecnico di Torino, Turin, Italy; Italian Institute for Genomic Medicine, Turin, Italy*

David Leon Brauer • *Department of Systems Biology and Bioinformatics, Institute of Computer Science, University of Rostock, Rostock, Germany*

Federica Calore • *Department of Cancer Biology and Genetics, Comprehensive Cancer Center, The Ohio State University, Columbus, OH, USA*

José Carlos Caris-Maldonado • *Advanced Center for Chronic Diseases—ACCDiS, Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile, Santiago, Chile*

Carlo M. Croce • *Department of Cancer Biology and Genetics, Comprehensive Cancer Center, The Ohio State University, Columbus, OH, USA*

Rosario Distefano • *Department of Cancer Biology and Genetics, Comprehensive Cancer Center, The Ohio State University, Columbus, OH, USA*

Alan Mitchell Durham • *Computer Science Department, Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, Brazil*

Almir José Ferreira • *Computer Science Department, Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, Brazil*

Marco Del Giudice • *DISAT, Politecnico di Torino, Turin, Italy; Italian Institute for Genomic Medicine, Turin, Italy*

Hanah Goetz • *School of Biological and Health Systems Engineering, Arizona State University, Tempe, AZ, USA*

Judith Hariprakash • *G.N. Ramachandran Knowledge Center for Bioinformatics, CSIR Institute of Genomics and Integrative Biology (CSIR-IGIB), Delhi, India*

Mohit Kumar Jolly • *Center for Theoretical Biological Physics, Rice University, Houston, TX, USA*

Kriti Kaushik • *Genomics and Molecular Medicine, CSIR Institute of Genomics and Integrative Biology (CSIR-IGIB), Delhi, India; Academy of Scientific and Innovative Research (AcSIR), New Delhi, India*

Manoj Kumar • *Bioinformatics Centre, Institute of Microbial Technology, Council of Scientific and Industrial Research, Chandigarh, India*

Herbert Levine • *Center for Theoretical Biological Physics, Rice University, Houston, TX, USA; Department of Chemistry, Rice University, Houston, TX, USA; Department of Bioengineering, Rice University, Houston, TX, USA; Department of Physics and Astronomy, Rice University, Houston, TX, USA*

Stella Logotheti • *Institute of Experimental Gene Therapy and Cancer Research, Rostock University Medical Center, Rostock, Germany*

Vinicius Maracaja-Coutinho • *Advanced Center for Chronic Diseases—ACCDiS, Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile, Santiago, Chile*

Gioacchino P. Marceca • *Department of Cancer Biology and Genetics, Comprehensive Cancer Center, The Ohio State University, Columbus, OH, USA; Department of Clinical and Experimental Medicine, University of Catania, Catania, CT, Italy*

Stephan Marquardt • *Institute of Experimental Gene Therapy and Cancer Research, Rostock University Medical Center, Rostock, Germany*

Annalisa Marsico • *Max-Planck-Institut für molekulare Genetik, Berlin, Germany; Institut für Informatik, Freie Universität Berlin, Berlin, Germany*

Andrea De Martino • *Soft & Living Matter Lab, CNR-NANOTEC, Rome, Italy; Italian Institute for Genomic Medicine, Turin, Italy*

Araks Martirosyan • *Laboratory of Glia Biology, VIB-KU Leuven Center for Brain and Disease Research, Leuven, Belgium; KU Leuven Department of Neuroscience, Leuven, Belgium*

Samatha Mathew • *Genomics and Molecular Medicine, CSIR Institute of Genomics and Integrative Biology (CSIR-IGIB), Delhi, India; Academy of Scientific and Innovative Research (AcSIR), New Delhi, India*

Isha Monga • *Bioinformatics Centre, Institute of Microbial Technology, Council of Scientific and Industrial Research, Chandigarh, India*

Jose C. Nacher • *Faculty of Science, Department of Information Science, Toho University, Chiba, Japan*

Serge Patrick Nana-Sinkam • *Division of Pulmonary Diseases and Critical Care Medicine, Virginia Commonwealth University, Richmond, VA, USA*

Giovanni Nigita • *Department of Cancer Biology and Genetics, Comprehensive Cancer Center, The Ohio State University, Columbus, OH, USA*

José Nelson Onuchic • *Center for Theoretical Biological Physics, Rice University, Houston, TX, USA; Department of Chemistry, Rice University, Houston, TX, USA; Department of Physics and Astronomy, Rice University, Houston, TX, USA; Department of Biosciences, Rice University, Houston, TX, USA*

Andrea Pagnani • *DISAT, Politecnico di Torino, Turin, Italy; Italian Institute for Genomic Medicine, Turin, Italy*

Alexandre Rossi Paschoal • *Department of Computer Science, Bioinformatics Graduation Program (PPGBIOINFO), Federal University of Technology—Paraná, Cornélio Procópio, Brazil*

Sushmita Paul • *Department of Bioscience & Bioengineering, Indian Institute of Technology Jodhpur, Jodhpur, Rajasthan, India*

Rosario Michael Piro • *Institut für Informatik, Freie Universität Berlin, Berlin, Germany; Institut für Medizinische Genetik und Humangenetik, Charite-Universitätsmedizin Berlin, Berlin, Germany*
Brigitte M. Pützer • *Institute of Experimental Gene Therapy and Cancer Research, Rostock University Medical Center, Rostock, Germany*
Jomar F. Rabajante • *Institute of Mathematical Sciences and Physics, University of the Philippines Los Baños, Laguna, Philippines*
Giulia Romano • *Division of Pulmonary Diseases and Critical Care Medicine, Virginia Commonwealth University, Richmond, VA, USA*
Ricardo C. H. del Rosario • *Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA*
Müşerref Duygu Saçar Demirci • *Faculty of Life and Natural Sciences, Department of Bioinformatics, Abdullah Gul University, Kayseri, Turkey*
Vinod Scaria • *G.N. Ramachandran Knowledge Center for Bioinformatics, CSIR Institute of Genomics and Integrative Biology (CSIR-IGIB), Delhi, India; Academy of Scientific and Innovative Research (AcSIR), CSIR Institute of Genomics and Integrative Biology (CSIR-IGIB), Delhi, India*
Bernd Schmeck • *Institute for Lung Research, Universities of Giessen and Marburg Lung Center, Member of the German Center for Lung Research (DZL), Philipps-University Marburg, Marburg, Germany*
Leon N. Schulte • *Institute for Lung Research, Universities of Giessen and Marburg Lung Center, Member of the German Center for Lung Research (DZL), Philipps-University Marburg, Marburg, Germany*
Paras Sehgal • *Academy of Scientific and Innovative Research (AcSIR), CSIR Institute of Genomics and Integrative Biology (CSIR-IGIB), Delhi, India; Genomics and Molecular Medicine, CSIR Institute of Genomics and Integrative Biology (CSIR-IGIB), Delhi, India*
Disha Sharma • *G.N. Ramachandran Knowledge Center for Bioinformatics, CSIR Institute of Genomics and Integrative Biology (CSIR-IGIB), Delhi, India; Academy of Scientific and Innovative Research (AcSIR), CSIR Institute of Genomics and Integrative Biology (CSIR-IGIB), Delhi, India*
Ambily Sivadas • *Academy of Scientific and Innovative Research (AcSIR), New Delhi, India; G.N. Ramachandran Knowledge Center for Bioinformatics, CSIR Institute of Genomics and Integrative Biology (CSIR-IGIB), Delhi, India*
Sridhar Sivasubbu • *Academy of Scientific and Innovative Research (AcSIR), CSIR Institute of Genomics and Integrative Biology (CSIR-IGIB), Delhi, India; Genomics and Molecular Medicine, CSIR Institute of Genomics and Integrative Biology (CSIR-IGIB), Delhi, India*
Christina Stielow • *Institute for Lung Research, Universities of Giessen and Marburg Lung Center, Member of the German Center for Lung Research (DZL), Philipps-University Marburg, Marburg, Germany*
Xiao-Jun Tian • *School of Biological and Health Systems Engineering, Arizona State University, Tempe, AZ, USA*
Luisa Tomasello • *Department of Cancer Biology and Genetics, Comprehensive Cancer Center, The Ohio State University, Columbus, OH, USA*
Manuela Vanegas Ferro • *School of Biological and Health Systems Engineering, Arizona State University, Tempe, AZ, USA*

Shamsudheen K. Vellarikkal • *Genomics and Molecular Medicine, CSIR Institute of Genomics and Integrative Biology (CSIR-IGIB), Delhi, India; Academy of Scientific and Innovative Research (AcSIR), New Delhi, India*

Dario Veneziano • *Department of Cancer Biology and Genetics, Comprehensive Cancer Center, The Ohio State University, Columbus, OH, USA*

Markus Wolfien • *Department of Systems Biology and Bioinformatics, Institute of Computer Science, University of Rostock, Rostock, Germany*

Olaf Wolkenhauer • *Department of Systems Biology and Bioinformatics, Institute of Computer Science, University of Rostock, Rostock, Germany; Stellenbosch Institute for Advanced Study (STIAS), Wallenberg Research Centre, Stellenbosch University, Stellenbosch, South Africa*

Malik Yousef • *Department of Community Information Systems, Zefat Academic College, Zefat, Israel*

Part I

Overview of Disease-Specific ncRNAs

Chapter 1

ncRNAs in Inflammatory and Infectious Diseases

Leon N. Schulte, Wilhelm Bertrams, Christina Stielow, and Bernd Schmeck

Abstract

Inflammatory and infectious diseases are among the main causes of morbidity and mortality worldwide. Inflammation is central to maintenance of organismal homeostasis upon infection, tissue damage, and malignancy. It occurs transiently in response to diverse stimuli (e.g., physical, radioactive, infective, pro-allergenic, or toxic), and in some cases may manifest itself in chronic diseases. To limit the potentially deleterious effects of acute or chronic inflammatory responses, complex transcriptional and posttranscriptional regulatory networks have evolved, often involving nonprotein-coding RNAs (ncRNA). MicroRNAs (miRNAs) are a class of posttranscriptional regulators that control mRNA translation and stability. Long ncRNAs (lncRNAs) are a very diverse group of transcripts >200 nt, functioning among others as scaffolds or decoys both in the nucleus and the cytoplasm. By now, it is well established that miRNAs and lncRNAs are implicated in all major cellular processes including control of cell death, proliferation, or metabolism. Extensive research over the last years furthermore revealed a fundamental role of ncRNAs in pathogen recognition and inflammatory responses. This chapter reviews and summarizes the current knowledge on regulatory ncRNA networks in infection and inflammation.

Key words miRNA, lncRNA, Infection, Inflammation, Immunity

1 Introduction

Gene expression is subjected to a vast variety of regulatory mechanisms taming transcriptional and translational noise and conferring adequate responses to cell-intrinsic and -extrinsic cues. While transcriptional and posttranscriptional regulation by protein factors has been a topic of research for a long time, the recent discovery of thousands of noncoding RNA genes in eukaryotic genomes has greatly expanded our understanding of gene expression control.

Ever since the discovery of microRNAs (miRNAs) as nonprotein-coding posttranscriptional regulators in eukaryotes in the early 1990s (*lin-4* in *Caenorhabditis elegans*, [1, 2]), the number of experimentally confirmed regulatory RNAs has continuously increased. In an attempt by the ENCODE project to map and classify the entire human transcriptome, thousands of previously

Xin Lai et al. (eds.), *Computational Biology of Non-Coding RNA: Methods and Protocols*, Methods in Molecular Biology, vol. 1912, https://doi.org/10.1007/978-1-4939-8982-9_1,

unknown long noncoding RNA genes were discovered [3]. While the miRNA biogenesis and target repression pathway is well understood to date, target identification is still a challenging venture. Gene regulation by miRNAs may reach a remarkable degree of network complexity: while one miRNA species may target different mRNAs through shared binding motives, typically in the 3′ UTR [4], a single mRNA may be targeted by multiple miRNAs at the same time. MiRNA-mediated control is complemented by long noncoding RNA-mediated transcriptional and posttranscriptional regulatory mechanisms, cumulating in highly complex gene-regulatory circuitries. This chapter summarizes the current knowledge about miRNAs and lncRNAs in mammalian inflammatory responses and infection.

2 MiRNAs

2.1 MiRNAs in Inflammation

For more than a decade, the role of miRNAs in hematopoietic cells has been in the focus of biomedical research. This research direction was pioneered by the David Bartel group, which reported on the role of miRNAs in the differentiation of hematopoietic progenitor cells in 2004 [5]. Ever since, the role of miRNAs in differentiation and function of immune cells has been the subject of countless research projects (reviewed in refs. [6–8]).

The innate immune system constitutes the first line of defense against pathogenic threats. The lung, with its huge surface area, is a primary site of pathogen attack and immunogen exposure due to its extensive contact with environmental particles, allergens, and airborne pathogens. Pattern recognition receptors (PRRs) on the cell surface (e.g., Toll-like receptors, TLRs) and in the cytosol (e.g., NOD-like receptors, NLRs, or RIG-like receptors, RLRs) serve as the sensors of the innate immune system, recognizing pathogen-associated molecular patterns (PAMPs), such as microbial cell wall components or bacterial or viral nucleic acids. Ligation of PAMPs to PRRs initiates an intracellular signaling cascade that culminates in the activation of specific transcription factors such as the inflammation master regulator NFκB. Activation of target genes, encoding protein factors such as pro-inflammatory cytokines, initiates a concerted local or systemic immune response. This response needs to be tightly controlled, to ensure successful pathogen clearance while preventing overshooting inflammation, which may cause serious side effects, such as severe tissue damage [9].

David Baltimore's group was the first to link PAMP recognition to expression changes of various miRNAs. Upon stimulation of THP-1 cells (a human acute monocytic leukemia cell line) with LPS from *Escherichia coli* [10], they observed increased expression of miR-146, -155, and -132. MiR-146a was identified as an immediate early-response gene, inducible by various microbial components and pro-inflammatory mediators (e.g., LPS, flagellin,