Advances in Experimental Medicine and Biology 799

Natalia Maltsev
Andrey Rzhetsky
T. Conrad Gilliam *Editors*

Systems Analysis of Human Multigene Disorders

Springer

Advances in Experimental Medicine and Biology

For further volumes:
http://www.springer.com/series/5584

Natalia Maltsev • Andrey Rzhetsky
T. Conrad Gilliam
Editors

Systems Analysis of Human Multigene Disorders

Springer

Editors
Natalia Maltsev
Department of Human Genetics
Institute for Genomics and Systems
Biology, Computation Institute
The University of Chicago
Chicago, IL, USA

T. Conrad Gilliam
Department of Human Genetics
Institute for Genomics and Systems
Biology, Computation Institute
The University of Chicago
Chicago, IL, USA

Andrey Rzhetsky
Deapartment of Medicine
Department of Human Genetics
Institute for Genomics and Systems
Biology, Computation Institute
The University of Chicago
Chicago, IL, USA

ISSN 0065-2598 ISSN 2214-8019 (electronic)
ISBN 978-1-4614-8777-7 ISBN 978-1-4614-8778-4 (eBook)
DOI 10.1007/978-1-4614-8778-4
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2013952921

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Understanding the genetic architecture underlying complex multigene disorders is one of the major goals of human genetics in the upcoming decades. Advances in whole genome sequencing and the success of high-throughput functional genomics allow supplementing conventional reductionist biology with systems-level approaches to human heredity and health as systems of interacting genetic, epigenetic, and environmental factors. This integrative approach holds the promise of unveiling yet unexplored levels of molecular organization and biological complexity. It may also hold the key to deciphering the multigene patterns of disease inheritance. Studies by countless groups have identified genes associated with many rare single gene (Mendelian) developmental disorders, but only limited progress has been made in finding the underlying causes for autism, schizophrenia, diabetes, and predisposition to cardiovascular disease, as they display complex patterns of inheritance and may result from many genetic variations, each contributing only weak effects to the disease phenotype. A major challenge to the detection and analysis of heritable patterns of disease susceptibility is the exponentially expanding search space required to explore all combinations of *m* genes or *m* genetic loci. Even the largest studies in human genetics are limited to the observation of several thousand meiotic events (i.e., the number of occasions in which the transmission of a given genetic variant from parents to offspring can be evaluated). Consequently, an exhaustive combinatorial search of even very small sets of multiple genetic loci leads to a huge burden of false-positive signals for every true-positive signal. This is because the number of statistical tests of significance performed on the same data set becomes too large to retain any statistical power. A biologically grounded approach is needed to constrain the plausible combinations of genomic regions that must be tested, drastically reducing the number of statistical tests.

The second obstacle to detecting multigenic inheritance is the need to understand relationships among the set of genes related to a disease and to determine how variations within each gene affect disease susceptibility. The influence of such diverse genetic interactions remains unknown. Genetic variations across multiple interacting genes may affect the phenotype in a linear (additive) or nonlinear (epistatic) manner. Groups of interacting genes are likely to affect disease phenotypes via as-yet-unknown mixtures of both types of interaction. Furthermore, disease susceptibility may increase incrementally with increasing genetic variation or dichotomously via a threshold effect. Finally, the genetic causes of a given disorder may differ, in whole or in part, in different affected families. Although it is tempting to test the entire spectrum of inheritance models, this is currently impractical. The total number of possible models of inheritance involving *m* genetically interacting genes grows exponentially with *m*, further amplifying the exponential growth of the number of distinct gene sets of size *m*. A biology-grounded plan of prioritizing genetic models by their likelihood and systematic analysis of model space is critically important.

The third obstacle is that it is extremely hard and expensive to design large-scale studies that account for interactions between environmental factors and genetic variation in relation to disease phenotypes; a typical large-scale genetic analysis avoids explicit modeling and/or extensive measuring of environmental factors.

The scientific community has made enormous investments in developing the scientific infrastructure necessary to enable breakthrough discoveries of the primary biological risk factors for common disorders, such as diabetes, autism, susceptibility to cardiovascular diseases, and cancer. These investments have made possible investigations to understand disease-associated risk factors, on a scale unpredictable even a few years ago. Studies such as those based on genome-wide association have become standard and have led to a substantial number of discoveries. Although progress has been made in understanding some of these complex traits, our grasp of the patterns of risk are reduced to simple, short lists of weakly associated, noninteracting genetic variants that explain only a very low percentage of the estimated heritability. Some other challenges in constructing disease risk models are as follows: multigenic models of inheritance are usually ignored; genetic heterogeneity of commonly investigated phenotypes can lead to inefficient studies; and the wealth of information available on the biological system is generally ignored in constructing models of disease risk.

The volume is structured to introduce the major perspectives on intellectual and technological challenges facing systems-level translational medicine.

Chapter 1 addresses the need for the integration of clinical and genomic profiling with preventative healthcare. In recent years it became exceedingly clear that genotypes alone are insufficient to predict health outcomes, since they fail to account for individualized responses to the environment and life history. Integrative genomic approaches incorporating whole genome sequencing, transcriptomics, and epigenomics should be combined with clinical interpretation in the light of the triggers, behaviors, and environment unique to each person. Such integration will allow for an accurate prediction of the disease progression for a particular patient and significant improvement of personalized treatment strategies. The chapter discusses some of the major obstacles to implementation of such an approach, from development of risk scores through integration of diverse omic data types, to presentation of results in a format that fosters development of personal health action plans.

Chapter 2 provides a comprehensive review of high-throughput data generation technologies employed by high-throughput systems-level biomedical studies with the emphasis on the next generation DNA sequencing platforms (NGS). NGS provides an inexpensive and scalable approach for detection of the molecular changes at the genetic, epigenetic, and transcriptional level. Furthermore, existing and developing single molecule sequencing platforms will soon allow direct RNA and protein measurements, thus increasing the specificity of current assays and making it possible to better characterize "epi-alterations" that occur in the epigenome and epitranscriptome. The authors describe novel approaches for generation and processing of genomic data, the development of the integrative models, and the increasing ubiquity of self-reporting and self-measured genomics and health data.

Chapter 3 addresses the challenges and best practices of high-throughput integrative medicine. Efficient mining of vast and complex data sets for the needs of biomedical research critically depends on seamless integration of clinical, genomic, and experimental information with prior knowledge about genotype–phenotype relationships accumulated in a plethora of publicly available databases. Furthermore, such experimental data should be accessible to a variety of algorithms and analytical pipelines that drive computational analysis and data mining. Translational projects require sophisticated approaches that coordinate and perform various analytical steps involved in extraction of useful knowledge from accumulated clinical and experimental data in an orderly semi-automated manner. The chapter explores cross-cutting requirements from multiple translational projects for data integration, management, and analysis.

Chapter 4 describes the algorithmic approaches for selecting and prioritizing disease candidate genes. The authors review the prioritization criteria and the algorithms along with some use cases that demonstrate how these tools can be used for identifying and ranking human disease candidate genes.

Chapter 5 presents a clinical perspective on systems-level translational research using lung cancer as an example. Lung cancer is no longer considered a single disease entity and is now being subdivided into molecular subtypes with dedicated targeted and chemotherapeutic strategies. The concept of using information from a patient's tumor to make therapeutic and treatment decisions has revolutionized the landscape for cancer care and research in general. Future directions will involve incorporation of molecular characteristics and next generation sequencing into screening strategies to improve early detection, while also having applications for joint treatment decision-making in the clinics with patients and practitioners.

This volume targets the readers who wish to learn about state-of-the-art approaches for systems-level analysis of complex human disorders.

The audience may range from graduate students embarking upon a research project to practicing biologists and clinicians working on systems biology of complex disorders and to bioinformaticians developing advanced databases, analytical tools, and integrative systems. With its interdisciplinary nature, this volume is expected to find a broad audience in pharmaceutical companies and in various academic departments in biological and medical sciences (such as molecular biology, genomics, systems biology, and clinical departments) and computational sciences and engineering (such as bioinformatics and computational biology, computer science, and biomedical engineering).

We thank all the authors and coauthors who have contributed to this volume. We would like to extend our thanks to Melanie Tucker and Meredith Clinton of Springer US for their help in the preparation of this book.

Chicago, IL

Natalia Maltsev
Andrey Rzhetsky
T. Conrad Gilliam

Contents

Contributors

Gady Agam Department of Computer Science, Illinois Institute of Technology, Chicago, IL, USA

Bruce J. Aronow Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

Sandhya Balasubramanian Department of Human Genetics, University of Chicago, Chicago, IL, USA

Rishi Batra 987400 Nebraska Medical Center, University of Nebraska Medical Center, Omaha, NE, USA

Eduardo Berrocal Department of Human Genetics, University of Chicago, Chicago, IL, USA

Department of Computer Science, Illinois Institute of Technology, Chicago, IL, USA

Daniela Börnigen Department of Human Genetics, University of Chicago, Chicago, IL, USA

Toyota Technological Institute at Chicago, Chicago, IL, USA

Apoorva Chawla Department of Medicine, Section of Hematology/Oncology, University of Chicago, Chicago, IL, USA

Bhadrachalam Chitturi Department of Computer Science, Amrita Vishwa Vidyapeetham University, Amritapuri Campus, Kollam, Kerala, India

Utpal Dave Computation Institute, University of Chicago/Argonne National Laboratory, Chicago, IL, USA

Inna Dubchak Genomics Division, Berkley National Laboratory, Walnut Creek, CA, USA

Bo Feng Department of Human Genetics, University of Chicago, Chicago, IL, USA

Department of Computer Science, Illinois Institute of Technology, Chicago, IL, USA

Greg Gibson Georgia Institute of Technology, Predictive Health Institute and School of Biology, Atlanta, GA, USA

T. Conrad Gilliam Department of Human Genetics, Institute for Genomics and Systems Biology, Computation Institute, The University of Chicago, Chicago, IL, USA

Thomas Hensing Department of Medicine, Section of Hematology/Oncology, NorthShore University Health System, Evanston, IL, USA

Department of Medicine, Section of Hematology/Oncology, University of Chicago, Chicago, IL, USA

Anil G. Jegga Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

Natalia Maltsev Department of Human Genetics, Institute for Genomics and Systems Biology, Computation Institute, The University of Chicago, Chicago, IL, USA

Christopher E. Mason Department of Physiology and Biophysics, Weill Cornell Medical College, New York, NY, USA

The HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medical College, New York, NY, USA

Sandra G. Porter Digital World Biology, Seattle, WA, USA

Ravi Salgia Department of Medicine, Section of Hematology/Oncology, University of Chicago, Chicago, IL, USA

Todd M. Smith PerkinElmer, Seattle, WA, USA

Dinanath Sulakhe Computation Institute, University of Chicago/Argonne National Laboratory, Chicago, IL, USA

Andrew Taylor Department of Human Genetics, University of Chicago, Chicago, IL, USA

Chao Wu Department of Computer Science, College of Engineering and Applied Science, University of Cincinnati, Cincinnati, OH, USA

Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Bingqing Xie Department of Human Genetics, University of Chicago, Chicago, IL, USA

Department of Computer Science, Illinois Institute of Technology, Chicago, IL, USA

Jinbo Xu Toyota Technological Institute at Chicago, Chicago, IL, USA

Cheng Zhu Department of Computer Science, College of Engineering and Applied Science, University of Cincinnati, Cincinnati, OH, USA

Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA