Translational Bioinformatics

Series Editor

Xiangdong Wang, MD, PhD
Professor of Clinical Bioinformatics, Lund University, Sweden
Professor of Medicine, Fudan University, China

Aims and Scope

The Book Series in Translational Bioinformatics is a powerful and integrative resource for understanding and translating discoveries and advances of genomic, transcriptomic, proteomic and bioinformatic technologies into the study of human diseases. The Series represents leading global opinions on the translation of bioinformatics sciences into both the clinical setting and descriptions to medical informatics. It presents the critical evidence to further understand the molecular mechanisms underlying organ or cell dysfunctions in human diseases, the results of genomic, transcriptomic, proteomic and bioinformatic studies from human tissues dedicated to the discovery and validation of diagnostic and prognostic disease biomarkers, essential information on the identification and validation of novel drug targets and the application of tissue genomics, transcriptomics, proteomics and bioinformatics in drug efficacy and toxicity in clinical research.

The Book Series in Translational Bioinformatics focuses on outstanding articles/chapters presenting significant recent works in genomic, transcriptomic, proteomic and bioinformatic profiles related to human organ or cell dysfunctions and clinical findings. The Series includes bioinformatics-driven molecular and cellular disease mechanisms, the understanding of human diseases and the improvement of patient prognoses. Additionally, it provides practical and useful study insights into and protocols of design and methodology.

Series Description

Translational bioinformatics is defined as the development of storage-related, analytic, and interpretive methods to optimize the transformation of increasingly voluminous biomedical data, and genomic data in particular, into proactive, predictive, preventive, and participatory health. Translational bioinformatics includes research on the development of novel techniques for the integration of biological and clinical data and the evolution of clinical informatics methodology to encompass biological observations. The end product of translational bioinformatics is the newly found knowledge from these integrative efforts that can be disseminated to a variety of stakeholders including biomedical scientists, clinicians, and patients. Issues related to database management, administration, or policy will be coordinated through the clinical research informatics domain. Analytic, storage-related, and interpretive methods should be used to improve predictions, early diagnostics, severity monitoring, therapeutic effects, and the prognosis of human diseases.

Translational Bioinformatics

Series Editor: Xiangdong Wang, MD, PhD, Professor of Clinical Bioinformatics,
Lund University, Sweden;
Professor of Medicine, Fudan University, China

Recently Published and Forthcoming Volumes

Applied Computational Genomics
Editor: Yin Yao Shugart
Volume 1

Pediatric Biomedical Informatics
Editor: John J. Hutton
Volume 2

Bioinformatics of Human Proteomics
Editor: Xiangdong Wang
Volume 3

For further volumes:
http://www.springer.com/series/11057

Translational Bioinformatics
Volume 2

Pediatric Biomedical Informatics

Computer Applications in Pediatric Research

Editor: John J. Hutton

Springer

Editor
John J. Hutton
Cincinnati Children's Hospital
Cincinnati Ohio
USA

ISSN 2213-2775 ISSN 2213-2783 (electronic)
ISBN 978-94-007-5148-4 ISBN 978-94-007-5149-1 (eBook)
DOI 10.1007/978-94-007-5149-1
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2012954669

Printed on acid-free paper

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

Series Foreword

Translating biomedical discoveries into better health for people continues to pose major and well-recognized challenges, although the pace of those discoveries over past decades has been truly startling. Among these challenges is analyzing the huge amounts of data generated by high throughput technologies and integrating it with data about patients that is now accessible through electronic health records. Converting data into knowledge is one aim of biomedical informatics.

In recognition of the major role informatics plays in accruing, integrating, and analyzing data in the biomedical sciences and translating it into clinical practice, a series of books on *Translational Bioinformatics* is being created by united scientific forces of the *International Society for Translational Medicine* (ISTM, www.istmed.org), *Journal of Clinical Bioinformatics* (JCBi, www.jclinbioinformatics.com), *Journal of Clinical and Translational Medicine* (CTM, www.clintransmed.com), and Springer Publisher. These will cover topics such as genomics, proteomics, metabolomics, systems immunology, and biomarkers. *Pediatric Biomedical Informatics: Computer Applications in Pediatric Research* is an important volume in this series and focuses on core resources in informatics that are necessary to support translational research in a research-intensive children's medical center. One key challenge is implementing interoperable research and clinical IT systems so that data can be exchanged to support translational research.

I, as the Editor of Series Books, am privileged and honored to have Professor John J. Hutton as the Editor of this special volume. I knew John by his outstanding contributions to biomedical informatics in clinical and translational medicine and his reputation as an academic manager and leader, as well as international and national opinion leader. The most important issue for me and other CTM fellows is his friendship, professional supervision, or endless support.

The first part of this book discusses important technical challenges that underlie most types of computer-based biomedical research, e.g. storing, mining, and transmitting huge quantities of data (giga-, tera-, and sometimes petabytes), while maintaining strict cybersecurity with identity management and access control. The second part discusses a broad range of research programs and demonstrates the wide variety of software applications that are necessary to support relevant types

of research. While different applications are needed to support different types of translational research, they all have in common the need to be interfaced with a well-designed hardware infrastructure. One of the most outstanding parts is that the book emphasizes new opportunities to investigate clinical phenomena through the use of data in electronic health records.

As has been emphasized by the US National Institutes of Health, the bench-to-bedside approach to translational research is really a two-way street. While translational researchers provide clinicians with new resources for prevention, diagnosis, and treatment of disease, it is the clinicians themselves who interface directly with patients and observe the onset and course of disease. Feedback from the clinicians provides essential information to the researchers and can provide new insights into ill understood elements of disease that are topics for future research.

Pediatric Biomedical Informatics: Computer Applications in Pediatric Research, as part of the Springer Series on Translational Bioinformatics, provides excellent and useful guidance to the implementation of an informatics infrastructure to support translational research in a research-intensive clinical environment. It shows how a wide variety of active research programs use this infrastructure to generate knowledge to improve human health.

Professor of Respiratory Medicine and Director of Biomedical Research Center
Fudan University, Zhongshan Hospital, Shanghai, China;
Professor (adj) of Clinical Bioinformatics,
Lund University Clinical Science, Lund, Sweden

Xiangdong Wang M.D., Ph.D.

Foreword

The Revolution in Biomedical Informatics

I was born before electronic computers existed. Therefore, I have witnessed the entire Revolution in Biomedical Informatics over the last 40 years. As a pediatric physician-scientist and a practicing pediatric cardiologist, for many years my goal was to generate data: experimental results in the laboratory to understand the structure and functions of proteins and genes, and observational and laboratory information to optimize care of patients. Generating sufficient data, cataloging the data rationally, and analyzing the limited data were my daily activities. And, there was never enough data. Until now.

Now, all of us working in medicine and science are deluged with data. Other descriptors of this Revolution of enormous information that I have heard are "avalanche" and "tsunami", as we are inundated with data. The goals for pediatricians and scientists have changed, moving beyond data creation. Although we must continue to generate data, the Revolution in Biomedical Informatics has created major challenges in the acquisition, management, storage, transmission, and analysis of data essential to accomplish our ultimate purposes of optimizing patient care, determining mechanisms underlying disease, and defining the best treatments for children.

I was fortunate to become the Chief Medical Officer at Cincinnati Children's Hospital Medical Center and Director of our Research Foundation in 2007, just as our leadership initiated the implementation of a single electronic medical or health record for all clinical care. We recognized the incredible potential of using the EMR data to assess and improve patient safety, to ascertain outcomes of our patients, and to serve as the basis for patient-oriented, clinical, translational, and health services research on a global scale. Under the guidance and tutelage of John Hutton, the Editor of this volume, and with the recruitment and expertise of the faculty and their colleagues who have contributed chapters, we have begun the journey toward maximizing this potential. The topics discussed include the major core informatics resources needed, from the EMR itself, to transmission

of information in it for storage and management, to security required for patient protection, creation of usable patient data warehouses, and integration of patient information with biobanked tissue and DNA for research, all critical infrastructure to optimize care and research. In addition, some intriguing applications in both patient-oriented research and basic science are provided, to illustrate how population-based studies, assessment of language, support for decisions, and generation of networks can be done.

Biomedical informatics is the key to our future, as we integrate clinical care, genomics, and basic science to improve outcomes and discover new therapeutics. With careful design and acquisition of information, we can tame the avalanche of data. We can link data across institutions to achieve greater power of analysis and increase the speed of discovery and evaluation of treatments. This book provides insight into how to use data generated during The Revolution in Biomedical Informatics to benefit children around our world.

Arnold W. Strauss, M.D.
Rachford Professor and Chair
Department of Pediatrics
University of Cincinnati College of Medicine
Director, Cincinnati Children's Research Foundation
Chief Medical Officer
Cincinnati Children's Hospital Medical Center

Preface

Cincinnati Children's Hospital Medical Center (CCHMC) is a major teaching affiliate of the University of Cincinnati College of Medicine. Its clinical and research programs in pediatrics rank among the largest and most highly respected in the United States. A decade ago, the Board of Trustees and leadership of CCHMC envisioned a future in which advances in information technology (IT) would radically change the infrastructure needed to support clinical care and biomedical research. The faculty developed and the leadership supported a strategic plan for information technology. Since then, Hospital Information Services has implemented an electronic health record to support clinical care throughout the medical center and a Division of Biomedical Informatics has been created to support research. The two units (Information Services and Biomedical Informatics) work very closely together to assure seamless support of medical center clinical and translational research programs.

Moving the institution from the past to the present has been an adventure. We have continuously assessed what core resources in informatics are necessary to support research programs and how these can best be integrated with hospital systems to receive clinical information that is necessary to conduct translational research in pediatrics. This book focuses on the core informatics resources that are now available to our faculty and provides examples of how they support specific types of research. While the focus is upon core resources in IT and computer applications in one research-intensive pediatric medical center, CCHMC is large and diverse so that all or portions of the lessons learned and services developed will be applicable to other institutions. Most of the contributors either are experts in the planning and implementation of IT services or are active users who participate in the assessment of needs of researchers now and in the future.

The first part of the book discusses core resources in informatics, both technical and knowledge-based, that are relevant to most translational research projects. Important technical challenges underlying computer-based biomedical research are storing, mining, and transmitting huge quantities of data (terabytes and sometimes petabytes), while maintaining strict cybersecurity with identity management and access control. Another topic of broad importance is the electronic health record

(EHR) and how information captured in the EHR, during the process of clinical care, can be moved into a research patient data warehouse. Data in the warehouse can then be mined to support research, including the identification of cohorts of patients for study of a wide range of diseases. These important topics are discussed in considerable detail.

Biosamples, such as blood and tissues, are key ingredients in biomedical research and can be collected in enterprise biobanks. The information management system of the biobank can be interfaced with the research patient data warehouse so that phenotypes of patients, who contributed the biosamples, can be derived. This is critically important when conducting genetic studies, where genomes of patients with specific phenotypes will be sequenced to identify variant genes associated with disease. Chapters devoted to biobanking, data warehousing, and privacy also discuss the very important topics of patient consent and ethical and legal issues surrounding human subjects research.

The second part of the book discusses a broad range of research programs supported by core informatics resources. The first example focuses on EHR-linked registries for studies of populations and how such registries can be federated to permit data sharing among members of a distributed research network. This is particularly important in pediatrics because most pediatric chronic conditions are rare. Frequently, no single center has a sufficient number of patients to support outcomes, comparative effectiveness, and other types of clinical research. Natural language processing (NLP) is a topic of increasing relevance and is the subject of two chapters. It is apparent that much of the information in EHRs is in free text, such as clinical notes by physicians and other health care professionals. While NLP is an active field of research, we have found it essential as a core capability to support patient phenotyping, prospective identification of safety events, and mining of publications for specific types of information to be incorporated into decision support applications. Perinatal and neonatal research are unique and critically important to pediatrics. They pose challenges in information management, as presented in a chapter addressing support of perinatal and neonatal research. Neonatal terminologies are not well developed. It is generally difficult to collect and integrate information, as infants transition from obstetrical to pediatric care. Other important topics in clinical research that require computer support and are discussed include decision support, calculation of outcomes measures, and automated detection and reporting of safety events.

Six chapters provide examples of informatics applications in translational research at the interface between basic science and the clinic. The first topic is network analysis of complex molecular interactions and biological processes to identify the underlying causes of pediatric disease and to develop innovative therapies. Such analyses are being applied, for example, to drug discovery and drug repurposing, i.e., discovering new targets for old drugs developed for some other purpose.

Developmental biology is particularly relevant to pediatrics and is a discipline that is well represented among research programs in our institution. One chapter is devoted to transcriptional networks that control lung maturation. Maturation of the lung is a late event during fetal development, yet is key to survival of the newborn

infant. Several decades ago, respiratory distress syndrome (RDS) was the major cause of morbidity and mortality of premature infants. Death from RDS is rare today because of research on lung maturation, which led to the development of antenatal glucocorticoid treatment and surfactant replacement therapy to prevent and treat it. A second chapter discusses how software applications and other tools can be applied to studies of the sequential action of genes during kidney and craniofacial development, i.e., define their genetic blueprint. Both chapters illustrate how software applications and computer-support permit studies of gene regulation in networks with consequent cell and organ development. Without tools of biomedical informatics, these complex, data-intensive studies could not be done.

The final chapters discuss informatics tools and applications that support genetics and genomics. Focus is upon studies of genetic variation and its effects upon human health. One chapter discusses single nucleotide polymorphism (SNP) genotyping and its integration into clinical information systems; the other discusses genome and exome sequencing to identify sequence variants that cause disease. In both cases, management of huge data sets, frequently requiring use of high performance computing, is a major informatics challenge.

Support of biomedical research in a pediatric medical center requires both a sophisticated IT infrastructure and a staff of knowledgeable professionals. One of the most important lessons we have learned during the process of building informatics support of clinical and translational research is that the information systems that support clinical care and those that support research must constantly communicate and share information. This is both a political and a technological challenge (staff from different organizational cultures must cooperate and to the extent possible, IT systems must be semantically and syntactically interoperable). These challenges can be overcome, as we hope our book demonstrates.

John J. Hutton, M.D.

Contents

Contributors

Bruce J. Aronow Ph.D. Department Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

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

Dominik Aronsky M.D., Ph.D., FACMI Departments of Biomedical Informatics and Emergency Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA

Michael G. Barnes Ph.D. Division of Rheumatology, Cincinnati BioBank Core Facility, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Jacek W. Biesiada Ph.D. Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Division of Management and Informatics, Technical University of Silesia, Gliwice, Poland

Eric W. Brunskill Ph.D. Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Ye Chen School of Electronics and Computing Systems, University of Cincinnati College of Engineering & Applied Science, Cincinnati, OH, USA

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

Jeremy J. Corsmo MPH, CIP, CHRC Research Compliance and Regulatory Affairs, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Louise Deleger Ph.D. Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Jingyuan Deng School of Computing Sciences and Informatics, University of Cincinnati College of Engineering & Applied Science, Cincinnati, OH, USA

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

Judith W. Dexheimer Ph.D. Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

Divisions of Emergency Medicine and Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Phillip Dexheimer M.S. Divisions of Biomedical Informatics and Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Gregory A. Grabowski M.D. Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Eric S. Hall Ph.D. Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

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

Nicholas Hunt MSIA, CISSP Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

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

John J. Hutton M.D., FACP Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

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

Anil G. Jegga DVM, M.S. Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

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

Laurie H. Johnson M.D. Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

Division of Emergency Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Eric S. Kirkendall M.D., MBI, FAAP Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

Divisions of Hospital Medicine and Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Michal Kouril Ph.D. Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

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

Long Jason Lu Ph.D. Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

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

John A. Lynch Ph.D. Department of Communication, McMicken College of Arts & Sciences, University of Cincinnati, Cincinnati, OH, USA

Center for Clinical and Translational Science and Training, University of Cincinnati College of Medicine, Cincinnati, OH, USA

Keith Marsolo Ph.D. Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

Jaroslaw Meller Ph.D. Department of Environmental Health, University of Cincinnati College of Medicine, Cincinnati, OH, USA

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

John P. Pestian Ph.D. Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

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

Andrew S. Potter M.S. Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

S. Steven Potter Ph.D. Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Senthilkumar Sadhasivam M.D., MPH Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

Perioperative and Acute Pain Service, Division of Pediatric Anesthesia, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Guergana K. Savova Ph.D. Boston Children's Hospital and Harvard Medical School, Harvard University, Boston, MA, USA

Imre Solti M.D., Ph.D., M.A. Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

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

S. Andrew Spooner M.D., M.S., FAAP Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

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

Paul E. Steele M.D. Departments of Pathology and Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

Clinical Laboratory, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Arnold W. Strauss M.D. Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

Cincinnati Children's Research Foundation, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Lirong Tan School of Computing Sciences and Informatics, University of Cincinnati College of Engineering & Applied Science, Cincinnati, OH, USA

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

Michael Wagner Ph.D. Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

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

Jeffrey A. Whitsett M.D. Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

Perinatal Institute, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

David P. Witte M.D. Departments of Pathology and Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

Division of Pathology and Laboratory Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Yan Xu Ph.D. Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

Divisions of Pulmonary Biology and Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Minlu Zhang Ph.D. School of Computing Sciences and Informatics, University of Cincinnati College of Engineering & Applied Science, Cincinnati, OH, USA

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

Cheng Zhu M.S. School of Computing Sciences and Informatics, University of Cincinnati College of Engineering & Applied Science, Cincinnati, OH, USA