CAMBRIDGE

Cambridge University Press
978-0-521-76800-9 - Bioinformatics and Computational Biology in Drug Discovery and Development
Edited by William T. Loging
Frontmatter
More information

Bioinformatics and Computational Biology in Drug Discovery and Development

Computational biology drives discovery through its use of high-throughput informatics approaches. This book provides a road map of the current drug development process, and how computational biology approaches play a critical role across the entire drug discovery pipeline.

Through the use of previously unpublished, real-life case studies, the impact of a range of computational approaches are discussed at various phases of the pipeline. Additionally, a focus section provides innovative visualization approaches, from both the drug discovery process as well as from other fields that utilize large data sets, recognizing the increasing use of such technology.

Serving the needs of early career and more experienced scientists, this up-to-date reference provides an essential introduction to the process and background of drug discovery, highlighting how computational researchers can contribute to that pipeline.

William T. Loging is Associate Professor of Genetics and Genomic Sciences, and Director of Production Bioinformatics at the Icahn School of Medicine at Mount Sinai, New York. An award-winning expert who has worked on many successful drug development projects, his current research focuses on generating treatments for diseases, with patents secured for his work in both oncology and immunology.

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76800-9 - Bioinformatics and Computational Biology in Drug Discovery and Development
Edited by William T. Loging
Frontmatter
More information

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76800-9 - Bioinformatics and Computational Biology in Drug Discovery and Development
Edited by William T. Loging
Frontmatter
More information

Bioinformatics and Computational Biology in Drug Discovery and Development

Edited by

William T. Loging, PhD

Associate Professor of Genetics and Genomic Sciences
Director of Production Bioinformatics
Icahn School of Medicine at Mount Sinai
New York, NY

CAMBRIDGE
UNIVERSITY PRESS

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76800-9 - Bioinformatics and Computational Biology in Drug Discovery and Development
Edited by William T. Loging
Frontmatter
More information

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9780521768009

© Cambridge University Press 2016

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2016

Printed in the United Kingdom by TJ International Ltd. Padstow Cornwall

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data
Name: Loging, William T., editor.
Title: Bioinformatics and computational biology in drug discovery and development / edited by William T. Loging.
Description: Cambridge, United Kingdom; New York: Cambridge University Press, 2016. | Includes bibliographical references and index.
Identifiers: LCCN 2015041412 | ISBN 9780521768009 (hardback)
Subjects: | MESH: Drug Discovery–methods. | Biomarkers, Pharmacological. | Computational Biology–methods. | Data Mining–methods.
Classification: LCC RS420 | NLM QV 745 | DDC 615.1/9–dc23
LC record available at http://lccn.loc.gov/2015041412

ISBN 978-0-521-76800-9 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Every effort has been made in preparing this book to provide accurate and up-to-date information which is in accord with accepted standards and practice at the time of publication. Although case histories are drawn from actual cases, every effort has been made to disguise the identities of the individuals involved. Nevertheless, the authors, editors and publishers can make no warranties that the information contained herein is totally free from error, not least because clinical standards are constantly changing through research and regulation. The authors, editors and publishers therefore disclaim all liability for direct or consequential damages resulting from the use of material contained in this book. Readers are strongly advised to pay careful attention to information provided by the manufacturer of any drugs or equipment that they plan to use.

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76800-9 - Bioinformatics and Computational Biology in Drug Discovery and Development
Edited by William T. Loging
Frontmatter
More information

Contents

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76800-9 - Bioinformatics and Computational Biology in Drug Discovery and Development
Edited by William T. Loging
Frontmatter
More information

Contributors

Mark Crawford
LS Pharma – Retired

Thomas B. Freeman
Capella Biosciences, Inc.

Jeff Handler
JAH Associates LLC

Marilyn Lewis
Pfizer Inc.

William T. Loging
Icahn School of Medicine at Mount Sinai

Pek Lum
Capella Biosciences, Inc.

Roy Mansfield
Pfizer Inc.

Jonathan Phillips
Boehringer-Ingelheim Pharma Inc.

Raul Rodriguez-Esteban
Roche Inc.

Telmo Silva
ClicData RCS

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76800-9 - Bioinformatics and Computational Biology in Drug Discovery and Development
Edited by William T. Loging
Frontmatter
More information

Yirong Wang
Mount Sinai Genetic Testing Lab
Connecticut Icahn School of Medicine at Mount Sinai

Bryn Williams-Jones
Connected Discovery Ltd.

Ke Xu
Bristol-Myers-Squibb Inc.

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76800-9 - Bioinformatics and Computational Biology in Drug Discovery and Development
Edited by William T. Loging
Frontmatter
More information

Foreword: The future of drug discovery and healthcare

The era of big data is transforming healthcare

The world we currently live in is unprecedented with respect to the rate of technology expansion and application of these technologies across a hierarchy of health-related problems. Businesses are generating more information, acquiring more data, and aggregating more data on customers than ever before, with the hope that the appropriate analytics applied to these data can inform on all aspects of their product's performance and aid in evolving products and in better targeting of products to specific customer groups. Just as climatologists use big, comprehensive data to predict weather patterns, and quantitative traders on Wall Street use Big Data to assess when to buy and when to sell stocks, so the medical field will soon harness that same power of big data to better understand patient populations, when to treat, and how to treat. However, today, very large-scale data and the integrative analytics that are applied to such data hold no greater promise perhaps than for the healthcare industry. The application of high-throughput informatics technologies to biological problems exemplifies the disciplines of Computational Biology, Bioinformatics, and more recently Clinical Genomics. When one views recent achievements following the hard-won successes of the Human Genome Project and the amount of genomic data being generated on populations of individuals to better understand phenotypic variation in these populations (from disease to the evolutionary architecture of human populations), it is clear that computational tools, advanced algorithms, and a diversity of highly skilled personnel are required to make the most of these data in order to deliver on the promise of the post-genomic era: improved diagnosis, treatment, and prevention of disease.

Understanding disease in a drug discovery setting

Human health is the product of many interacting, deeply complex systems: a person's physiology, their genome and other molecular features, the state of many different types of cells in their body, of their organs, their physical environment, the bacteria and viruses in that environment, their activities, their social interactions, and their medical care. Those who will prevail in the future of developing the most efficacious and safe interventions to treat and prevent disease will be those who can

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76800-9 - Bioinformatics and Computational Biology in Drug Discovery and Development
Edited by William T. Loging
Frontmatter
More information

obtain the best data, perform the most informed analyses to identify the patterns associated with disease and wellness, and build the most predictive models based on those patterns – where others may see only noise.

The typical drug discovery process spans 10–15 years, requires investment of roughly one billion dollars, as well as a great diversity of skill sets mainly engineered around the idea of designing a single molecule or biologic to activate or suppress the activity of a single protein. Drug discovery researchers over the last many decades have adhered to an overly simplified view of disease that is very much in need of a revolution in mindset, experimentation, and in approaches and workflows. A large gap often exists between informatics scientists and biologists, which very often limits the potential for advancement of our understanding of complex disease processes. To achieve the ultimate success in developing the most efficacious and safe interventions to treat and prevent disease, an intimate partnership must exist between the informatics and biology experts. Understanding informatics approaches to Big Data is only part of the answer; the application of these approaches and how they interface to solve the problems of delivering drugs to market are critical components as well. Several chapters in this book illuminate this. Often, management and biologists are not aware of the need for advanced computational approaches that can consider millions of variables of information simultaneously to better inform decision making, what exactly the computational approaches can deliver, what is required to enable this delivery and the limitations of such approaches. In the future, the companies and organizations that successfully integrate computational biology approaches with basic biomedical research and drug development will be best positioned to succeed in the quest of improving the lives of patients and well individuals. A major goal of this book is to help bridge these divides in perspective and understanding.

Laboratory bench to patient bedside

The healthcare industry is rapidly changing – payers, payees, providers, patients, and consumers more generally are taking control of data for use in healthcare outcomes, disease assessment and understanding, prediction, as well as health and lifestyle choices. Successful companies will seek to understand how to use these data to make more informed, more accurate, more precise decisions regarding the health of the individual, delivering the right intervention for any disease condition to the right patient at the right time (Schadt *et al.*, 2005), what we now refer to as precision medicine. Only recently have we had the technology to collect very high-dimensional (hundreds of thousands to millions of features), diverse complex molecular data on each individual in a cost-effective manner, and to analyze these data to understand individual variation in disease processes and responses to drugs and even wellness, where identifying features that have enabled individuals to buffer disease, live longer, healthier lives, promises to have the highest impact on overall human wellness (Friend and Schadt, 2014).

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76800-9 - Bioinformatics and Computational Biology in Drug Discovery and Development
Edited by William T. Loging
Frontmatter
More information

Healthcare data are coming from a growing spectrum of sources: electronic medical records (including family medical history), imaging data (MRI, PET scans), genetic information (whole genome sequencing of DNA or whole transcriptome sequencing of RNA), metabolite, protein, lipid and sugar information generated from tissue samples, pathology reports (on tumors), blood tests, urine tests, and so on. However, one of the more dramatic Big Data generators to explode onto the scene over the past several years are the network-enabled wearable and implantable devices that empower individuals to monitor extensively their own health, with devices such as the Apple Watch, FitBit, Nike FuelBand, and HexoSkin enabling a wide array of measurements that can be collected longitudinally, from activity levels to heart rate, blood pressure, and pulse oximetry. Increasingly, these types of devices and mobile apps run on smartphones are generating continuous, longitudinal data on individuals that can provide a far more comprehensive assessment of an individual's health compared to the extremely limited snapshot that gets taken in the doctor's office perhaps once per year. In the not so distant future, more information relating to the health of individuals will exist outside of their doctor's office than inside. Those biopharmaceutical organizations who can make use of these data, in both a clinical as well as a preclinical setting, will be well-placed for success in the future. That level of success will be marked by the number of life-saving therapies that are brought to market or the number of interventions that increase overall well-being in the human population.

Given the wide range of data types and the sheer scale of these data, the challenge is how best to collect them, to organize them in systematic ways, to effectively manage them, compute on them and mine them, so that we can derive meaningful insights across vast numbers of patients that ultimately will inform on tailored treatment paths.

Adaptive drug trials will evolve the standard of care

To get to this future of precision medicine, we will need to transform the way in which interventions are clinically validated and approved by regulatory agencies. We will need to run the type of adaptive trials that can identify and validate a new standard of care as evidence accumulates. For example, as more and more data are generated on patients, these data will be stored in big databases across any number of medical centers, with access to these data shared across medical centers. This information can then be connected to powerful supercomputers that crunch the data on many hundreds of thousands of patients to help predict the exact type of disease a given patient has, the molecular features that best define what that type of disease involves, and from this, identify the most appropriate treatments and/or clinical trials, including drugs, that in turn are communicated to the treating physician. As the physician makes decisions based on this knowledge to treat the patient (the physician is not obligated to follow the predictions made from the big data, but rather the predictions made help engage the physician's mind with the

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76800-9 - Bioinformatics and Computational Biology in Drug Discovery and Development
Edited by William T. Loging
Frontmatter
More information

most complete data to recognize patterns relating to the patient condition), the response of the patient can also be assessed, whether their condition improved, got worse, required additional treatments, and so on. Based on this feedback, which is also entered into the database, the models can be updated in an iterative fashion based on the accuracy of their predictions, and in the end the empirical studies will show that the model better predicts treatment path and as a result becomess adopted as the standard of care. In this way, the patient population becomes the clinical trial population, and the clinical trial employs advanced dynamic, adaptive, deep-learning procedures to maximally consider all data over time to reach the most accurate and validated conclusions.

To enable this new world of healthcare, we see now rising a new generation of biologists who are being trained simultaneously in traditional biomedical sciences, as well as in informatics and statistics. It is these new biologists that will help drive revolutionary progress in the biomedical sciences and in healthcare. I sincerely hope that the approaches and technologies described in this book will inspire this next generation of scientists on how to engage this emerging consumer-centric model of healthcare, to think and dream big about the future of healthcare and drug discovery, and in so doing positively impact the human race by enhancing our well-being.

Eric E. Schadt
Jean C. and James W. Crystal Professor of Genomics
Chairman and Professor, Department of Genetics and Genomic Sciences
Founding Director, Icahn Institute for Genomics and Multiscale Biology
Icahn School of Medicine at Mount Sinai in New York

References

Friend, S. and Schadt, E. Translational genomics. Clues from the resilient. *Science*. 2014;344:970–972.

Schadt, E., Sachs, A., and Friend, S. Embracing complexity, inching closer to reality. *Sci STKE*. 2005;295:pe40.

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-76800-9 - Bioinformatics and Computational Biology in Drug Discovery and Development
Edited by William T. Loging
Frontmatter
More information

Acknowledgments

A person is often the sum of experiences and the people one meets. That said, I have had the privilege of working with many amazing scientists and leaders throughout my time on this planet. To thank each of these people in depth would take an entire book in itself. However, I will try to provide brief credits. To the subject matter contributors of this book, for their contributions to the field and their work on constant revisions of both text and figures in their endless search of perfection; to members of Cambridge University Press, for their professionalism and patience on timelines. For my current and past team members; it has been said that if you want to be happy, surround yourself with people you enjoy working with. To my PhD mentor, Dr. David Reisman, for teaching me how to dig deep in the art of molecular sciences. For the departed Prof. Geoffrey Zubay (whom I never met) for inspiring me on how a textbook should be written and the inner workings of life. To Drs. Anton Fliri and Robert Volkmann for their mentorship and revealing to me how drug discovery works. To Dr. Lee Harland and fellow contributor Bryn Williams-Jones, for their friendship and long discussions of data science and the Edwardian era. Finally, and most importantly, to my family: Christie, Bri, and Liam, for their support, as well as their sacrifice, in supporting my path in the life sciences.

© in this web service Cambridge University Press www.cambridge.org