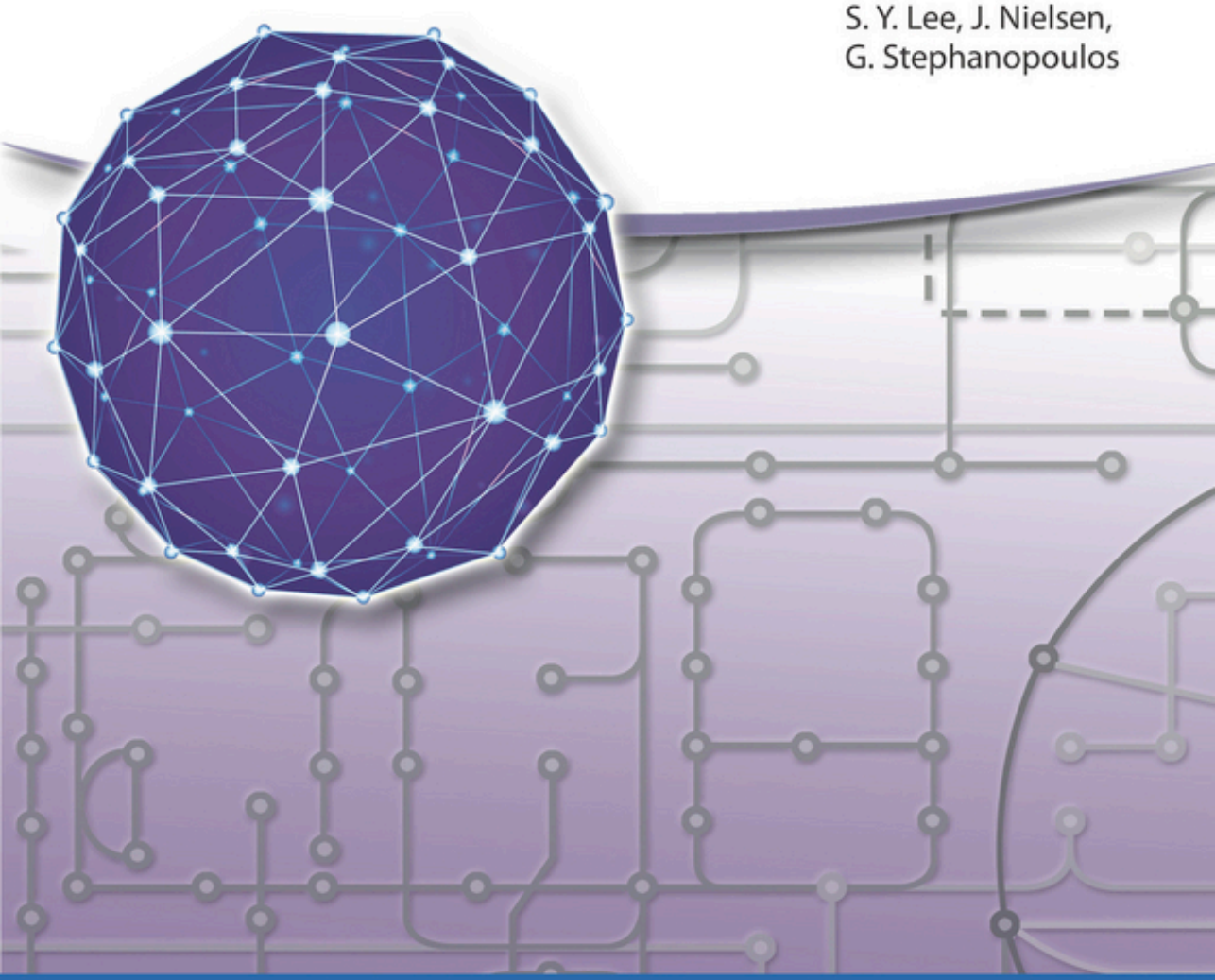
WILEY-VCH

Edited by Jens Nielsen and Stefan Hohmann

Systems Biology

Volume 6
Series Editors:
S. Y. Lee, J. Nielsen,
G. Stephanopoulos



Advanced Biotechnology

Edited by
Jens Nielsen and
Stefan Hohmann

Systems Biology

Related Titles

Meyers, R.A. (ed.)

Systems Biology

2012
Print ISBN: 978-3-527-32607-5

Dehmer, M., Emmert-Streib, F., Graber, A., Salvador, A. (eds.)

Applied Statistics for Network Biology

Methods in Systems Biology

2011
Print ISBN: 978-3-527-32750-8

Further Volumes of the "Advanved Biotechnology" Series:

Published:

Villadsen, J. (ed.)

Fundamental Bioengineering

2016
Print ISBN: 978-3-527-33674-6

Love, J. Ch. (ed.)

Micro- and Nanosystems for Biotechnology

2016
Print ISBN: 978-3-527-33281-6

Wittmann, Ch., Liao, J.C. (eds.)

Industrial Biotechnology

Microorganisms
(2 Volumes)

2017
Print ISBN: 978-3-527-34179-5

Wittmann, Ch., Liao, J.C. (eds.)

Industrial Biotechnology

Products and Processes

2017
Print ISBN: 978-3-527-34181-8

Coming soon:

Yoshida, T. (ed.)

Applied Bioengineering

2017
Print ISBN: 978-3-527-34075-0

Chang, H.N. (ed.)

Emerging Areas in Bioengineering

2017
Print ISBN: 9783527340880

Edited by Jens Nielsen and Stefan Hohmann

Systems Biology

WILEY-VCH
Verlag GmbH & Co. KGaA

Volume Editors

Jens Nielsen
Chalmers University of Technology
Department of Biology and Biological Engineering
Kemivägen 10
41296 Göteborg
Sweden

Stefan Hohmann
Chalmers University of Technology
Department of Biology and Biological Engineering
Kemigården 4, room 3054A
41296 Göteborg
Sweden

Series Editors

Sang Yup Lee
KAIST
373-1; Guseong-Dong
291 Daehak-ro, Yuseong-gu
305-701 Daejon
South Korea

Jens Nielsen
Chalmers University of Technology
Department of Biology and Biological Engineering
Kemivägen 10
41296 Göteborg
Sweden

Gregory Stephanopoulos
Massachusetts Instituts of Technology
Department of Chemical Engineering
77 Massachusetts Avenue
Cambridge, MA 02139
USA

Cover
Sphere – fotolia_©ket4up

Library of Congress Card No.: applied for

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at <http://dnb.d-nb.de>.

Print ISBN: 978-3-527-33558-9
ePDF ISBN: 978-3-527-69616-1
ePub ISBN: 978-3-527-69617-8
Mobi ISBN: 978-3-527-69615-4
oBook ISBN: 978-3-527-69613-0

Cover Design Adam Design
Typesetting SPi Global, Chennai, India
Printing and Binding

Printed on acid-free paper

Contents

List of Contributors

Elena Abad
Universitat Pompeu Fabra
Departament de Ciències
Experimentals i de la Salut
Barcelona Biomedical Research Park
Dr. Aiguader 88
08003 Barcelona
Spain

and

IDIBAPS-Hospital Clínic
Center for Neuroimmunology
Barcelona
Spain

Johannes Borgqvist
Chalmers University of Technology
and the University of Gothenburg
Department of Mathematical Sciences
Chalmers tvärgata 3
412 96 Gothenburg
Sweden

Siu H. J. Chan
The Pennsylvania State University
Department of Chemical Engineering
112 Fenske Laboratory
University Park, PA 16802
USA

Matthew W. Chang
National University of Singapore
Yong Loo Lin School of Medicine,
Department of Biochemistry
8 Medical Drive
Singapore 117597
Singapore

and

National University of Singapore
NUS Synthetic Biology for Clinical
and Technological Innovation
(SynCTI), Life Sciences Institute
28 Medical Drive
Singapore 117456
Singapore

Yi Ern Cheah
Vanderbilt University
Department of Chemical and
Biomolecular Engineering
Nashville, TN 37235
USA

Marija Cvijovic
Chalmers University of Technology
and University of Gothenburg
Department of Mathematical Sciences
Chalmers tvärgata 3
412 96 Gothenburg
Sweden

Riccardo Dainese
Chalmers University of Technology
and the University of Gothenburg
Department of Mathematical Sciences
Chalmers tvärgata 3
412 96 Gothenburg
Sweden

and

École Polytechnique Fédérale de
Lausanne (EPFL)
Laboratory of Systems Biology and
Genetics, Institute of Bioengineering,
School of Life Sciences
Lausanne
Switzerland

Bonnie V. Dougherty
University of Virginia
Department of Biomedical
Engineering
415 Lane Road
Charlottesville, VA 22908
USA

Jee L. Foo
National University of Singapore
Yong Loo Lin School of Medicine,
Department of Biochemistry
8 Medical Drive
Singapore 117597
Singapore

and

National University of Singapore
NUS Synthetic Biology for Clinical
and Technological Innovation
(SynCTI), Life Sciences Institute
28 Medical Drive
Singapore 117456
Singapore

Jordi García-Ojalvo
Universitat Pompeu Fabra
Departament de Ciències
Experimentals i de la Salut
Barcelona Biomedical Research Park
Dr. Aiguader 88
08003 Barcelona
Spain

Lada Georgieva
ETH Zurich
Department of Biosystems, Science
and Engineering (D-BSSE), Swiss
Institute of Bioinformatics (SIB)
Mattenstraße 26
4058 Basel
Switzerland

Harold F. Gómez
ETH Zurich
Department of Biosystems, Science
and Engineering (D-BSSE), Swiss
Institute of Bioinformatics (SIB)
Mattenstraße 26
4058 Basel
Switzerland

Diego R. Gomez-Alvarez
Chalmers University of Technology
and University of Gothenburg
Department of Mathematical Sciences
Chalmers tvärgata 3
412 96 Gothenburg
Sweden

Sanjan T. P. Gupta
University of Wisconsin – Madison
Department of Chemical and
Biological Engineering
3639 Engineering Hall
1415 Engineering Drive
Madison, WI 53706
USA

and

University of Wisconsin – Madison
Great Lakes Bioenergy Research
Center
Madison, WI
USA

Clinton M. Hasenour
Vanderbilt University
Department of Chemical and
Biomolecular Engineering
Nashville, TN 37235
USA

Dagmar Iber
ETH Zurich
Department of Biosystems, Science
and Engineering (D-BSSE), Swiss
Institute of Bioinformatics (SIB)
Mattenstraße 26
4058 Basel
Switzerland

Edda Klipp
Humboldt-Universität zu Berlin
Institute of Biology
Invalidenstr. 42
10115 Berlin
Germany

Marcus Krantz
Humboldt-Universität zu Berlin
Institute of Biology
Invalidenstr. 42
10115 Berlin
Germany

Erik Kristiansson
Chalmers University of
Technology/University of Gothenburg
Department of Mathematical Sciences
Chalmers tvärgata 3
412 96 Gothenburg
Sweden

Prashant Kumar
University of Wisconsin – Madison
Department of Chemical and
Biological Engineering
3639 Engineering Hall
1415 Engineering Drive
Madison, WI 53706
USA

Meiyappan Lakshmanan
Agency for Science, Technology and
Research (A*STAR)
Bioprocessing Technology Institute
20 Biopolis Way, #06-01, Centros
Singapore 138668
Singapore

Dong-Yup Lee
Agency for Science, Technology and Research (A*STAR)
Bioprocessing Technology Institute
20 Biopolis Way, #06-01, Centros
Singapore 138668
Singapore

and

National University of Singapore
NUS Synthetic Biology for Clinical and Technological Innovation (SynCTI), Life Sciences Institute
28 Medical Drive
Singapore 117456
Singapore

and

National University of Singapore
Department of Chemical and Biomolecular Engineering
4 Engineering Drive 4
Singapore 117585
Singapore

Na-Rae Lee
National University of Singapore
NUS Synthetic Biology for Clinical and Technological Innovation (SynCTI), Life Sciences Institute
28 Medical Drive
Singapore 117456
Singapore

Rune Linding
University of Copenhagen (UCPH)
Biotech Research and Innovation Centre (BRIC)
Ole Maaløes Vej 5, 4
2200 Copenhagen
Denmark

Hua Ling
National University of Singapore
Yong Loo Lin School of Medicine, Department of Biochemistry
8 Medical Drive
Singapore 117597
Singapore

and

National University of Singapore
NUS Synthetic Biology for Clinical and Technological Innovation (SynCTI), Life Sciences Institute
28 Medical Drive
Singapore 117456
Singapore

Timo Lubitz
Humboldt-Universität zu Berlin
Institute of Biology
Invalidenstr. 42
10115 Berlin
Germany

Min Ma
University of Lausanne
Department of Fundamental Microbiology
Biophore building
1015 Lausanne
Switzerland

Costas D. Maranas
The Pennsylvania State University
Department of Chemical Engineering
112 Fenske Laboratory
University Park, PA 16802
USA

Odysse Michos
ETH Zurich
Department of Biosystems, Science and Engineering (D-BSSE), Swiss Institute of Bioinformatics (SIB)
Mattenstraße 26
4058 Basel
Switzerland

Nadim Mira
University of Lausanne
Department of Fundamental Microbiology
Biophore building
1015 Lausanne
Switzerland

Thomas J. Moutinho Jr.
University of Virginia
Department of Biomedical Engineering
415 Lane Road
Charlottesville, VA 22908
USA

Ulrike Münzner
Humboldt-Universität zu Berlin
Institute of Biology
Invalidenstr. 42
10115 Berlin
Germany

Eduardo P. Olimpio
Delft University of Technology
Department of Bionanoscience, Kavli Institute of Nanoscience
Lorentzweg 1, Building 22
2628 CJ Delft
The Netherlands

Tobias Österlund
Chalmers University of Technology/University of Gothenburg
Department of Mathematical Sciences
Chalmers tvärgata 3
412 96 Gothenburg
Sweden

Jason Papin
University of Virginia
Department of Biomedical Engineering
415 Lane Road
Charlottesville, VA 22908
USA

Serge Pelet
University of Lausanne
Department of Fundamental Microbiology
Biophore building
1015 Lausanne
Switzerland

Jennifer L. Reed
University of Wisconsin – Madison
Department of Chemical and Biological Engineering
3639 Engineering Hall
1415 Engineering Drive
Madison, WI 53706
USA

and

University of Wisconsin – Madison
Great Lakes Bioenergy Research Center
Madison, WI
USA

Gourvendu Saxena
National University of Singapore
NUS Synthetic Biology for Clinical and Technological Innovation (SynCTI), Life Sciences Institute
28 Medical Drive
Singapore 117456
Singapore

and

National University of Singapore
Singapore Centre on Environmental Life Sciences Engineering (SCELSE), Life Sciences Institute
28 Medical Drive
Singapore 117456
Singapore

and

National University of Singapore
Faculty of Science, Department of Biological Sciences
14 Science Drive 4
Singapore 117543
Singapore

Margaret Simons
The Pennsylvania State University
Department of Chemical Engineering
112 Fenske Laboratory
University Park, PA 16802
USA

Sanjay Swarup
National University of Singapore
NUS Synthetic Biology for Clinical and Technological Innovation (SynCTI), Life Sciences Institute
28 Medical Drive
Singapore 117456
Singapore

and

National University of Singapore
Singapore Centre on Environmental Life Sciences Engineering (SCELSE), Life Sciences Institute
28 Medical Drive
Singapore 117456
Singapore

and

National University of Singapore
Faculty of Science, Department of Biological Sciences
14 Science Drive 4
Singapore 117543
Singapore

Mingyuan Tian
University of Wisconsin – Madison
Department of Chemical and Biological Engineering
3639 Engineering Hall
1415 Engineering Drive
Madison, WI 53706
USA

Pablo Villoslada
IDIBAPS-Hospital Clínic
Center for Neuroimmunology
Casanova, 143
08028 Barcelona
Spain

Franziska Voellmy
University of Copenhagen (UCPH)
Biotech Research and Innovation Centre (BRIC)
Ole Maaløes Vej 5, 4
2200 Copenhagen
Denmark

and

Technical University of Denmark
Denmark Novo Nordisk Foundation Center for Biosustainability
2970 Hørsholm
Denmark

and

Technical University of Denmark
Department of Systems Biology
2800 Lyngby
Denmark

Hyun Youk
Delft University of Technology
Department of Bionanoscience, Kavli Institute of Nanoscience
Lorentzweg 1, Building 22
2628 CJ Delft
The Netherlands

Jamey D. Young
Vanderbilt University
Department of Chemical and Biomolecular Engineering
Nashville, TN 37235
USA

and

Vanderbilt University
Department of Molecular Physiology and Biophysics
Nashville, TN
USA

About the Series Editors

Sang Yup Lee is Distinguished Professor at the Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology. At present, Professor Lee is the Director of the Center for Systems and Synthetic Biotechnology, Director of the BioProcess Engineering Research Center, and Director of the Bioinformatics Research Center. He has published more than 500 research papers, written 64 books or book chapters, and has more than 580 patents (either awarded or applied). He is a recipient of numerous awards, including the National Order of Merit, the Merck Metabolic Engineering Award, the ACS Marvin Johnson Award, the Charles Thom Award, the Amgen Biochemical Engineering Award, the Elmer Gaden Award, the POSCO TJ Park Prize, and the HoAm Prize. Professor Lee is a Fellow of the American Association for the Advancement of Science, the American Academy of Microbiology, the American Institute of Chemical Engineers, the Society for Industrial Microbiology and Biotechnology, the American Institute of Medical and Biological Engineering, the World Academy of Science, the Korean Academy of Science and Technology, and the National Academy of Engineering of Korea. He is also Foreign Member of the National Academy of Engineering, USA. In addition, he is Honorary Professor at the University of Queensland (Australia), the Chinese Academy of Sciences, Wuhan University (China), Hubei University of Technology (China), and the Beijing University of Chemical Technology (China), and Advisory Professor at the Shanghai Jiao Tong University (China). He is the Editor-in-Chief of the *Biotechnology Journal* and has been serving many other journals as Associate Editor or board member. He is a member of the Presidential Advisory Committee on Science and Technology, South Korea.

Jens Nielsen has been Professor and Director of Chalmers University of Technology, Sweden, since 2008. He received the M.Sc. degree in chemical engineering and the Ph.D. degree in biochemical engineering (1989) from the Technical University of Denmark (DTU). He established his independent research group and was appointed full Professor at DTU in 1998. He was Fulbright Visiting Professor at Massachusetts Institute of

Technology (MIT), USA, during 1995–1996. At DTU, he founded and directed the Center for Microbial Biotechnology. Professor Nielsen has published more than 350 research papers and co-authored more than 40 books. He is the inventor of more than 50 patents. He has founded several companies that have raised more than US$20 million in venture capital. He has received numerous Danish and international awards. Professor Nielsen is member of the Academy of Technical Sciences (Denmark), the National Academy of Engineering (USA), the Royal Danish Academy of Science and Letters, the American Institute for Medical and Biological Engineering, and the Royal Swedish Academy of Engineering Sciences.

Gregory Stephanopoulos is the W.H. Dow Professor of Chemical Engineering, Massachusetts Institute of Technology (MIT), USA, and Director of the MIT Metabolic Engineering Laboratory. He is also Instructor of Bioengineering at Harvard Medical School (since 1997). He received the B.S. degree from the National Technical University of Athens and the Ph.D. degree from the University of Minnesota, USA. He has co-authored nearly 400 research papers and is the inventor of 50 patents. He authored the first textbook on metabolic engineering. Professor Stephanopoulos is the recipient of numerous awards from professional societies such as American Institute of Chemical Engineers (AIChE), the American Chemical Society (ACS), and the Society of industrial Microbiology (SIM), and has won the BIO (Washington Carver Award) and the John Fritz Medal of the American Association of Engineering Societies, among others. In 2003, he was elected member of the National Academy of Engineering, USA. He was the President of AIChE in 2014.

1
Integrative Analysis of Omics Data

Tobias Österlund, Marija Cvijovic, and Erik Kristiansson

Summary

Data generation and analysis are essential parts of systems biology. Today, large amounts of omics data can be generated fast and cost-efficiently thanks to the development of modern high-throughput measurement techniques. Their interpretation is, however, challenging because of the high dimensionality and the often substantial levels of noise. Integrative analysis provides a framework for analysis of the omics data from a biological perspective, starting from the raw data, via preprocessing and statistical analysis, to the interpretation of the results. By integrating the data into structures created from biological information available in resources, databases, or genome-scale models, the focus moves from the individual transcripts or proteins to the entire pathways and other relevant biochemical functions present in the cell. The result provides a context-based interpretation of the omics data, which can be used to form a holistic and unbiased view of biological systems at a molecular level. The concept of integrative analysis can be used for many forms of omics data, including genome sequencing, transcriptomics, and proteomics, and can be applied to a wide range of fields within the life sciences.

1.1
Introduction

Systems biology is an interdisciplinary approach to biology and medicine that employs both experimentation and mathematical modeling to achieve a better understanding of biological systems by describing their shape, state, behavior, and evolutionary history. An important aim of systems biology is to deliver predictive and informative models that highlight the fundamental and presumably conserved relationships of biomolecular systems and thereby provide an improved insight into the many cellular processes [1]. Systems biology research methodology is a cyclical process fueled by quantitative experiments in combination with mathematical modeling (Figure 1.1) [2, 3]. In its most basic form, the cycle starts with the formulation of a set of hypotheses, which is followed by knowledge generation

Systems Biology, First Edition. Edited by Jens Nielsen and Stefan Hohmann.

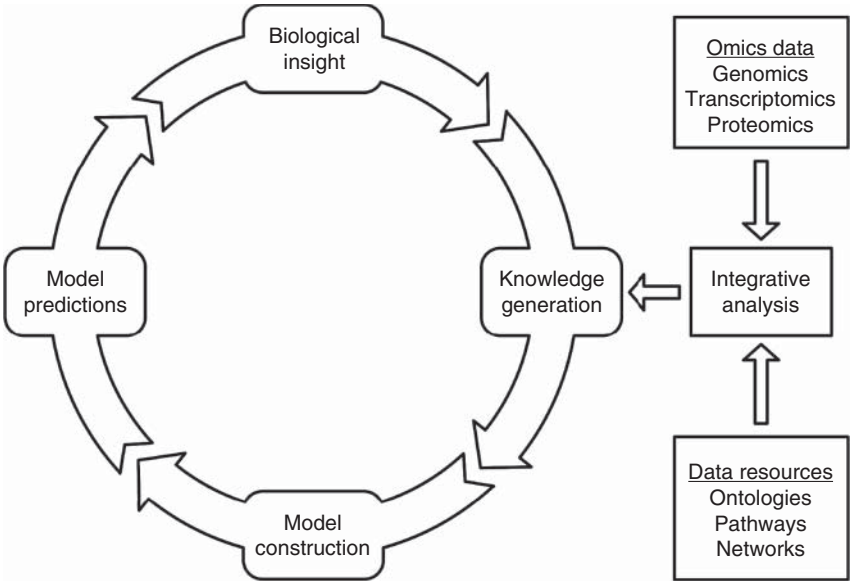


Figure 1.1 Systems biology research methodology. In the systems biology cycle, novel hypotheses are first formulated, which is followed by knowledge generation, model construction, and model predictions, which, in turn, leads to new biological insights. The development of high-throughput techniques have enabled rapid and cost-efficient generation of omics data from, for example, genome sequencing, transcriptomics, and proteomics. Integrative analysis provides a framework where omics data is systematically analyzed in a biological context, by data integration into known biological networks or other data resources, which enables improved interpretation and easier integration into quantitative models.

and model construction where an abstract description of the biological system (a model) is formulated and its parameters are estimated from data taken from the literature. The final step is defined by model predictions, where the constructed model is used to address the original hypotheses by providing a quantitative analysis of the system, which, in turn, generates new biological insight.

The development of high-throughput measurement techniques in the recent years has resulted in an unprecedented ability to rapidly and cost efficiently generate molecular data. Bioassays are today established for large-scale characterization of genes and their expression at the different layers defined by the central dogma: the genome, the transcriptome, and the proteome. The resulting data, which in this chapter will be referred to as *omics data*, is however complex because of its high dimensionality and is therefore hard to interpret and directly integrate into quantitative models. The concept of *integrative analysis* is a framework to systematically analyze the different components of omics data in relation to their corresponding biological functions and properties. The resulting biological interpretation can be used to form a holistic and unbiased view of biological systems at a molecular level. Thanks to the comprehensiveness of the

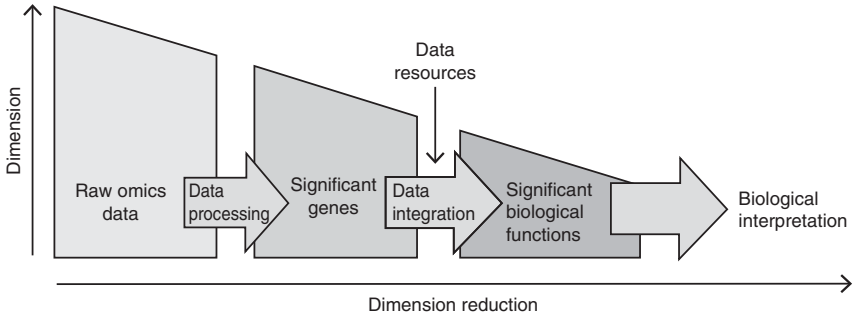


Figure 1.2 Description of the concept of integrative analysis as a tool for reduction of the dimension of omics data. Integrative analysis starts with raw omics data, which is typically affected by high levels of noise and errors. Computational and statistical approaches are first used to process the data to produce a ranked list of genes that are found to be of significant importance in the experiment. The gene list is used as input to the data integration, where known biological information is used as a basis for the interpretation of the data. During integrative analysis, the dimension of the data is significantly reduced, from potentially millions of data points to a limited number of significant biological functions and pathways, which considerably facilitates the interpretation.

omics data, all components (i.e., genes, transcripts, or proteins) can be measured simultaneously, which opens up opportunities for testing of existing hypotheses as well as generation of completely new hypotheses of the studied biological system.

The process of integrative analysis can be divided into two main steps: data processing and data integration (Figure 1.2). Integrative analysis starts from raw omics data and ends with the biological interpretation, and during this process the dimensionality of the data is reduced. The first step, the data processing, takes the high-dimensional omics data, and by applying computational and statistical tools, removes noise and errors while identifying genes and other components that contain information significant for the experiment. The next step, the data integration, uses the list of identified genes to pinpoint relevant functions and pathways by integrating the data on top of a "scaffold" built using established biological information collected from various resources and databases. The result, which is based on the combined analysis of the genes with similar functional properties, has a substantially reduced dimension, which considerably facilitates its interpretation.

Many studies in the life sciences aim to understand biological systems, often in relation to a perturbation caused by, for example, disease, genetic variability, changes in environmental parameters, or other factors introduced through laboratory experiments. A commonly used measurement technique is transcriptomics, where the transcriptional response is analyzed and the genes that are differentially expressed between investigated conditions are identified. In this setting, the data integration shifts the focus from what *genes* are differentially expressed to providing a *biological context* where activated and repressed pathways, functions, or subnetworks can be identified. This provides a more relevant view of the data, which paves the way toward more sound and detailed biological conclusions.

In this chapter, we provide a broad overview of integrative analysis of omics data. We will describe the general concept of integrative analysis and provide an outline of the many associated computational steps. It should, however, be pointed out that this topic has been extensively researched during the recent years and – due to the scope of the topic at hand – we will not be able to cover all aspects and details in a single chapter. We have therefore provided a comprehensive set of references throughout the text, which are the recommended starting points for further reading. Also, our main focus throughput this chapter will be on data generated by techniques from genomics, transcriptomics, and proteomics. This means that other types of data, which are commonly encountered in systems biology, such as metabolomics and lipidomics, will receive little attention, and here we instead refer the reader to the recent reviews by Robinson *et al*. [4] and Kim *et al*. [5].

The chapter is organized as follows. Section 1.2 contains an overview of some the types of omics data that are commonly used in integrative analysis. This is followed by Section 1.3, where we focus on the data processing, starting from the quality assessment of the raw data to statistical analysis. Section 1.4 explains the concepts of data integration and describes the different approaches and data resources that can be used. We end the chapter with an outlook discussing future challenges related to the continuous growth of biological information.

1.2 Omics Data and Their Measurement Platforms

In this section three commonly used types of omics data will be described, namely genome sequencing, transcriptomics (RNA sequencing and microarrays), and mass spectrometry (MS)-based proteomics.

1.2.1 Omics Data Types

Genome sequencing is used for determining the order of the complete set of nucleotides present in an organism. The comparative analysis of the genome of a strain or a multicellular organism in relation to a reference genome is referred to as "resequencing," which enables identification of the complete genotype and its variation between individuals. This includes both small mutations, such as single nucleotide polymorphisms (SNPs) and short insertions/deletions (indels), and larger structural variations such as genome rearrangements and copy number alterations [6]. The resulting information, containing a list of all identified genetic variants, is often subjected to integrative analysis in order to provide a biological context where the genotype can be linked to a phenotype [7]. Whole-genome and exome resequencing are important techniques for the study of human disease [8], and in, for example, cancer, the set of germline and somatic mutations are