

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 "Advanced 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 Daejeon
South Korea

Jens Nielsen

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

Gregory Stephanopoulos

Massachusetts Institute of Technology
Department of Chemical Engineering
77 Massachusetts Avenue
Cambridge, MA 02139
USA

Cover

Sphere – fotolia_@ket4up

All books published by **Wiley-VCH** are carefully produced. Nevertheless, authors, editors, and publisher do not warrant the information contained in these books, including this book, to be free of errors. Readers are advised to keep in mind that statements, data, illustrations, procedural details or other items may inadvertently be inaccurate.

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>>.

© 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Boschstr. 12, 69469 Weinheim, Germany

All rights reserved (including those of translation into other languages). No part of this book may be reproduced in any form – by photoprinting, microfilm, or any other means – nor transmitted or translated into a machine language without written permission from the publishers. Registered names, trademarks, etc. used in this book, even when not specifically marked as such, are not to be considered unprotected by law.

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 XV

About the Series Editors XXIII

- 1 Integrative Analysis of Omics Data 1**
Tobias Österlund, Marija Cvijovic, and Erik Kristiansson
 - Summary 1
 - 1.1 Introduction 1
 - 1.2 Omics Data and Their Measurement Platforms 4
 - 1.2.1 Omics Data Types 4
 - 1.2.2 Measurement Platforms 5
 - 1.3 Data Processing: Quality Assessment, Quantification, Normalization, and Statistical Analysis 6
 - 1.3.1 Quality Assessment 7
 - 1.3.2 Quantification 9
 - 1.3.3 Normalization 10
 - 1.3.4 Statistical Analysis 11
 - 1.4 Data Integration: From a List of Genes to Biological Meaning 12
 - 1.4.1 Data Resources for Constructing Gene Sets 13
 - 1.4.1.1 Gene Ontology Terms 13
 - 1.4.1.2 KEGG and Reactome 13
 - 1.4.1.3 Genome-Scale Metabolic Reconstructions 14
 - 1.4.2 Gene Set Analysis 14
 - 1.4.2.1 Gene Set Overenrichment Tests 16
 - 1.4.2.2 Rank-Based Enrichment Tests 16
 - 1.4.3 Networks and Network Topology 17
 - 1.5 Outlook and Perspectives 18
 - References 19

- 2 ¹³C Flux Analysis in Biotechnology and Medicine 25**
Yi Ern Cheah, Clinton M. Hasenour, and Jamey D. Young
 - 2.1 Introduction 25
 - 2.1.1 Why Study Metabolic Fluxes? 25
 - 2.1.2 Why are Isotope Tracers Important for Flux Analysis? 26

2.1.3	How are Fluxes Determined?	28
2.2	Theoretical Foundations of ^{13}C MFA	29
2.2.1	Elementary Metabolite Units (EMUs)	30
2.2.2	Flux Uncertainty Analysis	31
2.2.3	Optimal Design of Isotope Labeling Experiments	32
2.2.4	Isotopically Nonstationary MFA (INST-MFA)	34
2.3	Metabolic Flux Analysis in Biotechnology	36
2.3.1	^{13}C MFA for Host Characterization	36
2.3.2	^{13}C MFA for Pinpointing Yield Losses and Futile Cycles	39
2.3.3	^{13}C MFA for Bottleneck Identification	41
2.4	Metabolic Flux Analysis in Medicine	42
2.4.1	Liver Glucose and Oxidative Metabolism	43
2.4.2	Cancer Cell Metabolism	47
2.4.3	Fuel Oxidation and Anaplerosis in the Heart	48
2.4.4	Metabolism in Other Tissues: Pancreas, Brain, Muscle, Adipose, and Immune Cells	49
2.5	Emerging Challenges for ^{13}C MFA	50
2.5.1	Theoretical and Computational Advances: Multiple Tracers, Co-culture MFA, Dynamic MFA	50
2.5.2	Genome-Scale ^{13}C MFA	51
2.5.3	New Measurement Strategies	52
2.5.4	High-Throughput MFA	53
2.5.5	Application of MFA to Industrial Bioprocesses	53
2.5.6	Integrating MFA with Omics Measurements	54
2.6	Conclusion	55
	Acknowledgments	55
	Disclosure	55
	References	55
3	Metabolic Modeling for Design of Cell Factories	71
	<i>Mingyuan Tian, Prashant Kumar, Sanjan T. P. Gupta, and Jennifer L. Reed</i>	
	Summary	71
3.1	Introduction	71
3.2	Building and Refining Genome-Scale Metabolic Models	72
3.2.1	Generate a Draft Metabolic Network (Step 1)	74
3.2.2	Manually Curate the Draft Metabolic Network (Step 2)	75
3.2.3	Develop a Constraint-Based Model (Step 3)	77
3.2.4	Revise the Metabolic Model through Reconciliation with Experimental Data (Step 4)	79
3.2.5	Predicting the Effects of Genetic Manipulations	81
3.3	Strain Design Algorithms	83
3.3.1	Fundamentals of Bilevel Optimization	84
3.3.2	Algorithms Involving Only Gene/Reaction Deletions	94
3.3.3	Algorithms Involving Gene Additions	94
3.3.4	Algorithms Involving Gene Over/Underexpression	95

3.3.5	Algorithms Involving Cofactor Changes	98
3.3.6	Algorithms Involving Multiple Design Criteria	99
3.4	Case Studies	100
3.4.1	Strains Producing Lactate	100
3.4.2	Strains Co-utilizing Sugars	100
3.4.3	Strains Producing 1,4-Butanediol	102
3.5	Conclusions	103
	Acknowledgments	103
	References	104
4	Genome-Scale Metabolic Modeling and <i>In silico</i> Strain Design of <i>Escherichia coli</i>	109
	<i>Meiyappan Lakshmanan, Na-Rae Lee, and Dong-Yup Lee</i>	
4.1	Introduction	109
4.2	The COBRA Approach	110
4.3	History of <i>E. coli</i> Metabolic Modeling	111
4.3.1	Pre-genomic-era Models	111
4.3.2	Genome-Scale Models	112
4.4	<i>In silico</i> Model-Based Strain Design of <i>E. coli</i> Cell Factories	115
4.4.1	Gene Deletions	127
4.4.2	Gene Up/Downregulations	127
4.4.3	Gene Insertions	128
4.4.4	Cofactor Engineering	128
4.4.5	Other Approaches	128
4.5	Future Directions of Model-Guided Strain Design in <i>E. coli</i>	129
	References	130
5	Accelerating the Drug Development Pipeline with Genome-Scale Metabolic Network Reconstructions	139
	<i>Bonnie V. Dougherty, Thomas J. Moutinho Jr., and Jason Papin</i>	
	Summary	139
5.1	Introduction	139
5.1.1	Drug Development Pipeline	140
5.1.2	Overview of Genome-Scale Metabolic Network Reconstructions	140
5.1.3	Analytical Tools and Mathematical Evaluation	141
5.1.3.1	Flux Balance Analysis (FBA)	141
5.1.3.2	Flux Variability Analysis (FVA)	142
5.2	Metabolic Reconstructions in the Drug Development Pipeline	142
5.2.1	Target Identification	143
5.2.2	Drug Side Effects	145
5.3	Species-Level Microbial Reconstructions	146
5.3.1	Microbial Reconstructions in the Antibiotic Development Pipeline	146
5.3.1.1	Applications in the Drug Development Pipeline	146

5.3.2	Metabolic-Reconstruction-Facilitated Rational Drug Target Identification	147
5.3.2.1	Targeting Genes Essential for Biomass Production	147
5.3.2.2	Targeting Virulence Factors	147
5.3.2.3	Metabolite-centric Targeting	148
5.3.3	Repurposing and Expanding Utility of Antibiotics	149
5.3.3.1	Virtual Drug Screens Informed by Metabolic Reconstructions	149
5.3.3.2	Limiting Resistance with Drug Combinations	149
5.3.3.3	Improving Treatment Options by Increasing Sensitivity to Antibiotics	150
5.3.4	Improving Toxicity Screens with the Human Metabolic Network Reconstruction	150
5.4	The Human Reconstruction	151
5.4.1	Approaches for the Human Reconstruction	152
5.4.2	Target Identification	152
5.4.2.1	Drug Targeting in Cancer	152
5.4.2.2	Drug Targeting in Metabolic Diseases	153
5.4.3	Toxicity and Other Side Effects	154
5.5	Community Models	155
5.5.1	Host–Pathogen Community Models	155
5.5.2	Eukaryotic Community Models	156
5.6	Personalized Medicine	156
5.7	Conclusion	157
	References	158
6	Computational Modeling of Microbial Communities	163
	<i>Siu H. J. Chan, Margaret Simons, and Costas D. Maranas</i>	
	Summary	163
6.1	Introduction	163
6.1.1	Microbial Communities	163
6.1.2	Modeling Microbial Communities	165
6.1.3	Model Structures	165
6.1.4	Quantitative Approaches	166
6.2	Ecological Models	168
6.2.1	Generalized Predator–Prey Model	169
6.2.2	Evolutionary Game Theory	170
6.2.3	Models Including Additional Dimensions	171
6.2.4	Advantages and Disadvantages	171
6.3	Genome-Scale Metabolic Models	172
6.3.1	Introduction and Applications	172
6.3.2	Genome-Scale Metabolic Modeling of Microbial Communities	174
6.3.3	Simulation of Microbial Communities Assuming Steady State	175
6.3.3.1	Predicting Interactions Using FBA	175
6.3.3.2	Identifying Minimal Media by Mixed Integer Linear Programming	176

6.3.3.3	Pareto Optimality Analysis by FBA	176
6.3.3.4	Modeling Chemostat Co-culture	177
6.3.3.5	Community FBA with Community Mass Balance	177
6.3.4	Dynamic Simulation of Multispecies Models	177
6.3.5	Spatial and Temporal Modeling of Communities	178
6.3.6	Using Bilevel Optimization to Capture Multiple Objective Functions	179
6.3.6.1	OptCom	179
6.3.6.2	d-OptCom	181
6.3.6.3	CASINO Toolbox	181
6.3.6.4	Advantages and Disadvantages	182
6.3.6.5	Current Challenges and Future Directions	182
6.4	Concluding Remarks	183
	References	183
7	Drug Targeting of the Human Microbiome	191
	<i>Hua Ling, Jee L. Foo, Gourvindu Saxena, Sanjay Swarup, and Matthew W. Chang</i>	
	Summary	191
7.1	Introduction	191
7.2	The Human Microbiome	192
7.3	Association of the Human Microbiome with Human Diseases	194
7.3.1	Nasal–Sinus Diseases	194
7.3.2	Gut Diseases	194
7.3.3	Cardiovascular Diseases	196
7.3.4	Metabolic Disorders	196
7.3.5	Autoimmune Disorders	197
7.3.6	Lung Diseases	197
7.3.7	Skin Diseases	197
7.4	Drug Targeting of the Human Microbiome	198
7.4.1	Prebiotics	198
7.4.2	Probiotics	200
7.4.3	Antimicrobials	201
7.4.3.1	Antibiotics	201
7.4.3.2	Antimicrobial Peptides	202
7.4.4	Signaling Inhibitors	202
7.4.5	Metabolites	203
7.4.5.1	Short-Chain Fatty Acids	203
7.4.5.2	Bile Acids	203
7.4.6	Metabolite Receptors and Enzymes	204
7.4.6.1	Metabolite Receptors	204
7.4.6.2	Metabolic Enzymes	204
7.4.7	Microbiome-Aided Drug Metabolism	205
7.4.7.1	Drug Delivery and Release	205

7.4.7.2	Drug Toxicity	206
7.4.8	Immune Modulators	206
7.4.9	Synthetic Commensal Microbes	207
7.5	Future Perspectives	207
7.6	Concluding Remarks	208
	Acknowledgments	208
	References	209
8	Toward Genome-Scale Models of Signal Transduction Networks	215
	<i>Ulrike Münzner, Timo Lubitz, Edda Klipp, and Marcus Krantz</i>	
8.1	Introduction	215
8.2	The Potential of Network Reconstruction	219
8.3	Information Transfer Networks	222
8.4	Approaches to Reconstruction of ITNs	225
8.5	The rxncon Approach to ITNWR	230
8.6	Toward Quantitative Analysis and Modeling of Large ITNs	234
8.7	Conclusion and Outlook	236
	Acknowledgments	236
	Glossary	237
	References	238
9	Systems Biology of Aging	243
	<i>Johannes Borgqvist, Riccardo Dainese, and Marija Cvijovic</i>	
	Summary	243
9.1	Introduction	243
9.2	The Biology of Aging	245
9.3	The Mathematics of Aging	249
9.3.1	Databases Devoted to Aging Research	249
9.3.2	Mathematical Modeling in Aging Research	249
9.3.3	Distribution of Damaged Proteins during Cell Division: A Mathematical Perspective	256
9.3.3.1	Cell Growth	256
9.3.3.2	Cell Death	257
9.3.3.3	Cell Division	257
9.4	Future Challenges	260
	Conflict of Interest	262
	References	262
10	Modeling the Dynamics of the Immune Response	265
	<i>Elena Abad, Pablo Villoslada, and Jordi García-Ojalvo</i>	
10.1	Background	265
10.2	Dynamics of NF- κ B Signaling	266
10.2.1	Functional Role and Regulation of NF- κ B	266

10.2.2	Dynamics of the NF- κ B Response to Cytokine Stimulation	267
10.3	JAK/STAT Signaling	273
10.3.1	Functional Roles of the STAT Proteins	273
10.3.2	Regulation of the JAK/STAT Pathway	274
10.3.3	Multiplicity and Cross-talk in JAK/STAT Signaling	275
10.3.4	Early Modeling of STAT Signaling	276
10.3.5	Minimal Models of STAT Activation Dynamics	277
10.3.6	Cross-talk with Other Immune Pathways	279
10.3.7	Population Dynamics of the Immune System	281
10.4	Conclusions	282
	Acknowledgments	283
	References	283
11	Dynamics of Signal Transduction in Single Cells Quantified by Microscopy	289
	<i>Min Ma, Nadim Mira, and Serge Pelet</i>	
11.1	Introduction	289
11.2	Single-Cell Measurement Techniques	291
11.2.1	Flow Cytometry	291
11.2.2	Mass Cytometry	291
11.2.3	Single-Cell Transcriptomics	292
11.2.4	Single-Cell Mass Spectrometry	292
11.2.5	Live-Cell Imaging	292
11.3	Microscopy	293
11.3.1	Epi-Fluorescence Microscopy	294
11.3.2	Fluorescent Proteins	295
11.3.3	Relocation Sensors	295
11.3.4	Förster Resonance Energy Transfer	298
11.4	Imaging Signal Transduction	300
11.4.1	Quantifying Small Molecules	300
11.4.2	Monitoring Enzymatic Activity	301
11.4.2.1	Endogenous Relocation Sensors	301
11.4.2.2	Passive Relocation Sensors	302
11.4.2.3	Active Relocation Sensors	303
11.4.2.4	FRET Biosensors	304
11.4.3	Probing Protein–Protein Interactions	304
11.4.3.1	FRET in Protein Complexes	304
11.4.3.2	Bimolecular Fluorescence Complementation	305
11.4.3.3	Dimerization-Dependent FP	306
11.4.4	Measuring Protein Synthesis	307
11.4.4.1	mRNA Transcription	307
11.4.4.2	Protein Synthesis	308
11.4.4.3	Expression Dynamics Visualized by Protein Relocation	311

11.5	Conclusions	311
	References	312
12	Image-Based <i>In silico</i> Models of Organogenesis	319
	<i>Harold F. Gómez, Lada Georgieva, Odysse Michos, and Dagmar Iber</i>	
	Summary	319
12.1	Introduction	319
12.2	Typical Workflow of Image-Based <i>In silico</i> Modeling Experiments	320
12.2.1	<i>In silico</i> Models of Organogenesis	322
12.2.2	Imaging as a Source of (Semi-)Quantitative Data	323
12.2.2.1	Imaging a Growing Organ	324
12.2.3	Image Analysis and Quantification	326
12.2.4	Computational Simulations of Models Describing Organogenesis	328
12.2.5	Image-Based Parameter Estimation	329
12.2.6	<i>In silico</i> Model Validation and Exchange	329
12.2.6.1	<i>In silico</i> Model Validation	329
12.2.6.2	Model Exchange via the Systems Biology Markup Language (SBML)	330
12.3	Application: Image-Based Modeling of Branching Morphogenesis	331
12.3.1	Image-Based Model Selection	331
12.4	Future Avenues	334
	References	334
13	Progress toward Quantitative Design Principles of Multicellular Systems	341
	<i>Eduardo P. Olimpio, Diego R. Gomez-Alvarez, and Hyun Youk</i>	
	Summary	341
13.1	Toward Quantitative Design Principles of Multicellular Systems	341
13.2	Breaking Multicellular Systems into Distinct Functional and Spatial Modules May Be Possible	342
13.3	Communication among Cells as a Means of Cell–Cell Interaction	346
13.4	Making Sense of the Combinatorial Possibilities Due to Many Ways that Cells Can Be Arranged in Space	350
13.5	From Individual Cells to Collective Behaviors of Cell Populations	352
13.6	Tuning Multicellular Behaviors	355
13.7	A New Framework for Quantitatively Understanding Multicellular Systems	359
	Acknowledgments	361
	References	362

14	Precision Genome Editing for Systems Biology – A Temporal Perspective	367
	<i>Franziska Voellmy and Rune Linding</i>	
	Summary	367
14.1	Early Techniques in DNA Alterations	367
14.2	Zinc-Finger Nucleases	369
14.3	TALENs	369
14.4	CRISPR-Cas9	370
14.5	Considerations of Gene-Editing Nuclease Technologies	372
14.5.1	Repairing Nuclease-Induced DNA Damage	372
14.5.2	Nuclease Specificity	373
14.6	Applications	376
14.6.1	CRISPR Nuclease Genome-Wide Loss-of-Function Screens (CRISPRn)	377
14.6.2	CRISPR Interference: CRISPRi	378
14.6.3	CRISPR Activation: CRISPRa	378
14.6.4	Further Scalable Additions to the CRISPR-Cas Gene Editing Tool Arsenal	379
14.6.5	<i>In vivo</i> Applications	379
14.6.5.1	Animal Disease Models	379
14.6.5.2	Gene Therapy	379
14.7	A Focus on the Application of Genome-Engineering Nucleases on Chromosomal Rearrangements	380
14.7.1	Introduction to Chromosomal Rearrangements: The First Disease-Related Translocation	380
14.7.2	A Global Look at the Mechanisms behind Chromosomal Rearrangements	382
14.7.3	Creating Chromosomal Rearrangements Using CRISPR-Cas	383
14.8	Future Perspectives	384
	References	384
	Index	393

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

Gourvindu 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

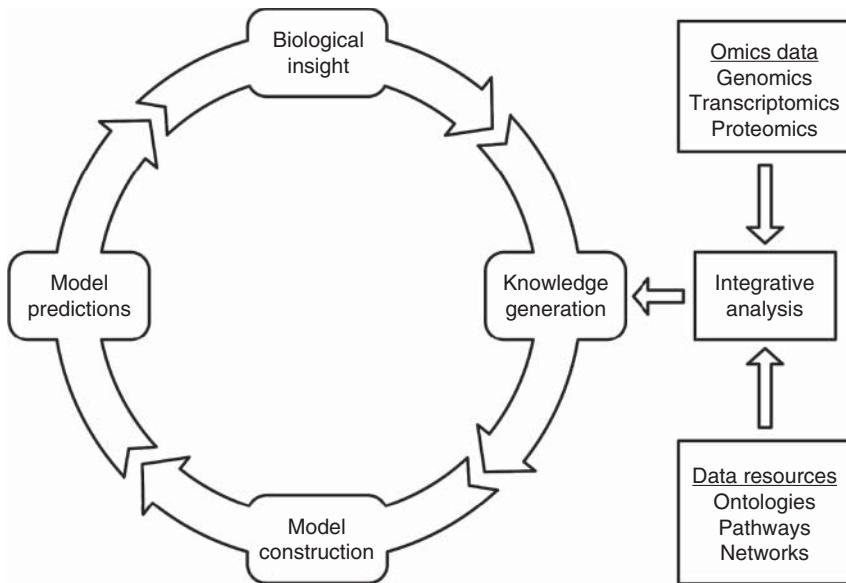


Figure 1.1 Systems biology research methodology. In the systems biology cycle, novel hypotheses are first formulated, which is followed by knowledge generation, model construction, and model predictions, which, in turn, leads to new biological insights. The development of high-throughput techniques have 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

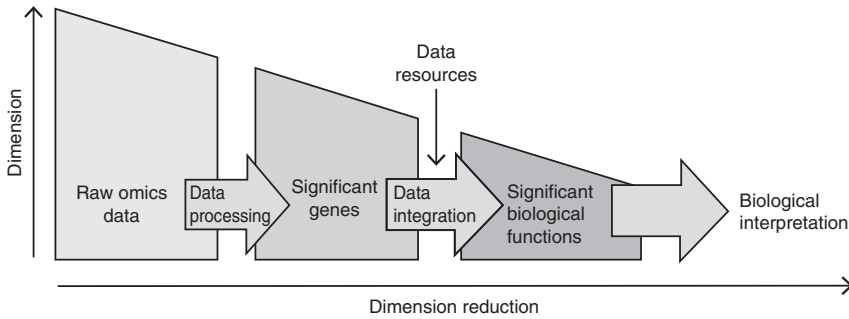


Figure 1.2 Description of the concept of integrative analysis as a tool for reduction of the dimension of omics data. Integrative analysis starts with raw omics data, which is typically affected by high levels of noise and errors. Computational and statistical approaches are first used to process the 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 throughout 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