Contents

Preface *XVII*
List of Contributors *XIX*

Part One **Modeling, Simulation, and Meaning of Gene Networks** *1*

1 **Network Analysis to Interpret Complex Phenotypes** *3*
Hong Yu, Jialiang Huang, Wei Zhang, and Jing-Dong J. Han
1.1 Introduction *3*
1.2 Identification of Important Genes based on Network Topologies *5*
1.2.1 Degree *5*
1.2.2 Betweenness *6*
1.2.3 Network Motifs *6*
1.2.4 Hierarchical Structure *7*
1.3 Inferring Information from Known Networks *8*
1.3.1 Understanding Biological Functions based on Network Modularity *8*
1.3.2 Inferring Functional Relationships and Novel Functional Genes Through Networks *8*
1.3.3 Unraveling Transcriptional Regulations from Expression Data through Transcriptional Networks *9*
1.3.4 Extracting the Pathway-Linked Regulators and Effectors based on Network Flows *10*
1.4 Conclusions *10*
References *11*

2 **Stochastic Modeling of Gene Regulatory Networks** *13*
Tianhai Tian
2.1 Introduction *13*
2.2 Discrete Stochastic Simulation Methods *14*
2.2.1 SSA *15*
2.2.2 Accelerating τ-Leap Methods *16*
2.2.3 Langevin Approach *19*

2.3 Discrete Stochastic Modeling *20*
2.3.1 Stochastic Modeling Method *20*
2.3.2 Toggle Switch with the SOS Pathway *22*
2.3.3 Other Models for the Genetic Toggle Switch *24*
2.4 Continuous Stochastic Modeling *26*
2.4.1 Deterministic Models for the λ Phage Network *26*
2.4.2 Stochastic Models for External Noise *28*
2.4.3 Deterministic Models with Threshold Values *29*
2.4.4 Stochastic Switching *30*
2.5 Stochastic Models for Both Internal and External Noise *31*
2.5.1 Stochastic Models for Microarray Gene Expression Data *33*
2.6 Conclusions *34*
References *35*

3 Modeling Expression Quantitative Trait Loci in Multiple Populations *39*
Ching-Lin Hsiao and Cathy S. J. Fann
3.1 Introduction *39*
3.1.1 Data Structure in eQTL Studies *39*
3.1.2 Current eQTL Studies *40*
3.1.2.1 eQTL Studies in a Single Human Population *40*
3.1.2.2 eQTL Studies in Multiple Human Populations *43*
3.1.3 An Illustrated Example *45*
3.1.4 Notations *46*
3.2 IGM Method *47*
3.2.1 Modeling SNP–GE Association in a Single Population *47*
3.2.2 Integrating Hypotheses to Identify Common eQTL *48*
3.2.3 Applying the IGM Method to HapMap Data *48*
3.2.3.1 Characterizing Putative eQTL Identified by the IGM *49*
3.3 CTWM *50*
3.3.1 Modeling SNP–GE Association in Pooled Data by CTWM *50*
3.3.2 Applying CTWM to HapMap Data *52*
3.3.2.1 Characterizing Putative eQTL Identified by CTWM *52*
3.3.2.2 Justification of Model Assumptions *53*
3.4 CTWM-GS Method *53*
3.4.1 Solving Normal Equations in CTWM *54*
3.4.2 Estimators of BD and GS *55*
3.4.3 Testing BD and GS *56*
3.4.4 Applying CTWM-GS to HapMap Data *56*
3.4.4.1 Applying the GS to Population Studies *57*
3.5 Discussion *60*
References *61*

Part Two Inference of Gene Networks *67*

4 Transcriptional Network Inference Based on Information Theory *69*
Patrick E. Meyer, Catharina Olsen, and Gianluca Bontempi
4.1 Introduction *69*
4.1.1 Notation *69*
4.1.2 Formalization *71*
4.1.3 Performance Measures in Undirected Network Inference *72*
4.1.3.1 Precision–Recall (PR) Curves *73*
4.1.3.2 *F*-Scores *74*
4.1.4 Causal Subset Selection *74*
4.2 Inference Based on Conditional Mutual Information *76*
4.2.1 Constraint-Based Methods *77*
4.2.2 Approximated Conditional Mutual Information *78*
4.2.3 Variable Selection Algorithms *78*
4.3 Inference Based on Pairwise Mutual Information *80*
4.3.1 Relevance Network (RELNET) *80*
4.3.2 Context Likelihood of Relatedness (CLR) *81*
4.3.3 Chow–Liu Tree *81*
4.3.4 Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNE) *82*
4.3.5 Minimum Redundancy Networks (MRNET) *83*
4.4 Arc Orientation *84*
4.4.1 Assessing Arc Orientation Methods *86*
4.5 Conclusions *87*
References *87*

5 Elucidation of General and Condition-Dependent Gene Pathways Using Mixture Models and Bayesian Networks *91*
Sandra Rodriguez-Zas and Younhee Ko
5.1 Introduction *91*
5.2 Methodology *92*
5.2.1 Network Learning Algorithms: Frequentist- and Bayesian MCMC-based algorithms *93*
5.3 Applications *95*
5.3.1 Elucidation of Gene Networks *95*
5.3.2 Discovery of Condition-Dependent Gene Relationships *96*
5.3.3 MCMC Mixture Bayesian Network *99*
5.3.4 Computational Considerations *101*
5.4 Conclusions *101*
References *102*

6 Multiscale Network Reconstruction from Gene Expression Measurements: Correlations, Perturbations, and "*A Priori* Biological Knowledge" *105*
Daniel Remondini and Gastone Castellani
6.1 Introduction *105*
6.1.1 Complex Networks *105*
6.1.2 Gene Interaction Networks from Gene Expression Measurements *107*
6.2 "Perturbation Method" *108*
6.3 Network Reconstruction by the Correlation Method from Time-Series Gene Expression Data *109*
6.4 Network Reconstruction from Gene Expression Data by *A Priori* Biological Knowledge *110*
6.5 Examples and Methods of Correlation Network Analysis on Time-Series Data *112*
6.6 Examples and Methods for Pathway Network Analysis *117*
6.6.1 Gene Selection and Pathway Grouping *118*
6.6.2 Pathway Significance and Pathway Network *118*
6.6.3 Results *121*
6.7 Discussion *126*
6.8 Conclusions *127*
References *128*

7 Gene Regulatory Networks Inference: Combining a Genetic Programming and H_∞ Filtering Approach *133*
Lijun Qian, Haixin Wang, and Xiangfang Li
7.1 Introduction *133*
7.2 Background *134*
7.2.1 Noise in Gene Expression *134*
7.2.2 Modeling of Gene Regulatory Networks with Noise *136*
7.2.2.1 Boolean Networks Model with Noise *136*
7.2.2.2 Bayesian Networks Model with Noise *136*
7.2.2.3 Linear Additive Regulation Model with Noise *137*
7.2.2.4 Neural Networks Model with Noise *137*
7.2.3 Proposed Nonlinear ODE Model with Noise *138*
7.3 Methodology for Identification and Algorithm Description *139*
7.3.1 GP *140*
7.3.2 H_∞ Filter *142*
7.4 Simulation Evaluation *144*
7.4.1 Synthetic Data *144*
7.4.2 Microarray Data *145*
7.5 Conclusions *146*
References *150*

8 Computational Reconstruction of Protein Interaction Networks *155*
Konrad Mönks, Irmgard Mühlberger, Andreas Bernthaler, Raul Fechete, Paul Perco, Rudolf Freund, Arno Lukas, and Bernd Mayer
8.1 Introduction *155*
8.1.1 Selecting Relevant Features from Omics Profiles *156*
8.1.2 Analyzing Omics Data on a Network Level *157*
8.2 Protein Interaction Networks *159*
8.2.1 Network Categories *159*
8.2.1.1 Metabolic Networks *159*
8.2.1.2 Paralog Networks *160*
8.2.1.3 Physical Interaction Networks *160*
8.2.2 Parameters for Protein Annotation *161*
8.2.2.1 Gene Expression Profiles *161*
8.2.2.2 Subcellular Location *161*
8.2.2.3 Gene Annotation *161*
8.2.2.4 Transcription Factors *162*
8.2.2.5 MicroRNA *162*
8.2.2.6 Pathways *162*
8.2.3 Data Preparation *163*
8.2.3.1 Integration of Data Sources *163*
8.2.3.2 Obtaining Edge Weights *164*
8.2.3.3 Data Completeness *166*
8.2.3.4 Data Normalization *166*
8.2.4 Deriving Models *166*
8.2.4.1 Basic Considerations *167*
8.2.4.2 Choosing an Algorithm *169*
8.2.5 Validation Procedures *169*
8.2.5.1 Model Performance Evaluation *169*
8.2.5.2 Network Structure Assessment *170*
8.3 Characterization of Computed Networks *171*
8.3.1 Evaluation of the Specific Protein–Protein Interactions *171*
8.3.2 Application of the Specific Protein–Protein Interactions *175*
8.4 Conclusions *177*
References *178*

Part Three Analysis of Gene Networks *181*

9 What if the Fit is Unfit? Criteria for Biological Systems Estimation Beyond Residual Errors *183*
Eberhard O. Voit
9.1 Introduction *183*
9.2 Model Design *184*
9.3 Concepts and Challenges of Parameter Estimation *187*
9.3.1 Typical Parameter Estimation Problems *190*
9.3.1.1 Data Fit is Unacceptable *190*

9.3.1.2 Differently Structured Candidate Models are Difficult to Compare *191*
9.3.1.3 Fit is Acceptable, But... *192*
9.3.1.4 Needed: A Better Fit! Or Not? *195*
9.4 Conclusions *197*
References *198*

10 Machine Learning Methods for Identifying Essential Genes and Proteins in Networks *201*
Kitiporn Plaimas and Rainer König
10.1 Introduction *201*
10.2 Definitions and Constructions of the Network *202*
10.3 Network Descriptors *203*
10.3.1 Network Topological Features for Undirected Graphs *204*
10.3.1.1 Connectivity *205*
10.3.1.2 Clustering Coefficient *205*
10.3.1.3 Centrality Measures *205*
10.3.2 Network Topological Features for a Bipartite Graph of Metabolic Networks *206*
10.3.2.1 Stoichiometric Properties *206*
10.3.2.2 Chokepoints *207*
10.3.2.3 Load Scores *207*
10.3.2.4 Deviations *207*
10.3.2.5 Damage in Global Networks *208*
10.4 Machine Learning *208*
10.5 Some Examples of Applications *210*
10.5.1 Validating an Experimental Knock-Out Screen *210*
10.5.2 Training with Data from One Organism to Predict Essential Genes for Another Organism *211*
10.5.3 Further Reported Investigations *211*
10.6 Conclusions *212*
References *213*

11 Gene Coexpression Networks for the Analysis of DNA Microarray Data *215*
Matthew T. Weirauch
11.1 Introduction *215*
11.2 Background *216*
11.2.1 Gene Transcription *216*
11.2.2 DNA Microarrays *217*
11.2.3 Networks *218*
11.3 Construction of GCNs *218*
11.3.1 Data Format and Representation *219*
11.3.2 Calculating Pairwise Gene Scores *219*
11.3.2.1 Overview *219*

11.3.2.2 Euclidean Distance *220*
11.3.2.3 Mutual Information *220*
11.3.2.4 Pearson's Correlation Coefficient *221*
11.3.2.5 Spearman's Rank Correlation Coefficient *221*
11.3.3 Choosing a Threshold *222*
11.3.3.1 Simple Thresholding Methods *222*
11.3.3.2 Fisher's *Z*-Transformation *222*
11.3.3.3 Permutation-Based Thresholds *223*
11.3.3.4 Other Methods *223*
11.4 Integration of GCNs with Other Data *224*
11.4.1 Integration of Multiple Expression Datasets *225*
11.4.1.1 Integrating Data within a Species *226*
11.4.1.2 Integrating Data across Species *226*
11.4.2 Integration of Heterogeneous Data Sources *227*
11.4.2.1 Union and Intersection-Based Methods *227*
11.4.2.2 Probabilistic Methods *228*
11.4.2.3 Integration of Expression Data with Specific Other Data Types *229*
11.5 Analysis of GCNs *230*
11.5.1 Network Hubs *231*
11.5.2 Network Motifs *232*
11.5.3 Gene Modules *232*
11.5.3.1 Gene Vector Clustering Methods for the Identification of Gene Modules *233*
11.5.3.2 Network-Based Methods for the Identification of Gene Modules *233*
11.6 GCNs for the Study of Cancer *237*
11.7 Conclusions *237*
References *238*

12 Correlation Network Analysis and Knowledge Integration *251*
Thomas N. Plasterer, Robert Stanley, and Erich Gombocz
12.1 Introduction *251*
12.2 Systems Biology Data Quandaries *252*
12.3 Semantic Web Approaches *252*
12.4 Correlation Network Analysis *253*
12.4.1 Selecting Nodes and Edges for Networks *255*
12.4.2 Distributions of Correlation Statistics *258*
12.5 Knowledge Annotation for Networks *259*
12.5.1 HRP and the Paired-Plaque Study *260*
12.5.2 Annotation with Public Sources and Ontologies *261*
12.5.3 Results and Benefits of the Approach *262*
12.5.3.1 Integral Informatics Approach *263*
12.6 Future Developments *274*
12.6.1 Improved Background Corrections *274*

12.6.2 Better Tools for Stratifying Key Observations *274*
12.6.3 Integration of Specialized Content: Chemical Structure and Images *275*
12.6.4 Expanded Sharing and Integration of Public Datasets *275*
12.6.5 Improved Integration of Text and Structured Data *276*
12.6.6 New Classes of Knowledge-Based Applications Such as Network Pattern Based Screening and Prediction *277*
References *278*

13 Network Screening: A New Method to Identify Active Networks from an Ensemble of Known Networks *281*
Shigeru Saito and Katsuhisa Horimoto
13.1 Introduction *281*
13.2 Methods *282*
13.2.1 Causal Graph *283*
13.2.2 Quantification of Consistency Between a Model and Data *284*
13.2.3 Statistical Methods for Evaluation *284*
13.2.3.1 GEV Distribution *284*
13.2.3.2 Nonparametric Statistic *285*
13.2.4 Simulation Study *286*
13.2.5 Network Screening *287*
13.3 Example Applications *289*
13.3.1 Evaluation of the *E. coli* SOS Network *289*
13.3.2 Network Screening for *E. coli* Networks Under Anaerobic Conditions *290*
13.3.3 Network Screening for ChIP networks on iPSCs *292*
13.4 Discussion *295*
References *295*

14 Community Detection in Biological Networks *299*
Gautam S. Thakur
14.1 Introduction *299*
14.2 Centrality Measures *300*
14.2.1 Degree Centrality *300*
14.2.2 Eigenvector Centrality *300*
14.2.3 Betweenness Centrality *301*
14.2.4 Closeness Centrality *301*
14.3 Study of Complex Systems *302*
14.4 Overview *302*
14.5 Proposed Algorithm *304*
14.5.1 Overview *304*
14.5.2 Algorithm Description *305*
14.5.3 Random Walk *306*
14.5.4 Local Optimization *308*
14.5.4.1 Vertex Ranking *308*

14.5.4.2 Community Ranking *310*
14.5.5 Mutual Exchange *311*
14.5.5.1 Dynamic Optimization *311*
14.6 Experiments *312*
14.6.1 Benchmarking *312*
14.6.1.1 Evaluation *312*
14.6.2 Computer-Generated Experiments *314*
14.6.3 Application to Real Networks *317*
14.6.3.1 Zachary Karate Club *318*
14.6.3.2 Neurotransmitter Receptor Complexes *319*
14.6.4 Study of Wireless Mobile Users *321*
14.7 Further Improvements *323*
14.8 Conclusions *324*
References *325*

15 On Some Inverse Problems in Generating Probabilistic Boolean Networks *329*
Xi Chen, Wai-Ki Ching, and Nam-Kiu Tsing
15.1 Introduction *329*
15.2 Reviews on BNs and PBNs *331*
15.2.1 BNs *331*
15.2.2 PBNs *332*
15.3 Construction of PBNs from a Prescribed Stationary Distribution *333*
15.3.1 CG Method Approach *334*
15.3.2 Estimation of $C_j^{(i)}$ *335*
15.3.3 Numerical Example *336*
15.3.4 Computational Cost Analysis *338*
15.4 Construction of PBNs from a Prescribed Transition Probability Matrix *338*
15.4.1 Heuristic Algorithms *339*
15.4.2 Numerical Demonstration *340*
15.4.3 Computational Cost Analysis *341*
15.4.4 Modifications of Algorithms 15.1 and 15.2 *341*
15.4.5 Numerical Example *342*
15.5 Conclusions *345*
References *346*

16 Boolean Analysis of Gene Expression Datasets *349*
Debashis Sahoo
16.1 Introduction *349*
16.2 Boolean Analysis *350*
16.3 Main Organization *351*
16.4 StepMiner *351*
16.5 StepMiner Algorithm *352*

16.5.1 Fitting One- or Two-Step Functions *352*
16.5.2 Selecting the Best Step Function *353*
16.5.3 Degrees of Freedom *354*
16.5.4 FDR *355*
16.6 BooleanNet *355*
16.6.1 Boolean Implications in Gene Expression Microarray Data *357*
16.6.2 Biological Interpretations of Boolean Implications *360*
16.6.3 Conserved Boolean Implications *361*
16.6.4 Comparison against Correlation Network *364*
16.6.5 Boolean Implication Networks are Not Scale-Free *365*
16.6.6 Computational Efficiency of BooleanNet *367*
16.7 BooleanNet Algorithm *368*
16.7.1 Data Collection and Preprocessing *368*
16.7.2 Discovery of Boolean Relationships *368*
16.7.3 Computation of FDR *371*
16.7.4 Correlation Network for Human CD Genes *371*
16.7.5 Discovery of Conserved Boolean Relationships *371*
16.7.6 Connected Component Analysis *371*
16.8 Conclusions *371*
References *373*

Part Four Systems Approach to Diseases *377*

17 Representing Cancer Cell Trajectories in a Phase-Space Diagram: Switching Cellular States by Biological Phase Transitions *379*
Mariano Bizzarri and Alessandro Giuliani
17.1 Introduction *379*
17.2 Beyond Reductionism *380*
17.3 Cell Shape as a Diagram of Forces *381*
17.4 Morphologic Phenotypes and Phase Transitions *382*
17.5 Cancer as an Anomalous Attractor *386*
17.6 Shapes as System Descriptors *388*
17.7 Fractals of Living Organisms *389*
17.8 Fractals and Cancer *390*
17.9 Modifications in Cell Shape Precede Tumor Metabolome Reversion *391*
17.10 Conclusions *395*
References *396*

18 Protein Network Analysis for Disease Gene Identification and Prioritization *405*
Jing Chen and Anil G. Jegga
18.1 Introduction *405*
18.2 Protein Networks and Human Disease *405*

18.2.1 Ranking Algorithms for Network Analyses *406*
18.2.2 Prioritization Methods *407*
18.2.3 Evaluation of Protein Interaction Network Topological Features *409*
18.3 ToppGene Suite of Applications *409*
18.4 Conclusions *410*
References *411*

19 Pathways and Networks as Functional Descriptors for Human Disease and Drug Response Endpoints *415*
Yuri Nikolsky, Marina Bessarabova, Eugene Kirillov, Zoltan Dezso, Weiwei Shi, and Tatiana Nikolskaya
19.1 Introduction *415*
19.2 Gene Content Classifiers and Functional Classifiers *416*
19.3 Biological Pathways and Networks Have Different Properties as Functional Descriptors *418*
19.4 Applications of Pathways as Functional Classifiers *420*
19.5 Single Pathway Learning for Identifying Functional Descriptor Pathways *425*
19.6 Multiple-Path Learning (MPL) Algorithm for Pathway Descriptors *427*
19.7 Applications of MPL-Deduced Pathway Descriptors *428*
19.7.1 Cross-Tissue Prediction of Compound Toxicity Using Functional Descriptors *428*
19.7.2 Cross-Platform Reproducibility Using Pathways (MPL Algorithms) *429*
19.7.3 Predictive Classifiers for Breast Cancer Endpoints *430*
19.7.4 Using Pathways as the Measure of Congruency Between Datasets *431*
19.8 Combining Advantages of Pathways and Networks *432*
19.9 Key Upstream and Downstream Interactions of Genetically Altered Genes and "Universal Cancer Genes" *435*
19.10 Conclusions *437*
References *438*

Index *443*