
Contents

Author and guest contributor biographies	xi
Acknowledgements	xv
Preface	xvii
1 Biomarkers and bioinformatics	1
1.1 Bioinformatics, translational research and personalized medicine	1
1.2 Biomarkers: fundamental definitions and research principles	2
1.3 Clinical resources for biomarker studies	5
1.4 Molecular biology data sources for biomarker research	6
1.5 Basic computational approaches to biomarker discovery: key applications and challenges	7
1.6 Examples of biomarkers and applications	10
1.7 What is next?	12
2 Review of fundamental statistical concepts	15
2.1 Basic concepts and problems	15
2.2 Hypothesis testing and group comparison	19
2.3 Assessing statistical significance in multiple-hypotheses testing	20
2.4 Correlation	23
2.5 Regression and classification: basic concepts	23
2.6 Survival analysis methods	26
2.7 Assessing predictive quality	28
2.8 Data sample size estimation	32
2.9 Common pitfalls and misinterpretations	34

3	Biomarker-based prediction models: design and interpretation principles	37
3.1	Biomarker discovery and prediction model development	37
3.2	Evaluation of biomarker-based prediction models	38
3.3	Overview of data mining and key biomarker-based classification techniques	40
3.4	Feature selection for biomarker discovery	47
3.5	Critical design and interpretation factors	52
4	An introduction to the discovery and analysis of genotype-phenotype associations	57
4.1	Introduction: sources of genomic variation	57
4.2	Fundamental biological and statistical concepts	60
4.3	Multi-stage case-control analysis	64
4.4	SNPs data analysis: additional concepts, approaches and applications	64
4.5	CNV data analysis: additional concepts, approaches and applications	68
4.6	Key problems and challenges	69
	Guest commentary on chapter 4: Integrative approaches to genotype-phenotype association discovery	73
	<i>Ana Dopazo</i>	
	References	76
5	Biomarkers and gene expression data analysis	77
5.1	Introduction	77
5.2	Fundamental analytical steps in gene expression profiling	79
5.3	Examples of advances and applications	82
5.4	Examples of the roles of advanced data mining and computational intelligence	84
5.5	Key limitations, common pitfalls and challenges	85
	Guest commentary on chapter 5: Advances in biomarker discovery with gene expression data	89
	<i>Haiying Wang, Huiru Zheng</i>	
	Unsupervised clustering approaches	90
	Module-based approaches	91
	Final remarks	92
	References	92
6	Proteomics and metabolomics for biomarker discovery: an introduction to spectral data analysis	93
6.1	Introduction	93
6.2	Proteomics and biomarker discovery	94
6.3	Metabolomics and biomarker discovery	97
6.4	Experimental techniques for proteomics and metabolomics: an overview	99
6.5	More on the fundamentals of spectral data analysis	100

6.6	Targeted and global analyses in metabolomics	101
6.7	Feature transformation, selection and classification of spectral data	102
6.8	Key software and information resources for proteomics and metabolomics	106
6.9	Gaps and challenges in bioinformatics	107
Guest commentary on chapter 6: Data integration in proteomics and metabolomics for biomarker discovery		111
<i>Kenneth Bryan</i>		
	Data integration and feature selection	112
	References	114
7	Disease biomarkers and biological interaction networks	115
7.1	Network-centric views of disease biomarker discovery	115
7.2	Basic concepts in network analysis	118
7.3	Fundamental approaches to representing and inferring networks	119
7.4	Overview of key network-driven approaches to biomarker discovery	120
7.5	Network-based prognostic systems: recent research highlights	124
7.6	Final remarks: opportunities and obstacles in network-based biomarker research	127
Guest commentary on chapter 7: Commentary on 'disease biomarkers and biological interaction networks'		131
<i>Zhongming Zhao</i>		
	Integrative approaches to biomarker discovery	132
	Pathway-based analysis of GWA data	133
	Integrative analysis of networks and pathways	134
	References	134
8	Integrative data analysis for biomarker discovery	137
8.1	Introduction	137
8.2	Data aggregation at the model input level	141
8.3	Model integration based on a single-source or homogeneous data sources	141
8.4	Data integration at the model level	144
8.5	Multiple heterogeneous data and model integration	145
8.6	Serial integration of source and models	148
8.7	Component- and network-centric approaches	151
8.8	Final remarks	152
Guest commentary on chapter 8: Data integration: The next big hope?		155
<i>Yves Moreau</i>		
	References	158
9	Information resources and software tools for biomarker discovery	159
9.1	Biomarker discovery frameworks: key software and information resources	159

9.2	Integrating and sharing resources: databases and tools	161
9.3	Data mining tools and platforms	166
9.4	Specialized information and knowledge resources	168
9.5	Integrative infrastructure initiatives and inter-institutional programmes	168
9.6	Innovation outlook: challenges and progress	169
10	Challenges and research directions in bioinformatics and biomarker discovery	173
10.1	Introduction	173
10.2	Better software	175
10.3	The clinical relevance of new biomarkers	176
10.4	Collaboration	177
10.5	Evaluating and validating biomarker models	178
10.6	Defining and measuring phenotypes	181
10.7	Documenting and reporting biomarker research	181
10.8	Intelligent data analysis and computational models	184
10.9	Integrated systems and infrastructures for biomedical computing	185
10.10	Open access to research information and outcomes	186
10.11	Systems-based approaches	187
10.12	Training a new generation of researchers for translational bioinformatics	188
10.13	Maximizing the use of public resources	189
10.14	Final remarks	189
	Guest commentary (1) on chapter 10: Towards building knowledge-based assistants for intelligent data analysis in biomarker discovery	193
	<i>Riccardo Bellazzi</i>	
	References	196
	Guest commentary (2) on chapter 10: Accompanying commentary on 'challenges and opportunities of bioinformatics in disease biomarker discovery'	197
	<i>Gary B. Fogel</i>	
	Introduction	197
	Biocytberinfrastructure	198
	Government Regulations on biomarker discovery	198
	Computational intelligence approaches for biomarker discovery	199
	Open source data, intellectual property, and patient privacy	199
	Conclusions	200
	References	200
	References	203
	Index	223