

Contents

<i>List of Contributors</i>	<i>page xi</i>
<i>Preface</i>	xv
1 An Introduction to High-Throughput Bioinformatics Data	
<i>Keith A. Baggerly, Kevin R. Coombes, and Jeffrey S. Morris</i>	1
1.1 Introduction	1
1.2 Microarrays	2
1.3 SAGE	19
1.4 Mass Spectrometry	24
1.5 Finding Data	34
2 Hierarchical Mixture Models for Expression Profiles	
<i>Michael A. Newton, Ping Wang, and Christina Kendziorski</i>	40
2.1 Introduction	40
2.2 Dual Character of Posterior Probabilities	43
2.3 Differential Expression as Independence	45
2.4 The Multigroup Mixture Model	47
2.5 Improving Flexibility	49
3 Bayesian Hierarchical Models for Inference in Microarray Data	
<i>Anne-Mette K. Hein, Alex Lewin, and Sylvia Richardson</i>	53
3.1 Introduction	53
3.2 Bayesian Hierarchical Modeling of Probe Level GeneChip Data	56
3.3 Bayesian Hierarchical Model for Normalization and Differential Expression	67
3.4 Predictive Model Checking	70

vi	<i>Contents</i>	
4	Bayesian Process-Based Modeling of Two-Channel Microarray Experiments: Estimating Absolute mRNA Concentrations <i>Mark A. van de Wiel, Marit Holden, Ingrid K. Glad, Heidi Lyng, and Arnaldo Frigessi</i>	75
4.1	Introduction	75
4.2	The Hierarchical Model	78
4.3	Reparameterization and Identifiability	82
4.4	MCMC-Based Inference	84
4.5	Validation	85
4.6	Illustration	85
4.7	TransCount Web Site and Computing Times	91
4.8	A Statistical Discussion of the Model	91
4.9	Discussion	93
5	Identification of Biomarkers in Classification and Clustering of High-Throughput Data <i>Mahlet G. Tadesse, Naijun Sha, Sinae Kim, and Marina Vannucci</i>	97
5.1	Introduction	97
5.2	Bayesian Variable Selection in Linear Models	100
5.3	Bayesian Variable Selection in Classification	101
5.4	Bayesian Variable Selection in Clustering via Finite Mixture Models	103
5.5	Bayesian Variable Selection in Clustering via Dirichlet Process Mixture Models	106
5.6	Example: Leukemia Gene Expression Data	108
5.7	Conclusion	113
6	Modeling Nonlinear Gene Interactions Using Bayesian MARS <i>Veerabhadran Baladandayuthapani, Chris C. Holmes, Bani K. Mallick, and Raymond J. Carroll</i>	116
6.1	Introduction	116
6.2	Bayesian MARS Model for Gene Interaction	118
6.3	Computation	121
6.4	Prediction and Model Choice	122
6.5	Examples	123
6.6	Discussion and Summary	131
7	Models for Probability of Under- and Overexpression: The POE Scale <i>Elizabeth Garrett-Mayer and Robert Scharpf</i>	137
7.1	POE: A Latent Variable Mixture Model	137
7.2	The POE Model	138
7.3	Unsupervised versus Semisupervised POE	144

<i>Contents</i>		vii
7.4	Using POE Scale	145
7.5	Example: POE as Applied to Lung Cancer Microarray Data	148
7.6	Discussion	152
8	Sparse Statistical Modelling in Gene Expression Genomics <i>Joseph Lucas, Carlos Carvalho, Quanli Wang, Andrea Bild, Joseph R. Nevins, and Mike West</i>	155
8.1	Perspective	156
8.2	Sparse Regression Modelling	157
8.3	Sparse Regression for Artifact Correction with Affymetrix Expression Arrays	162
8.4	Sparse Latent Factor Models and Latent Factor Regressions	167
8.5	Concluding Comments	173
9	Bayesian Analysis of Cell Cycle Gene Expression Data <i>Chuan Zhou, Jon C. Wakefield, and Linda L. Breeden</i>	177
9.1	Introduction	177
9.2	Previous Studies	178
9.3	Data	180
9.4	Bayesian Analysis of Cell Cycle Data	182
9.5	Discussion	197
10	Model-Based Clustering for Expression Data via a Dirichlet Process Mixture Model <i>David B. Dahl</i>	201
10.1	Introduction	201
10.2	Model	203
10.3	Inference	208
10.4	Simulation Study	209
10.5	Example	212
10.6	Conclusion	216
11	Interval Mapping for Expression Quantitative Trait Loci <i>Meng Chen and Christina Kendziorski</i>	219
11.1	Introduction	219
11.2	eQTL Mapping Experiments	221
11.3	QTL Mapping Methods	222
11.4	Currently Available eQTL Mapping Methods	223
11.5	MOM Interval Mapping	225
11.6	Discussion	231
12	Bayesian Mixture Models for Gene Expression and Protein Profiles <i>Michele Guindani, Kim-Anh Do, Peter Müller, and Jeffrey S. Morris</i>	238
12.1	Introduction	238

viii	<i>Contents</i>	
12.2	A Nonparametric Bayesian Model for Differential Gene Expression	240
12.3	A Mixture of Beta Model for MALDI-TOF Data	243
12.4	A Semiparametric Mixture Model for SAGE Data	247
12.5	Summary	250
13	Shrinkage Estimation for SAGE Data Using a Mixture Dirichlet Prior	
	<i>Jeffrey S. Morris, Keith A. Baggerly, and Kevin R. Coombes</i>	254
13.1	Introduction	254
13.2	Overview of SAGE	255
13.3	Methods for Estimating Relative Abundances	257
13.4	Mixture Dirichlet Distribution	260
13.5	Implementation Details	263
13.6	Simulation Study	264
13.7	Conclusion	267
14	Analysis of Mass Spectrometry Data Using Bayesian Wavelet-Based Functional Mixed Models	
	<i>Jeffrey S. Morris, Philip J. Brown, Keith A. Baggerly, and Kevin R. Coombes</i>	269
14.1	Introduction	270
14.2	Overview of MALDI-TOF	270
14.3	Functional Mixed Models	274
14.4	Wavelet-Based Functional Mixed Models	276
14.5	Analyzing Mass Spectrometry Data Using Wavelet-Based Functional Mixed Models	280
14.6	Conclusion	288
15	Nonparametric Models for Proteomic Peak Identification and Quantification	
	<i>Merlise A. Clyde, Leanna L. House, and Robert L. Wolpert</i>	293
15.1	Introduction	293
15.2	Kernel Models for Spectra	294
15.3	Prior Distributions	296
15.4	Likelihood	301
15.5	Posterior Inference	302
15.6	Illustration	303
15.7	Summary	305
16	Bayesian Modeling and Inference for Sequence Motif Discovery	
	<i>Mayetri Gupta and Jun S. Liu</i>	309
16.1	Introduction	309
16.2	Biology of Transcription Regulation	311

<i>Contents</i>		ix
16.3	Problem Formulation, Background, and General Strategies	312
16.4	A Bayesian Approach to Motif Discovery	316
16.5	Extensions of the Product-Multinomial Motif Model	320
16.6	HMM-Type Models for Regulatory Modules	321
16.7	Model Selection through a Bayesian Approach	327
16.8	Discussion: Motif Discovery Beyond Sequence Analysis	329
17	Identification of DNA Regulatory Motifs and Regulators by Integrating Gene Expression and Sequence Data <i>Deukwo Kwon, Sinae Kim, David B. Dahl, Michael Swartz, Mahlet G. Tadesse, and Marina Vannucci</i>	333
17.1	Introduction	333
17.2	Integrating Gene Expression and Sequence Data	335
17.3	A Model for the Identification of Regulatory Motifs	337
17.4	Identification of Regulatory Motifs and Regulators	340
17.5	Conclusion	344
18	A Misclassification Model for Inferring Transcriptional Regulatory Networks <i>Ning Sun and Hongyu Zhao</i>	347
18.1	Introduction	347
18.2	Methods	348
18.3	Simulation Results	355
18.4	Application to Yeast Cell Cycle Data	360
18.5	Discussion	361
19	Estimating Cellular Signaling from Transcription Data <i>Andrew V. Kossenkov, Ghislain Bidaut, and Michael F. Ochs</i>	366
19.1	Introduction	366
19.2	Bayesian Decomposition	370
19.3	Key Biological Databases	373
19.4	Example: Signaling Activity in <i>Saccharomyces cerevisiae</i>	376
19.5	Conclusion	380
20	Computational Methods for Learning Bayesian Networks from High-Throughput Biological Data <i>Bradley M. Broom and Devika Subramanian</i>	385
20.1	Introduction	385
20.2	Bayesian Networks	387
20.3	Learning Bayesian Networks	389
20.4	Algorithms for Learning Bayesian Networks	391
20.5	Example: Learning Robust Features from Data	395
20.6	Conclusion	398

21	Bayesian Networks and Informative Priors: Transcriptional Regulatory Network Models	
	<i>Alexander J. Hartemink</i>	401
	21.1 Introduction	401
	21.2 Bayesian Networks and Bayesian Network Inference	403
	21.3 Adding Informative Structure Priors	407
	21.4 Applications of Informative Structure Priors	409
	21.5 Adding Informative Parameter Priors	418
	21.6 Discussion	419
	21.7 Availability of Papers and Banjo Software	421
	21.8 Acknowledgments	421
22	Sample Size Choice for Microarray Experiments	
	<i>Peter Müller, Christian Robert, and Judith Rousseau</i>	425
	22.1 Introduction	425
	22.2 Optimal Sample Size as a Decision Problem	428
	22.3 Monte Carlo Evaluation of Predictive Power	431
	22.4 The Probability Model	432
	22.5 Pilot Data	435
	22.6 Example	435
	22.7 Conclusion	436