

Contents

Preface	v
Contributors	xvii
Color Insert	(facing page 236)

1 The Analysis of Gene Expression Data: An Overview of Methods and Software	1
<i>Giovanni Parmigiani, Elizabeth S. Garrett, Rafael A. Irizarry, and Scott L. Zeger</i>	
1.1 Measuring Gene Expression Using Microarrays	1
1.1.1 Microarray Technologies	1
1.1.2 Sources of Variation in Gene Expression Measurements Using Microarrays	4
1.1.3 Phases of Microarray Data Analysis	5
1.2 Design of Microarray Experiments	7
1.2.1 Replication and Sample Size Considerations	7
1.2.2 Design of Two-Channel Arrays	9
1.3 Data Storage	9
1.3.1 Databases	9
1.3.2 Standards	10
1.3.3 Statistical Analysis Languages	11
1.4 Preprocessing	12
1.4.1 Image Analysis	12
1.4.2 Visualizations for Quality Control	12
1.4.3 Background Subtraction	13
1.4.4 Probe-level Analysis of Oligonucleotide Arrays	14
1.4.5 Within-Array Normalization of cDNA Arrays	15
1.4.6 Normalization Across Arrays	15
1.5 Screening for Differentially Expressed Genes	16
1.5.1 Estimation or Selection?	16
1.5.2 One Problem or Many?	17
1.5.3 Selection and False Discovery Rates	18
1.5.4 Beyond Two Groups	19
1.6 Challenges of Genome Biometry Analyses	19

1.7	Visualization and Unsupervised Analyses	21
1.7.1	Profile Visualization	21
1.7.2	Why Clustering?	22
1.7.3	Hierarchical Clustering	23
1.7.4	k -Means Clustering and Self-Organizing Maps	25
1.7.5	Model-Based Clustering	26
1.7.6	Principal Components Analysis	26
1.7.7	Multidimensional Scaling	27
1.7.8	Identifying Novel Molecular Subclasses	27
1.7.9	Time Series Analysis	28
1.8	Prediction	29
1.8.1	Prediction Tools	29
1.8.2	Dimension Reduction	30
1.8.3	Evaluation of Classifiers	30
1.8.4	Regression-Based Approaches	31
1.8.5	Classification Trees	31
1.8.6	Probabilistic Model-Based Classification	32
1.8.7	Discriminant Analysis	33
1.8.8	Nearest-Neighbor Classifiers	33
1.8.9	Support Vector Machines	33
1.9	Free and Open-Source Software	33
1.9.1	Whitehead Institute Tools	34
1.9.2	Eisen Lab Tools	34
1.9.3	TIGR Tools	34
1.9.4	GeneX and CyberT	35
1.9.5	Projects at NCBI	35
1.9.6	BRB	35
1.9.7	The OOML library	36
1.9.8	MatArray	36
1.9.9	BASE	36
1.10	Conclusion	36

2 Visualization and Annotation of Genomic Experiments

46

Robert Gentleman and Vincent Carey

2.1	Introduction	46
2.2	Motivations for Component-Based Software	47
2.3	Formalism	49
2.4	Bioconductor Software for Filtering, Exploring, and Interpreting Microarray Experiments	50
2.4.1	Formal Data Structures and Methods for Multiple Microarrays	50
2.4.2	Tools for Filtering Gene Expression Data: The Closure Concept	54

2.4.3	Expression Density Diagnostics: High-Throughput Exploratory Data Analysis for Microarrays	55
2.4.4	Annotation	57
2.5	Visualization	58
2.5.1	Chromosomes	59
2.6	Applications	64
2.6.1	A Case Study of Gene Filtering	64
2.6.2	Application of Expression Density Diagnostics	67
2.7	Conclusions	70
3	Bioconductor R Packages for Exploratory Analysis and Normalization of cDNA Microarray Data	73
	<i>Sandrine Dudoit and Jean Yee Hwa Yang</i>	
3.1	Introduction	73
3.1.1	Overview of Packages	73
3.1.2	Two-Color cDNA Microarray Experiments	75
3.2	Methods	76
3.2.1	Standards for Microarray Data	76
3.2.2	Object-Oriented Programming: Microarray Classes and Methods	77
3.2.3	Diagnostic Plots	78
3.2.4	Normalization Using Robust Local Regression	79
3.3	Application: Swirl Microarray Experiment	80
3.4	Software	81
3.4.1	Package <code>marrayClasses</code> —Classes and Methods for cDNA Microarray Data	81
3.4.2	Package <code>marrayInput</code> —Data Input for cDNA Microarrays	89
3.4.3	Package <code>marrayPlots</code> —Diagnostic Plots for cDNA Microarray Data	91
3.4.4	Package <code>marrayNorm</code> —Location and Scale Normalization for cDNA Microarray Data	96
3.5	Discussion	99
4	An R Package for Analyses of Affymetrix Oligonucleotide Arrays	102
	<i>Rafael A. Irizarry, Laurent Gautier, and Leslie M. Cope</i>	
4.1	Introduction	102
4.2	Methods	103
4.2.1	Notation	103
4.2.2	The CEL/CDF Convention	104
4.2.3	Probe Pair Sets	106
4.2.4	Probe-Level Objects	107

4.2.5	Normalization	108
4.2.6	Exploratory Data Analysis of Probe-Level Data	111
4.3	Application	113
4.3.1	Expression Measures	113
4.4	Software	115
4.4.1	A Case Study	115
4.4.2	Extending the Package	118
4.5	Conclusion	118
5	DNA-Chip Analyzer (dChip)	120
	<i>Cheng Li and Wing Hung Wong</i>	
5.1	Introduction	120
5.2	Methods	121
5.2.1	Normalization of Arrays Based on an “Invariant Set”	121
5.2.2	Model-Based Analysis of Oligonucleotide Arrays	122
5.2.3	Confidence Interval for Fold Change	122
5.2.4	Pooling Replicate Arrays Considering Measurement Accuracy	124
5.3	Software and Applications	125
5.3.1	Reading in Array Data Files	125
5.3.2	Viewing an Array Image	127
5.3.3	Normalizing Arrays	129
5.3.4	Viewing PM/MM Data	129
5.3.5	Calculating Model-Based Expression Indexes	131
5.3.6	Filter Genes	132
5.3.7	Hierarchical Clustering	133
5.3.8	Comparing Samples	135
5.3.9	Mapping Genes to Chromosomes	137
5.3.10	Sample Classification by Linear Discriminant Analysis	138
5.4	Discussion	139
6	Expression Profiler	142
	<i>Jaak Vilo, Misha Kapushesky, Patrick Kemmeren, Ugis Sarkans, and Alvis Brazma</i>	
6.1	Introduction	142
6.2	EPCLUST	143
6.2.1	EPCLUST: Data Import	143
6.2.2	EPCLUST: Data Filtering	144
6.2.3	EPCLUST: Data Annotation	146
6.2.4	EPCLUST: Data Environment	147

6.2.5	EPCLUST: Data Analysis	148
6.3	URLMAP: Cross-Linking of the Analysis Results Between the Tools and Databases	151
6.4	EP:GO GeneOntology Browser	152
6.5	EP:PPI: Comparison of Protein Pairs and Expression . .	153
6.6	Pattern Discovery, Pattern Matching, and Visualization Tools	154
6.7	An Example of the Data Analysis and Visualizations Performed by the Tools in Expression Profiler	154
6.8	Integration of Expression Profiler with Public Microarray Databases	159
6.9	Conclusions	160
7	An S-PLUS Library for the Analysis and Visualization of Differential Expression	163
	<i>Jae K. Lee and Michael O'Connell</i>	
7.1	Introduction	163
7.2	Assessment of Differential Expression	164
7.2.1	Local Pooled Error	165
7.2.2	Tests for Differential Expression	169
7.2.3	Cluster Analysis and Visualization	171
7.3	Analysis of Melanoma Expression	174
7.3.1	Tests for Differential Expression	175
7.3.2	Cluster Analysis and Visualization	178
7.3.3	Annotation	180
7.4	Discussion	181
8	DRAGON and DRAGON View: Methods for the Annotation, Analysis, and Visualization of Large-Scale Gene Expression Data	185
	<i>Christopher M.L.S. Bouton, George Henry, Carlo Colantuoni, and Jonathan Pevsner</i>	
8.1	Introduction	185
8.2	System and Methods	189
8.2.1	Overview of DRAGON	189
8.2.2	DRAGON's Hardware, Software, and Database Architecture	190
8.2.3	Cross-Referencing Information in DRAGON . .	192
8.2.4	The DRAGON Search and Annotate Tools . .	193
8.2.5	The DRAGON View Data Visualization Tools .	196
8.2.6	DRAGON Gram: A Novel Visualization Tool . .	198
8.3	Implementation	199
8.4	Discussion and Conclusion	204

6.2.5	EPCLUST: Data Analysis	148
6.3	URLMAP: Cross-Linking of the Analysis Results Between the Tools and Databases	151
6.4	EP:GO GeneOntology Browser	152
6.5	EP:PPI: Comparison of Protein Pairs and Expression . .	153
6.6	Pattern Discovery, Pattern Matching, and Visualization Tools	154
6.7	An Example of the Data Analysis and Visualizations Performed by the Tools in Expression Profiler	154
6.8	Integration of Expression Profiler with Public Microarray Databases	159
6.9	Conclusions	160
7	An S-PLUS Library for the Analysis and Visualization of Differential Expression	163
	<i>Jae K. Lee and Michael O'Connell</i>	
7.1	Introduction	163
7.2	Assessment of Differential Expression	164
7.2.1	Local Pooled Error	165
7.2.2	Tests for Differential Expression	169
7.2.3	Cluster Analysis and Visualization	171
7.3	Analysis of Melanoma Expression	174
7.3.1	Tests for Differential Expression	175
7.3.2	Cluster Analysis and Visualization	178
7.3.3	Annotation	180
7.4	Discussion	181
8	DRAGON and DRAGON View: Methods for the Annotation, Analysis, and Visualization of Large-Scale Gene Expression Data	185
	<i>Christopher M.L.S. Bouton, George Henry, Carlo Colantuoni, and Jonathan Pevsner</i>	
8.1	Introduction	185
8.2	System and Methods	189
8.2.1	Overview of DRAGON	189
8.2.2	DRAGON's Hardware, Software, and Database Architecture	190
8.2.3	Cross-Referencing Information in DRAGON . .	192
8.2.4	The DRAGON Search and Annotate Tools . .	193
8.2.5	The DRAGON View Data Visualization Tools .	196
8.2.6	DRAGON Gram: A Novel Visualization Tool . .	198
8.3	Implementation	199
8.4	Discussion and Conclusion	204

17.4.4	Interpretation	402
17.4.5	Advantages of Bayesian Decomposition	403
17.5	Discussion	403
18	Bayesian Clustering of Gene Expression Dynamics	409
	<i>Paola Sebastiani, Marco Ramoni, and Isaac S. Kohane</i>	
18.1	Introduction	409
18.2	Methods	411
18.2.1	Modeling Time	412
18.2.2	Probabilistic Scoring Metric	413
18.2.3	Heuristic Search	415
18.2.4	Statistical Diagnostics	416
18.3	Software	417
18.3.1	Screen 0: Welcome Screen	417
18.3.2	Screen 1: Getting Started	418
18.3.3	Screen 2: Analysis	418
18.3.4	Screen 3: Cluster Model	419
18.3.5	Screen 4: Pack and Go!	419
18.4	Application	420
18.4.1	Analysis	420
18.4.2	Statistical Diagnostics	421
18.4.3	Understanding the Model	421
18.5	Conclusions	424
19	Relevance Networks: A First Step Toward Finding Genetic Regulatory Networks Within Microarray Data	428
	<i>Atul J. Butte and Isaac S. Kohane</i>	
19.1	Introduction	428
19.1.1	Advantages of Relevance Networks	429
19.2	Methodology	431
19.2.1	Formal Definition of Relevance Networks	431
19.2.2	Finding Regulatory Networks in Phenotypic Data	432
19.2.3	Using Entropy and Mutual Information to Evaluate Gene–Gene Associations	434
19.3	Applications	437
19.3.1	Finding Pharmacogenomic Regulatory Networks	437
19.3.2	Setting the Threshold	439
19.4	Software	440
	Index	447