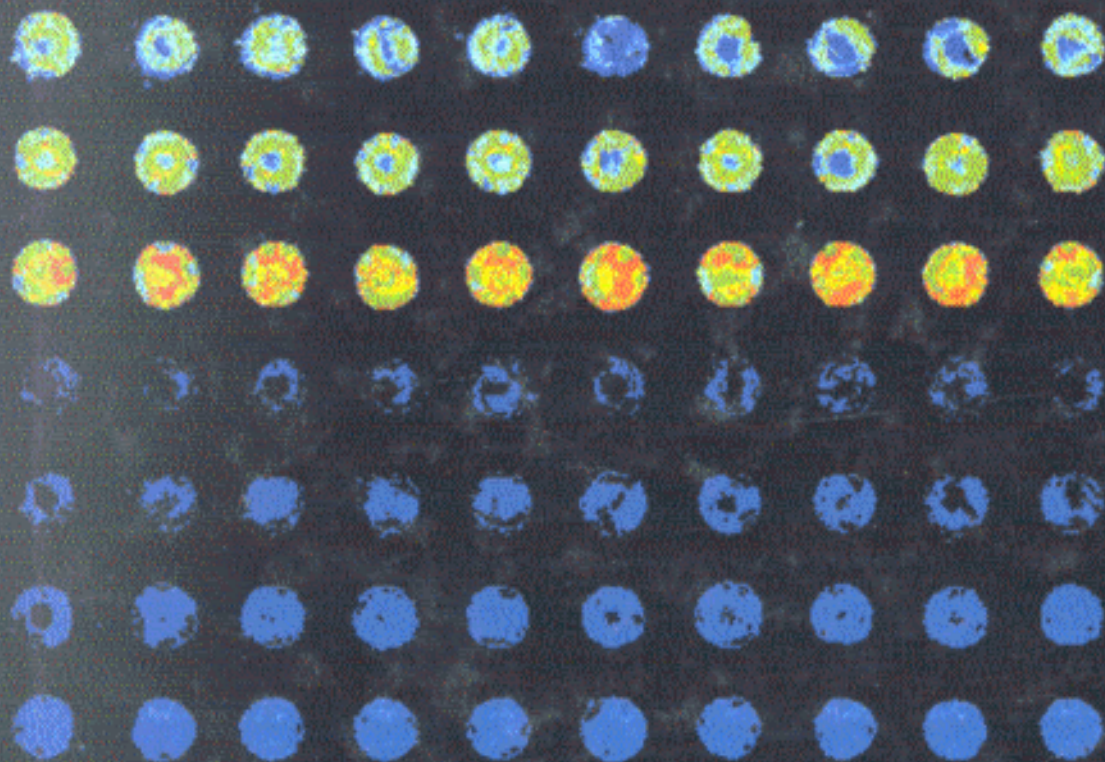


DNA Microarrays and Gene Expression

From Experiments to Data Analysis and Modeling

Pierre Baldi and G. Wesley Hatfield



CAMBRIDGE

Contents

<i>Preface</i>	<i>page</i> viii
1 A brief history of genomics	1
2 DNA array formats	7
<i>In situ</i> synthesized oligonucleotide arrays	7
Pre-synthesized DNA arrays	11
Filter-based DNA arrays	12
Non-conventional gene expression profiling technologies	13
3 DNA array readout methods	17
Reading data from a fluorescent signal	17
Reading data from a radioactive signal	21
4 Gene expression profiling experiments: Problems, pitfalls, and solutions	29
Primary sources of experimental and biological variation	29
Special considerations for gene expression profiling in bacteria	32
Problems associated with target preparation with polyadenylated mRNA from eukaryotic cells	42
A total RNA solution for target preparation from eukaryotic cells	44
Target cDNA synthesis and radioactive labeling for pre-synthesized DNA arrays	45
Data acquisition for nylon filter experiments	45
Data acquisition for Affymetrix GeneChip™ glass slide experiments	48
Normalization methods	49

5	Statistical analysis of array data: Inferring changes	53
	Problems and common approaches	53
	Probabilistic modeling of array data	55
	Simulations	65
	Extensions	69
6	Statistical analysis of array data: Dimensionality reduction, clustering, and regulatory regions	73
	Problems and approaches	73
	Visualization, dimensionality reduction, and principal component analysis	75
	Clustering overview	78
	Hierarchical clustering	82
	K-means, mixture models, and EM algorithms	85
	DNA arrays and regulatory regions	90
7	The design, analysis, and interpretation of gene expression profiling experiments	97
	Experimental design	99
	Identification of differentially expressed genes	100
	Determination of the source of errors in DNA array experiments	101
	Estimation of the global false positive level for a DNA array experiment	103
	Improved statistical inference from DNA array data using a Bayesian statistical framework	109
	Nylon filter data vs. Affymetrix GeneChip™ data	109
	Modeling probe pair set data from Affymetrix GeneChips™	124
	Application of clustering and visualization methods	124
	Identification of differential gene expression patterns resulting from two-variable perturbation experiments	125
8	Systems biology	135
	Introduction	135
	The molecular world: Representation and simulation	137
	Computational models of regulatory networks	146
	Software and databases	163
	The search for general principles	165

Contents

vii

Appendix A	Experimental protocols	177
Appendix B	Mathematical complements	185
Appendix C	Internet resources	195
Appendix D	CyberT: An online program for the statistical analysis of DNA array data	199
Index		207