

Contents

Foreword	xi
Preface	xv
Contributors	xvii
Glossary	xix
 SECTION I AN INTRODUCTION TO BIOINFORMATICS FOR THE GENETICIST	 1
1 Bioinformatics challenges for the geneticist	3
<i>Michael R. Barnes</i>	
1.1 Introduction	3
1.2 The role of bioinformatics in genetics research	4
1.3 Genetics in the post-genome era	5
1.4 Conclusions	12
References	15
 2 Managing and manipulating genetic data	 17
<i>Karl W. Broman and Simon C. Heath</i>	
2.1 Introduction	17
2.2 Basic principles	18
2.3 Data entry and storage	20
2.4 Data manipulation	21
2.5 Examples of code	22
2.6 Resources	30
2.7 Summary	31
References	31

SECTION II	MASTERING GENES, GENOMES AND GENETIC VARIATION DATA	33
3	The HapMap – A haplotype map of the human genome	35
	<i>Ellen M. Brown and Bryan J. Barratt</i>	
3.1	Introduction	35
3.2	Accessing the data	38
3.3	Application of HapMap data in association studies	42
3.4	Future perspectives	54
	References	54
4	Assembling a view of the human genome	59
	<i>Colin A. M. Semple</i>	
4.1	Introduction	59
4.2	Genomic sequence assembly	60
4.3	Annotation from a distance: the generalities	64
4.4	Annotation up close and personal: the specifics	70
4.5	Annotation: the next generation	78
	References	80
5	Finding, delineating and analysing genes	85
	<i>Christopher Southan and Michael R. Barnes</i>	
5.1	Introduction	85
5.2	Why learn to predict and analyse genes in the complete genome era?	86
5.3	The evidence cascade for gene products	88
5.4	Dealing with the complexities of gene models	95
5.5	Locating known genes in the human genome	97
5.6	Genome portal inspection	100
5.7	Analysing novel genes	101
5.8	Conclusions and prospects	102
	References	103
6	Comparative genomics	105
	<i>Martin S. Taylor and Richard R. Copley</i>	
6.1	Introduction	105
6.2	The genomic landscape	106
6.3	Concepts	109
6.4	Practicalities	113
6.5	Technology	118
6.6	Applications	132
6.7	Challenges and future directions	137
6.8	Conclusion	138
	References	139

SECTION III BIOINFORMATICS FOR GENETIC STUDY DESIGN AND ANALYSIS	145
7 Identifying mutations in single gene disorders	147
<i>David P. Kelsell, Diana Blaydon and Charles A. Mein</i>	
7.1 Introduction	147
7.2 Clinical ascertainment	147
7.3 Genome-wide mapping of monogenic diseases	148
7.4 The nature of mutation in monogenic diseases	152
7.5 Considering epigenetic effects in mendelian traits	160
7.6 Summary	162
References	162
8 From Genome Scan to Culprit Gene	165
<i>Ian C. Gray</i>	
8.1 Introduction	165
8.2 Theoretical and practical considerations	166
8.3 A stepwise approach to locus refinement and candidate gene identification	176
8.4 Conclusion	180
8.5 A list of the software tools and Web links mentioned in this chapter	181
References	182
9 Integrating Genetics, Genomics and Epigenomics to Identify Disease Genes	185
<i>Michael R. Barnes</i>	
9.1 Introduction	185
9.2 Dealing with the (draft) human genome sequence	186
9.3 Progressing loci of interest with genomic information	187
9.4 <i>In silico</i> characterization of the IBD5 locus – a case study	191
9.5 Drawing together biological rationale – hypothesis building	209
9.6 Identification of potentially functional polymorphisms	211
9.7 Conclusions	212
References	213
10 Tools for statistical genetics	217
<i>Aruna Bansal, Charlotte Vignal and Ralph McGinnis</i>	
10.1 Introduction	217
10.2 Linkage analysis	217
10.3 Association analysis	223
10.4 Linkage disequilibrium	229
10.5 Quantitative trait locus (QTL) mapping in experimental crosses	235
10.6 Closing remarks	239
References	241

SECTION IV	MOVING FROM ASSOCIATED GENES TO DISEASE ALLELES	247
11	Predictive functional analysis of polymorphisms: An overview	249
	<i>Mary Plumpton and Michael R. Barnes</i>	
11.1	Introduction	249
11.2	Principles of predictive functional analysis of polymorphisms	252
11.3	The anatomy of promoter regions and regulatory elements	256
11.4	The anatomy of genes	258
11.5	Pseudogenes and regulatory mRNA	266
11.6	Analysis of novel regulatory elements and motifs in nucleotide sequences	266
11.7	Functional analysis of non-synonymous coding polymorphisms	268
11.8	Integrated tools for functional analysis of genetic variation	273
11.9	A note of caution on the prioritization of <i>in silico</i> predictions for further laboratory investigation	275
11.10	Conclusions	275
	References	276
12	Functional <i>in silico</i> analysis of gene regulatory polymorphism	281
	<i>Chaolin Zhang, Xiaoyue Zhao, Michael Q. Zhang</i>	
12.1	Introduction	281
12.2	Predicting regulatory regions	282
12.3	Modelling and predicting transcription factor-binding sites	288
12.4	Predicting regulatory elements for splicing regulation	295
12.5	Evaluating the functional importance of regulatory polymorphisms	300
	References	302
13	Amino-acid properties and consequences of substitutions	311
	<i>Matthew J. Betts and Robert B. Russell</i>	
13.1	Introduction	311
13.2	Protein features relevant to amino-acid behaviour	312
13.3	Amino-acid classifications	316
13.4	Properties of the amino acids	318
13.5	Amino-acid quick reference	321
13.6	Studies of how mutations affect function	334
13.7	A summary of the thought process	339
	References	340
14	Non-coding RNA bioinformatics	343
	<i>James R. Brown, Steve Deharo, Barry Dancis, Michael R. Barnes and Philippe Sanseau</i>	
14.1	Introduction	343
14.2	The non-coding (nc) RNA universe	344
14.3	Computational analysis of ncRNA	349

14.4 ncRNA variation in disease	356
14.5 Assessing the impact of variation in ncRNA	362
14.6 Data resources to support small ncRNA analysis	363
14.7 Conclusions	363
References	364

SECTION V ANALYSIS AT THE GENETIC AND GENOMIC DATA INTERFACE **369**

15 What are microarrays? **371**

Catherine A. Ball and Gavin Sherlock

15.1 Introduction	371
15.2 Principles of the application of microarray technology	373
15.3 Complementary approaches to microarray analysis	377
15.4 Differences between data repository and research database	377
15.5 Descriptions of freely available research database packages	377
References	385

16 Combining quantitative trait and gene-expression data **389**

Elissa J. Chesler

16.1 Introduction: the genetic regulation of endophenotypes	389
16.2 Transcript abundance as a complex phenotype	390
16.3 Scaling up genetic analysis and mapping models for microarrays	394
16.4 Genetic correlation analysis	397
16.5 Systems genetic analysis	400
16.6 Using expression QTLs to identify candidate genes for the regulation of complex phenotypes	403
16.7 Conclusions	408
References	408

17 Bioinformatics and cancer genetics **413**

Joel Greshock

17.1 Introduction	413
17.2 Cancer genomes	414
17.3 Approaches to studying cancer genetics	415
17.4 General resources for cancer genetics	418
17.5 Cancer genes and mutations	420
17.6 Copy number alterations in cancer	425
17.7 Loss of heterozygosity in cancer	431
17.8 Gene-expression data in cancer	432
17.9 Multiplatform gene target identification	435
17.10 The epigenetics of cancer	438
17.11 Tumour modelling	438
17.12 Conclusions	439
References	439

18 Needle in a haystack? Dealing with 500 000 SNP genome scans	447
<i>Michael R. Barnes and Paul S. Derwent</i>	
18.1 Introduction	447
18.2 Genome scan analysis issues	449
18.3 Ultra-high-density genome-scanning technologies	459
18.4 Bioinformatics for genome scan analysis	469
18.5 Conclusions	489
References	490
19 A bioinformatics perspective on genetics in drug discovery and development	495
<i>Christopher Southan, Magnus Ulvsbäck and Michael R. Barnes</i>	
19.1 Introduction	495
19.2 Target genetics	498
19.3 Pharmacogenetics (PGx)	508
19.4 Conclusions: toward 'personalized medicine'	525
References	525
Appendix I	529
Appendix II	531
Index	537