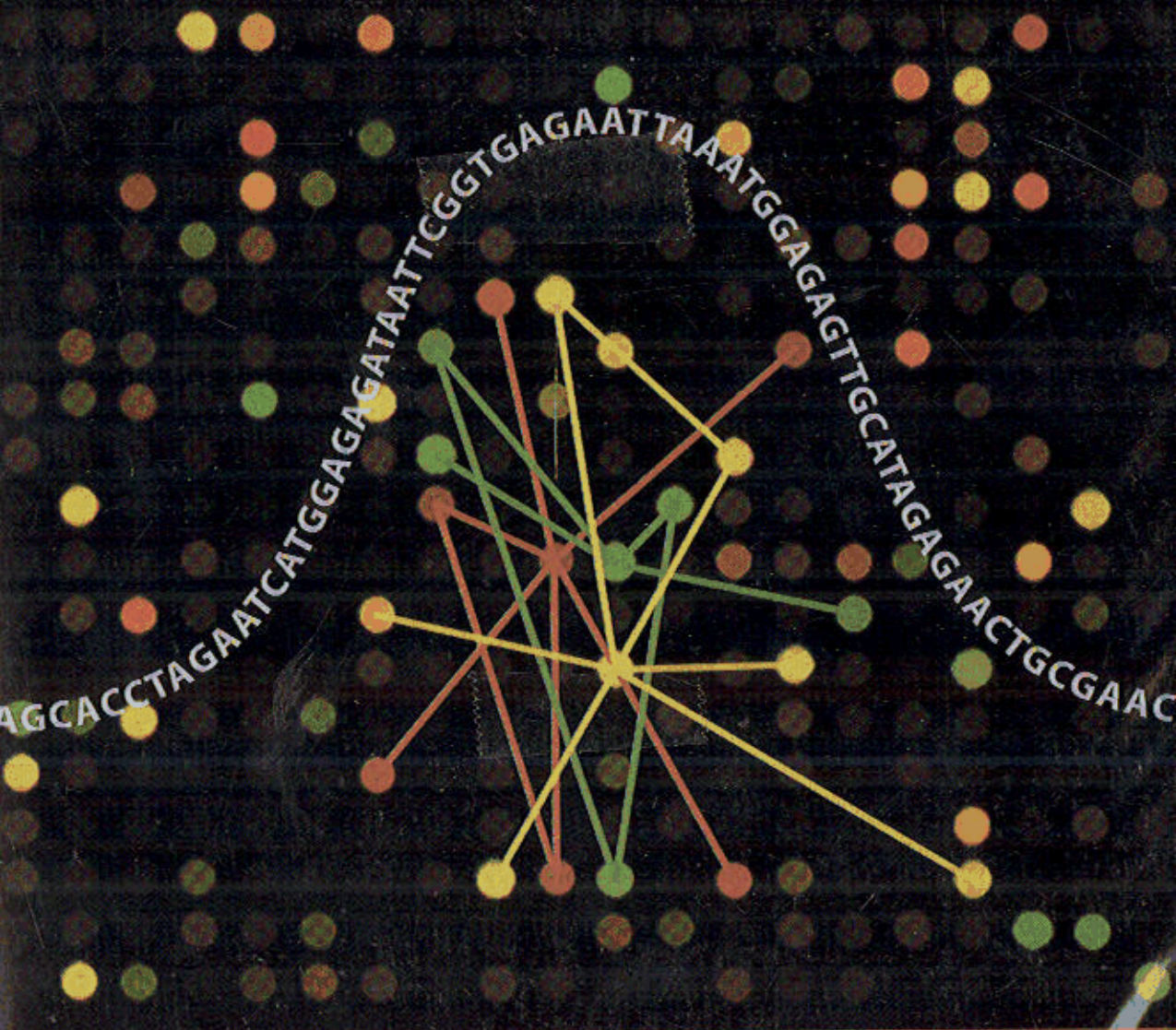


A PRIMER OF GENOME SCIENCE

SECOND EDITION



GREG GIBSON • SPENCER V. MUSE

Contents

Preface xi

1 Genome Projects: Organization and Objectives

The Core Aims of Genome Science 1

Mapping Genomes 4

Genetic Maps 4

EXERCISE 1.1 *Constructing a genetic map* 7

Physical Maps 8

Cytological Maps 8

Comparative Genomics 10

The Human Genome Project 13

Objectives 13

The Content of the Human Genome 16

BOX 1.1 *The Ethical, Legal, and Social Implications of the Human Genome Project* 18

Internet Resources 22

EXERCISE 1.2 *Use the NCBI and Ensembl genome browsers to examine a human disease gene* 23

Animal Genome Projects 25

BOX 1.2 *GenBank Files* 26

Rodent Genome Projects 29

EXERCISE 1.3 *Compare the structure of a gene in a mouse and a human* 32

Other Vertebrate Biomedical Models 33

Animal Breeding Projects 34

Invertebrate Model Organisms 35

BOX 1.3 *Managing and Distributing Genome Data* 38

Plant Genome Projects 41

Arabidopsis Thaliana 41

Grasses and Legumes 43

Other Flowering Plants 46

Microbial Genome Projects 47

The Minimal Genome 47

Sequenced Microbial Genomes 51

EXERCISE 1.4 *Compare two microbial genomes using the CMR* 52

Environmental Sequencing 53

Yeast 55

EXERCISE 1.5 *Examining a gene in the Saccharomyces Genome Database* 56

Parasite Genomics 57

Summary • Discussion Questions

• *Literature Cited* 59

2 Genome Sequencing and Annotation

Automated DNA Sequencing 65

The Principle of Sanger Sequencing 65

High-Throughput Sequencing 66

Reading Sequence Traces 68

EXERCISE 2.1 *Reading a sequence trace* 71

Contig Assembly 72

BOX 2.1 *Pairwise Sequence Alignment* 74

EXERCISE 2.2 *Computing an optimal sequence alignment* 78

Emerging Sequencing Methods 79

Genome Sequencing 81

Hierarchical Sequencing 82

Shotgun Sequencing 86

BOX 2.2 *Searching Sequence Databases Using BLAST* 88

Sequence Verification 92

GENOME ANNOTATION 93

EST Sequencing 93

Ab Initio Gene Discovery 96

BOX 2.3 *Hidden Markov Models and Gene Finding* 97

Regulatory Sequences 101

Non-Protein Coding Genes 102

Structural Features of Genome Sequences 105

Functional Annotation and Gene Family Clusters 110

EXERCISE 2.3 *Perform a BLAST search* 111

Clustering of Genes by Sequence Similarity 112

Clusters of Orthologous Genes 114

Phylogenetic Classification of Genes 115

BOX 2.4 *Phylogenetics* 116

EXERCISE 2.4 *A simple phylogenetic analysis* 120

Gene Ontology 122

BOX 2.5 *Gene Ontologies* 124

Summary • Discussion Questions

• *Web Site Exercises*

• *Literature Cited* 126

3 SNPs and Variation

The Nature of Single Nucleotide Polymorphisms 133

Classification of SNPs 133

Distribution of SNPs 136

Linkage Disequilibrium and Haplotype Maps 138

BOX 3.1 *Disequilibrium between Alleles at Two Loci* 140

EXERCISE 3.1 *Quantifying heterozygosity and LD* 143

Applications of SNP Technology 147

Population Genetics 147

BOX 3.2 *The Coalescent* 150

Recombination Mapping 153

EXERCISE 3.2 *Inferring haplotype structure* 154

QTL Mapping 156

Linkage Disequilibrium Mapping 159

BOX 3.3 *Case-Control Association Studies* 164

BOX 3.4 *Family-Based Association Tests* 168

EXERCISE 3.3 *Perform a case-control association test* 171

SNP Discovery 172

Resequencing 172

Sequence-Free Polymorphism Detection 174

SNP Genotyping 176

Low-Technology Methods 178

Minisequencing Methods 178

EXERCISE 3.4 *Designing a genotyping assay for a double polymorphism* 184

Homogeneous Fluorogenic Dye-Based Methods 184

Haplotype Phasing Methods 188

Summary • Discussion Questions

• *Web Site Exercises*

• *Literature Cited* 188

4 Gene Expression and the Transcriptome

Parallel Analysis of Gene

Expression: Microarrays 195

Applications of Microarray Technology 196

Experimental Design 198

EXERCISE 4.1 *Design a microarray experiment* 200

Microarray Technologies 202

Statistical Analysis of cDNA Microarray Data 211

BOX 4.1 *Microarray Image Processing* 214

EXERCISE 4.2 *Calculate which genes are differentially expressed* 216

BOX 4.2 *Basic Statistical Methods* 220

EXERCISE 4.3 *Evaluate the significance of the following gene expression differences* 223

Microarray Data Mining 224

BOX 4.3 *Clustering Methods* 226

EXERCISE 4.4 *Perform a cluster analysis on gene expression profiles* 229

ChIP Chips and Gene Regulation 230

SAGE and Microbeads 232

Serial Analysis of Gene Expression 232

BOX 4.4 *Motif Detection in Promoter Sequences* 234

Microbead-Based Expression Profiling 239

Single-Gene Analyses 239

Northern Blots 239

Quantitative PCR 241

Properties of Transcriptomes 242

Microbial Transcriptomics 243

Cancer and Clinical Applications 247

Development, Physiology, and Behavior 250

Evolutionary and Ecological Functional Genomics 253

Gene Expression Databases 255

Summary • Discussion Questions

• *Web Site Exercises*

• *Literature Cited* 258

5 Proteomics and Functional Genomics

Functional Proteomics 265

Protein Annotation 265

EXERCISE 5.1 *Structural annotation of a protein* 268

BOX 5.1 *Hidden Markov Models in Domain Profiling* 270

Protein Separation and 2D-PAGE 273

Mass Spectrometry and Protein Identification 277

EXERCISE 5.2 *Identification of a protein on the basis of a MS profile* 280

Protein Microarrays 283

Protein Interaction Maps 285

EXERCISE 5.3 *Formulating a network of protein interactions* 287

Structural Proteomics 289

Objectives of Structural Proteomics 289

BOX 5.2 *Biological Networks In Genome Science* 290

Protein Structure Determination 294

Protein Structure Prediction and Threading 297

Functional Genomics 300

Saturation Forward Genetics 301

High-Throughput Reverse Genetics 306

BOX 5.3 *Transgenic Animals and Plants* 312

Fine-Structure Genetics 314

EXERCISE 5.4 *Designing a genetic screen* 316

Genetic Fingerprinting 321

Summary • Discussion Questions

• *Web Site Exercises*

• *Literature Cited* 322

6 Integrative Genomics

Metabolomics 333

Analysis of Cellular Constituents 333

Metabolic and Biochemical Databases 337

In Silico Genomics 339

Metabolic Control Analysis 339

*Systems-Level Modeling of Gene
Networks 343*

Summary • Discussion Questions

• Literature Cited 348

Glossary 351

List of Abbreviations 363

Index 365