

## CONTENTS

Preface | xi

Acknowledgements | xiv

Design and conventions of this book | xvi

- 1 Introduction: working with the molecules of life in the computer | 1
  - Life on Earth and evolution | 2
  - The machinery of genetic information: more about DNA | 4
  - Genes and genomes | 7
  - Genes at work: transcription and translation | 8
  - Organization of the human genome | 11
  - Inferring products of DNA replication | 12
  - Inferring RNA products of transcription | 14
  - Inferring protein products of translation | 14
  - Exercises | 17
- 2 Gene technology: cutting DNA | 19
  - Early days of restriction enzymes | 20
  - Properties of restriction enzymes | 21
  - Pattern matching | 23
  - Identifying restriction enzyme sites with Perl | 24
  - Exercises | 29
- 3 Gene technology: knocking genes down | 31
  - Interfering with gene expression | 31
  - Small silencing RNAs | 32
  - RNAi: functions and applications | 34
  - Silencing RNAs and design principles | 35
  - Identifying siRNA candidates | 37
  - Exercises | 42
- 4 Gene technology: amplifying DNA | 44
  - What is PCR? | 44
  - PCR applications | 46
  - Primer design | 46

**CONTENTS**

- Reverse translation | 48  
Good manners during Perl programming | 51  
Exercises | 54
- 5** Human disease: when DNA sequences are toxic | 55  
Inherited disease and changes in DNA | 55  
Huntington's disease and CAG repeat expansion | 57  
Identifying mRNAs with CAG repeats | 59  
Exercises | 65
- 6** Human disease: iron imbalance and the iron responsive element | 66  
An inherited disease affecting the iron-binding protein ferritin | 66  
Many proteins of iron metabolism are regulated at the level of translation | 66  
Structure of the iron responsive element | 68  
Identifying the iron responsive element | 69  
Exercises | 72
- 7** Human disease: cancer as a result of aberrant proteins | 74  
Cancer as a genetic disease | 74  
Cancer and DNA repair | 75  
Chromosome rearrangements and the Philadelphia chromosome | 76  
Dotplots and alignments | 77  
BLAST | 81  
Using BLAST to examine the BCR-ABL fusion protein | 83  
Exercises | 90
- 8** Evolution: what makes us human? | 92  
Genetic differences between humans and chimpanzees | 92  
A protein related to human speech | 93  
FOXP2 in other animals | 94  
Comparing FOXP2 in different animals | 95  
Changes in FOXP2 specific to humans | 101  
Exercises | 102
- 9** Evolution: resolving a criminal case | 105  
A fatal injection | 105  
Methods of molecular phylogeny | 106  
Examination of the criminal case | 113  
Exercises | 120

- 10** Evolution: the sad case of the Tasmanian tiger | 121  
Extinction | 121  
The thylacine | 122  
Tiger history | 123  
Recent DNA analysis | 125  
Inferring the phylogeny of marsupials | 126  
Examining taxonomy | 129  
Exercises | 135
- 11** A function to every gene: termites, metagenomics and learning about the function of a sequence | 137  
Assigning function based on sequence similarity | 137  
Metagenomics | 138  
The other genome | 139  
Termites and cellulose digestion | 140  
Assigning function to termite sequences | 141  
Exercises | 147
- 12** A function to every gene: royal blood and order in the sequence universe | 150  
Royal disease | 150  
Blood-clotting pathways | 151  
Protein domain architecture | 153  
Bioinformatics of protein domains | 154  
Bioinformatic analysis of blood-clotting proteins | 156  
Evolution of blood clotting | 160  
Exercises | 162
- 13** A function to every gene: a slimy molecule | 164  
Extensive sugar decoration | 164  
Mucins and repeats | 165  
Computational identification of mucin domains | 168  
Exercises | 171
- 14** Information resources: learning about flu viruses | 173  
Short history of sequence databases | 173  
Features of nucleotide sequence databases | 175  
Comparative genomics | 177  
Protein sequence and structure data | 178  
Exploring databases at the NCBI | 180

## CONTENTS

- NCBI databases in eUtils | 181
- NCBI query syntax | 182
- The Entrez Programming Utilities | 185
- Parameters supplied to eUtils: scripts and construction of URLs | 185
- EFetch | 187
- ESearch | 188
- Further analysis of influenza viruses: extracting and filtering information with Perl and Unix tools | 192
- ELink | 195
- Exercises | 196
- NCBI documentation resources | 196
  
- 15** Finding genes: going ashore at CpG islands | 198
  - Transcription and its regulation | 198
  - DNA sequences that influence transcription | 199
  - CpG islands | 200
  - Finding CpG islands | 202
  - Exercises | 206
  
- 16** Finding genes: in the world of snurps | 208
  - Methods of gene prediction | 208
  - The splicing machinery | 209
  - Constructing a PSSM | 213
  - Scoring with a PSSM | 216
  - Exercises | 219
  
- 17** Finding genes: hunting for the distant RNA relatives | 222
  - The RNA world | 222
  - Properties of RNA and computational RNA finding | 223
  - An RNA involved in protein transport | 227
  - The organelles and their evolution | 227
  - The quest for chloroplast RNAs | 229
  - Automating tasks with Unix and Perl | 237
  - Exercises | 238
  
- 18** Personal genomes: the differences between you and me | 241
  - Personal genomes | 241
  - Individual variation and SNPs | 242
  - Counting SNPs | 244
  - Exercises | 250

<b>19</b>	<b>Personal genomes: what's in my genome?</b>	<b>252</b>
	Human roots	252
	An SNP dataset of South African individuals	253
	What SNPs are unique to the Bushmen?	256
	What SNPs are in coding regions?	257
	A bitter taste	260
	Constructing your own modules	263
	Exercises	264
<b>20</b>	<b>Personal genomes: details of family genetics</b>	<b>266</b>
	Basic principles of genetic inheritance	266
	Analysis of a family quartet	269
	Where are the crossing-over sites?	272
	Exercises	277
	Appendix I: brief Unix reference	278
	Appendix II: a selection of biological sequence analysis software	289
	Appendix III: a short Perl reference	300
	Appendix IV: a brief introduction to R	323
	Index	330