

Contents

Preface XIII

List of Contributors XVII

Part One Sanger DNA Sequencing I

1 Sanger DNA Sequencing 3

Artem E. Men, Peter Wilson, Kirby Siemering, and Susan Forrest

- 1.1 The Basics of Sanger Sequencing 3
- 1.2 Into the Human Genome Project (HGP) and Beyond 6
- 1.3 Limitations and Future Opportunities 7
- 1.4 Bioinformatics Holds the Key 8
- 1.5 Where to Next? 9
- References 10

Part Two Next-Generation Sequencing: Toward Personalized Medicine 13

2 Illumina Genome Analyzer II System 15

Abizar Lakdawalla and Harper VanSteenhouse

- 2.1 Library Preparation 15
- 2.2 Cluster Creation 17
- 2.3 Sequencing 19
- 2.4 Paired End Reads 19
- 2.5 Data Analysis 20
- 2.6 Applications 21
- 2.6.1 Genome Sequencing Applications 23
- 2.6.2 Epigenomics 23
- 2.6.3 Transcriptome Analysis 23
- 2.6.4 Protein–Nucleic Acid Interactions 26
- 2.6.5 Multiplexing 26

2.7	Conclusions	26
	References	27
3	Applied Biosystems SOLiD™ System: Ligation-Based Sequencing	29
	<i>Vicki Pandey, Robert C. Nutter, and Ellen Prediger</i>	
3.1	Introduction	29
3.2	Overview of the SOLiD™ System	29
3.2.1	The SOLiD Platform	30
3.2.1.1	Library Generation	30
3.2.1.2	Emulsion PCR	31
3.2.1.3	Bead Purification	31
3.2.1.4	Bead Deposition	33
3.2.1.5	Sequencing by Ligation	33
3.2.1.6	Color Space and Base Calling	35
3.3	SOLiD™ System Applications	35
3.3.1	Large-Scale Resequencing	35
3.3.2	<i>De novo</i> Sequencing	35
3.3.3	Tag-Based Gene Expression	36
3.3.4	Whole Transcriptome Analysis	37
3.3.5	Whole Genome Resequencing	38
3.3.6	Whole Genome Methylation Analysis	38
3.3.7	Chromatin Immunoprecipitation	39
3.3.8	MicroRNA Discovery	39
3.3.9	Other Tag-Based Applications	40
3.4	Conclusions	40
	References	41
4	The Next-Generation Genome Sequencing: 454/Roche GS FLX	43
	<i>Lei Du and Michael Egholm</i>	
4.1	Introduction	43
4.2	Technology Overview	44
4.3	Software and Bioinformatics	47
4.3.1	Whole Genome Assembly	47
4.3.2	Resequencing and Mutation Detection	47
4.3.3	Ultradeep Sequencing	47
4.4	Research Applications	49
	References	51
5	Polony Sequencing: History, Technology, and Applications	57
	<i>Jeremy S. Edwards</i>	
5.1	Introduction	57
5.2	History of Polony Sequencing	57
5.2.1	Introduction to Polonies	58
5.2.2	Evolution of Polonies	59
5.2.3	Current Applications of the Original Polonies Method	61

5.3	Polony Sequencing	62
5.3.1	Constructing a Sequencing Library	63
5.3.2	Loading the Library onto Beads Using BEAMing	64
5.3.3	Immobilizing the Beads in the Sequencing Flow Cell	65
5.3.4	Sequencing	66
5.3.5	Data Analysis	68
5.4	Applications	69
5.4.1	Human Genome Sequencing	69
5.4.1.1	Requirements of an Ultrahigh-Throughput Sequencing Technology	69
5.4.2	Challenges of Sequencing the Human Genome with Short Reads	70
5.4.2.1	Chromosome Sequencing	72
5.4.2.2	Exon Sequencing	72
5.4.2.3	Impact on Medicine	72
5.4.3	Transcript Profiling	73
5.4.3.1	Polony SAGE	73
5.4.3.2	Transcript Characterization with Polony SAGE	73
5.4.3.3	Digital Karyotyping	75
5.5	Conclusions	75
	References	76

Part Three The Bottleneck: Sequence Data Analysis 77

6	Next-Generation Sequence Data Analysis	79
	<i>Leonard N. Bloksberg</i>	
6.1	Why Next-Generation Sequence Analysis is Different?	79
6.2	Strategies for Sequence Searching	80
6.3	What is a "Hit," and Why it Matters for NGS?	82
6.3.1	Word Hit	82
6.3.2	Segment Hit	82
6.3.3	SeqID Hit or Gene Hit	82
6.3.4	Region Hit	82
6.3.5	Mapped Hit	83
6.3.6	Syntenic Hit	83
6.4	Scoring: Why it is Different for NGS?	83
6.5	Strategies for NGS Sequence Analysis	84
6.6	Subsequent Data Analysis	86
	References	87
7	DNASTAR's Next-Generation Software	89
	<i>Tim Durfee and Thomas E. Schwei</i>	
7.1	Personalized Genomics and Personalized Medicine	89
7.2	Next-Generation DNA Sequencing as the Means to Personalized Genomics	89

7.3	Strengths of Various Platforms	90
7.4	The Computational Challenge	90
7.5	DNASTAR's Next-Generation Software Solution	91
7.6	Conclusions	94
	References	94

Part Four Emerging Sequencing Technologies 95

8	Real-Time DNA Sequencing	97
	<i>Susan H. Hardin</i>	
8.1	Whole Genome Analysis	97
8.2	Personalized Medicine and Pharmacogenomics	97
8.3	Biodefense, Forensics, DNA Testing, and Basic Research	98
8.4	Simple and Elegant: Real-Time DNA Sequencing	98
	References	101
9	Direct Sequencing by TEM of Z-Substituted DNA Molecules	103
	<i>William K. Thomas and William Glover</i>	
9.1	Introduction	103
9.2	Logic of Approach	104
9.3	Identification of Optimal Modified Nucleotides for TEM Visual Resolution of DNA Sequences Independent of Polymerization	106
9.4	TEM Substrates and Visualization	107
9.5	Incorporation of Z-Tagged Nucleotides by Polymerases	108
9.6	Current and New Sequencing Technology	109
9.7	Accuracy	111
9.8	Advantages of ZSG's Proposed DNA Sequencing Technology	111
9.9	Advantages of Significantly Longer Read Lengths	112
9.9.1	<i>De novo</i> Genome Sequencing	112
9.9.2	Transcriptome Analysis	113
9.9.3	Haplotype Analysis	114
	References	115
10	A Single DNA Molecule Barcoding Method with Applications in DNA Mapping and Molecular Haplotyping	117
	<i>Ming Xiao and Pui-Yan Kwok</i>	
10.1	Introduction	117
10.2	Critical Techniques in the Single DNA Molecule Barcoding Method	118
10.3	Single DNA Molecule Mapping	120
10.3.1	Sequence Motif Maps of Lambda DNA	121
10.3.2	Identification of Several Viral Genomes	123

10.4	Molecular Haplotyping	124
10.4.1	Localization of Polymorphic Alleles Tagged by Single Fluorescent Dye Molecules Along DNA Backbones	125
10.4.2	Direct Haplotype Determination of a Human DNA Sample	127
10.5	Discussion	129
	References	131
11	Optical Sequencing: Acquisition from Mapped Single-Molecule Templates	133
	<i>Shiguo Zhou, Louise Pape, and David C. Schwartz</i>	
11.1	Introduction	133
11.2	The Optical Sequencing Cycle	135
11.2.1	Optical Sequencing Microscope and Reaction Chamber Setup	137
11.2.1.1	Microscope Setup	137
11.2.1.2	Optical Sequencing Reaction Chamber Setup	137
11.2.2	Surface Preparation	137
11.2.3	Genomic DNA Mounting/Overlay	139
11.2.4	Nicking Large Double-Stranded Template DNA Molecules	139
11.2.4.1	Nicking Mounted DNA Template Molecules	139
11.2.4.2	Gapping Nick Sites	139
11.2.5	Optical Sequencing Reactions	140
11.2.5.1	Basic Process	140
11.2.5.2	Choices of DNA Polymerases	140
11.2.5.3	Polymerase-Mediated Incorporations of Multiple Fluorochrome-Labeled Nucleotides	140
11.2.5.4	Washes to Remove Unincorporated Labeled Free Nucleotides and Reduce Background	141
11.2.6	Imaging Fluorescent Nucleotide Additions and Counting Incorporated Fluorochromes	141
11.2.7	Photobleaching	147
11.2.8	Demonstration of Optical Sequencing Cycles	147
11.3	Future of Optical Sequencing	148
	References	149
12	Microchip-Based Sanger Sequencing of DNA	153
	<i>Ryan E. Forster, Christopher P. Fredlake, and Annelise E. Barron</i>	
12.1	Integrated Microfluidic Devices for Genomic Analysis	154
12.2	Improved Polymer Networks for Sanger Sequencing on Microfluidic Devices	156
12.2.1	Poly(<i>N,N</i> -dimethylacrylamide) Networks for DNA Sequencing	156
12.2.2	Hydrophobically Modified Polyacrylamides for DNA Sequencing	159
12.3	Conclusions	160
	References	160

Part Five Next-Generation Sequencing: Truly Integrated Genome Analysis 165

- 13 Multiplex Sequencing of Paired End Ditags for Transcriptome and Genome Analysis 167**
Chia-Lin Wei and Yijun Ruan
 - 13.1 Introduction 167
 - 13.2 The Development of Paired End Ditag Analysis 168
 - 13.3 GIS-PET for Transcriptome Analysis 170
 - 13.4 ChIP-PET for Whole Genome Mapping of Transcription Factor Binding Sites and Epigenetic Modifications 173
 - 13.5 ChIA-PET for Whole Genome Identification of Long-Range Interactions 175
 - 13.6 Perspective 179
 - References 180
- 14 Paleogenomics Using the 454 Sequencing Platform 183**
M. Thomas P. Gilbert
 - 14.1 Introduction 183
 - 14.2 The DNA Degradation Challenge 184
 - 14.3 The Effects of DNA Degradation on Paleogenomics 185
 - 14.4 Degradation and Sequencing Accuracy 185
 - 14.5 Sample Contamination 189
 - 14.6 Solutions to DNA Damage 191
 - 14.7 Solutions to Contamination 192
 - 14.8 What Groundwork Remains, and What Does the Future Hold? 195
 - References 196
- 15 ChIP-seq: Mapping of Protein-DNA Interactions 201**
Anthony Peter Fejes and Steven J.M. Jones
 - 15.1 Introduction 201
 - 15.2 History 202
 - 15.3 ChIP-seq Method 202
 - 15.4 Sanger Dideoxy-Based Tag Sequencing 203
 - 15.5 Hybridization-Based Tag Sequencing 205
 - 15.6 Application of Sequencing by Synthesis 206
 - 15.7 Medical Applications of ChIP-seq 209
 - 15.8 Challenges 209
 - 15.9 Future Uses of ChIP-seq 211
 - References 213
- 16 MicroRNA Discovery and Expression Profiling using Next-Generation Sequencing 217**
Eugene Berezikov and Edwin Cuppen
 - 16.1 Background on miRNAs 217
 - 16.2 miRNA Identification 218

16.3	Experimental Approach	219
16.3.1	Sample Collection	219
16.3.2	Library Construction	221
16.3.3	Massively Parallel Sequencing	222
16.3.4	Bioinformatic Analysis	223
16.3.4.1	MicroRNA Discovery	223
16.3.4.2	miRNA Expression Profiling	225
16.4	Validation	225
16.5	Outlook	226
	References	226
17	DeepSAGE: Tag-Based Transcriptome Analysis Beyond Microarrays	229
	<i>Kåre L. Nielsen, Annabeth H. Petersen, and Jeppe Emmersen</i>	
17.1	Introduction	229
17.2	DeepSAGE	231
17.3	Data Analysis	235
17.4	Comparing Tag-Based Transcriptome Profiles	235
17.5	Future Perspectives	238
	References	239
18	The New Genomics and Personal Genome Information: Ethical Issues	245
	<i>jeantine E. Lunshof</i>	
18.1	The New Genomics and Personal Genome Information: Ethical Issues	245
18.2	The New Genomics: What Makes it Special?	245
18.3	Innovation in Ethics: Why do We Need it?	246
18.4	A Proviso: Global Genomics and Local Ethics	247
18.5	Medical Ethics and Hippocratic Confidentiality	247
18.6	Principles of Biomedical Ethics	248
18.7	Clinical Research and Informed Consent	248
18.8	Large-Scale Research Ethics: New Concepts	249
18.9	Personal Genomes	250
18.9.1	What is a Personal Genome and What is New About It?	250
18.9.2	But, Can Making Promises that Cannot be Substantiated be Ever Morally Justifiable?	251
18.10	The Personal Genome Project: Consenting to Disclosure	251
	References	252
	Index	255