

Contents

Preface *XIX*

List of Contributors *XXI*

Part One Tag-Based Nucleic Acid Analysis 7

1 **DeepSuperSAGE: High-Throughput Transcriptome Sequencing with Now- and Next-Generation Sequencing Technologies 3**

Hideo Matsumura, Carlos Molina, Detlev H. Krüger, Ryohei Terauchi, and Günter Kahl

- 1.1 Introduction 3
- 1.2 Overview of the Protocols 5
 - 1.2.1 Principle of the SuperSAGE Method 5
 - 1.2.2 Power of the SuperSAGE Tag 5
 - 1.2.3 Development of DeepSuperSAGE 6
 - 1.2.4 Ditag-Based DeepSuperSAGE (for 454 Pyrosequencing) 7
 - 1.2.5 Single-Tag-Based DeepSuperSAGE (HT-SuperSAGE) 8
- 1.3 Methods and Protocols 9
 - 1.3.1 Linker or Adapter Preparation 9
 - 1.3.2 RNA Samples 10
 - 1.3.3 cDNA Synthesis and NlaIII Digestion 10
 - 1.3.4 Tag Extraction from cDNA 10
 - 1.3.5 Tag Extraction from cDNA 11
 - 1.3.6 Purification of Linker–Tag Fragments 12
 - 1.3.7 Ditag or Adapter–Tag Formation and Amplification 12
 - 1.3.8 Preparation of Templates for Sequencing 14
- 1.4 Applications 14
 - 1.4.1 Applications of DeepSuperSAGE in Combination with 454 Pyrosequencing 14
 - 1.4.2 Practical Analysis of HT-SuperSAGE 18
- 1.5 Perspectives 19
- References 20

2 **DeepCAGE: Genome-Wide Mapping of Transcription Start Sites 23**

Matthias Harbers, Mitchell S. Dushay, and Piero Carninci

- 2.1 Introduction 23
- 2.2 What is CAGE? 24
- 2.3 Why CAGE? 26
- 2.4 Methods and Protocols 28
 - 2.4.1 Key Reagents and Consumables 28
 - 2.4.2 Precautions 30

2.4.3	RNA Samples Used for DeepCAGE Library Preparation	30
2.4.4	DeepCAGE Library Preparation	32
2.5	Applications	43
2.6	Perspectives	44
	References	45
3	Definition of Promotome–Transcriptome Architecture Using CAGEscan	47
	<i>Nicolas Bertin, Charles Plessy, Piero Carninci, and Matthias Harbers</i>	
3.1	Introduction	47
3.2	What is CAGEscan?	48
3.3	Why CAGEscan?	50
3.4	Methods and Protocols	51
3.4.1	Key Reagents and Consumables	51
3.4.2	Precautions	53
3.4.3	RNA Samples Used for CAGEscan Library Preparation	53
3.4.4	Considerations on Pooling CAGEscan Libraries	54
3.4.5	CAGEscan Library Preparation	54
3.5	Applications and Perspectives	59
	References	61
4	RACE: New Applications of an Old Method to Connect Exons	63
	<i>Charles Plessy</i>	
4.1	Introduction	63
4.2	Deep-RACE	65
4.2.1	Choice of the Sequencer	65
4.2.2	Validation of Promoter Studies	65
4.2.3	Other Applications of Deep-RACE	66
4.2.4	Limitations of Deep-RACE	66
4.3	Methods Outline	67
4.3.1	Primer Design	67
4.3.2	Molecular Biology of Deep-RACE Library Preparation	67
4.3.3	Sequencing of Deep-RACE Libraries	68
4.3.4	Analysis	68
4.4	Perspectives	70
	References	71
5	RNA-PET: Full-Length Transcript Analysis Using 5'- and 3'-Paired-End Tag Next-Generation Sequencing	73
	<i>Xiaoan Ruan and Yijun Ruan</i>	
5.1	Introduction	73
5.2	Methods and Protocols	75
5.2.1	Key Reagents and Consumables	75
5.2.2	Protocol	78
5.3	Applications	88
5.3.1	PET Sequencing with SOLiD	88
5.3.2	Mapping of the PETs	88
5.3.3	PET Clustering, Annotation, and Genome Browser Visualization	89
5.4	Perspectives	90
	References	90
6	Stranded RNA-Seq: Strand-Specific Shotgun Sequencing of RNA	97
	<i>Alistair R.R. Forrest</i>	
6.1	Introduction	97
6.1.1	Before Starting	93
6.2	Methods and Protocols	93
6.2.1	Preface	93
6.2.2	Materials and Consumables	94

6.2.3	Protocol	95
6.3	Bioinformatic Considerations	103
6.4	Applications	104
6.5	Perspectives	105
	References	107
7	Differential RNA Sequencing (dRNA-Seq): Deep-Sequencing-Based Analysis of Primary Transcriptomes	109
	<i>Anne Borries, Jörg Vogel, and Cynthia M. Sharma</i>	
7.1	Introduction	109
7.2	What is dRNA-Seq?	111
7.3	Why dRNA-Seq?	112
7.4	Methods and Protocols	115
7.4.1	Materials and Consumables	115
7.4.2	Precautions	116
7.4.3	RNA Samples Used for dRNA-Seq Library Preparation	116
7.4.4	dRNA-Seq Library Preparation	116
7.5	Applications	119
7.6	Perspectives	120
	References	121
8	Identification and Expression Profiling of Small RNA Populations Using High-Throughput Sequencing	123
	<i>Javier Armisen, W. Robert Shaw, and Eric A. Miska</i>	
8.1	Introduction	123
8.1.1	miRNAs	123
8.1.2	piRNAs	125
8.1.3	siRNAs	126
8.1.4	Other Small RNAs	126
8.2	HTS/NGS	127
8.3	Methods and Protocols	128
8.3.1	Key Reagents and Solutions	128
8.3.2	Total RNA Isolation	129
8.3.3	Small RNA Isolation	129
8.3.4	Ligation of Adapters	131
8.4	Troubleshooting	134
8.5	Applications	134
8.6	Perspectives	136
	References	138
9	Genome-Wide Mapping of Protein–DNA Interactions by ChIP-Seq	139
	<i>Joshua W.K. Ho, Artyom A. Alekseyenko, Mitzi I. Kuroda, and Peter J. Park</i>	
9.1	Introduction	139
9.2	Methods and Protocols	141
9.2.1	Antibody Validation	141
9.2.2	ChIP	141
9.2.3	Sequencing Library Preparation	144
9.2.4	Data Analysis	146
9.3	Applications	147
9.3.1	Deciphering the Transcriptional Regulatory Program	148
9.3.2	Unraveling Epigenetic Regulation	148
9.3.3	Comparative Interindividual or Interspecies Analysis	149
9.3.4	Study of Human Diseases and Clinical Applications	149
9.3.5	Advantages and Challenges of ChIP-Seq	149
9.4	Perspectives	150
	References	151

10	Analysis of Protein–RNA Interactions with Single-Nucleotide Resolution Using iCLIP and Next-Generation Sequencing 153
	<i>Julian König, Nicholas J. McGlincy, and Jernej Ule</i>
10.1	Introduction 153
10.2	Procedure Overview 154
10.3	Antibody and Library Preparation Quality Controls 155
10.4	Oligonucleotide Design 156
10.5	Recent Modifications of the iCLIP Protocol 158
10.6	Troubleshooting 158
10.7	Methods and Protocols 159
	References 169
11	Massively Parallel Tag Sequencing Unveils the Complexity of Marine Protistan Communities in Oxygen-Depleted Habitats 171
	<i>Virginia Edgcomb and Thorsten Stoeck</i>
11.1	Introduction 171
11.2	Cariaco Basin 173
11.3	Framvaren Fjord 176
11.4	Comparison of Cariaco Basin to Framvaren Fjord 177
11.5	Perspectives on Interpretation of Microbial Eukaryote 454 Data 179
	References 182
12	Chromatin Interaction Analysis Using Paired-End Tag Sequencing (ChIA-PET) 185
	<i>Xiaolan Ruan and Yijun Ruan</i>
12.1	Introduction 185
12.1.1	Development of the ChIA-PET Method 186
12.1.2	Applications of the ChIA-PET Method 187
12.1.3	Experimental Design of ChIA-PET Analysis 187
12.1.3.1	ChIP Sample Preparation 187
12.1.3.2	ChIA-PET Library Construction 189
12.1.3.3	ChIA-PET Library Sequencing and Mapping 190
12.1.3.4	Control Libraries 191
12.2	Methods and Protocols 192
12.2.1	Key Reagents and Consumables 192
12.2.2	Protocol 195
12.3	Timeline 206
12.4	Anticipated Results 207
12.4.1	Verification of Sonicated Chromatin DNA Size Range 207
12.4.2	ChIP Quality Control: Yield and Enrichment 207
12.4.3	ChIA-PET Library Quality Control 207
12.4.4	ChIA-PET Sequencing and Mapping Analysis 207
12.5	Perspectives 209
	References 209
13	Tag-Seq: Next-Generation Tag Sequencing for Gene Expression Profiling 211
	<i>Sorana Morrissy, Yongjun Zhao, Allen Delaney, Jennifer Asano, Noreen Dhalla, Irene Li, Helen McDonald, Pawan Pandoh, Anna-Liisa Prabhu, Angela Tam, Martin Hirst, and Marco Marra</i>
13.1	Introduction 211
13.2	Protocol Details 212
13.3	Protocol Overview and Timeline 213
13.4	Critical Parameters and Troubleshooting 214
13.5	Methods and Protocols 215
13.5.1	Basic Protocol 1: First- and Second-Strand cDNA Synthesis for Tag-Seq Library Construction 215

13.5.2	Basic Protocol 2: Tag Generation	219
13.5.3	Basic Protocol 3: PCR and Fragment Isolation	223
13.5.4	Basic Protocol 4: Preparing the Library for Illumina Sequencing	226
13.5.5	Alternate Protocol: Amplified Tag-Seq library construction (Tag-SeqLite)	227
13.5.6	Basic Protocol 5: Data Analysis	232
13.6	Applications	239
13.7	Perspectives	240
	References	241
14	Isolation of Active Regulatory Elements from Eukaryotic Chromatin Using FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements)	243
	<i>Paul G. Giresi and Jason D. Lieb</i>	
14.1	Introduction	243
14.2	Methods and Protocols	245
14.2.1	FAIRE Procedure	245
14.2.2	Optimization of the FAIRE Procedure	245
14.2.3	Equipment and Reagents	246
14.2.4	Detection of FAIRE DNA	250
14.2.5	High-Throughput Sequencing	252
14.3	Applications	254
14.4	Perspectives	254
	References	255
15	Identification of Nucleotide Variation in Genomes Using Next-Generation Sequencing	257
	<i>Hendrik-Jan Megens and Martien A.M. Groenen</i>	
15.1	Introduction	257
15.1.1	SNP Discovery and Nucleotide Variation Assessment	259
15.1.2	Sequence and Library Preparation Strategies	259
15.2	Methods	261
15.2.1	Preprocessing of Reads	262
15.2.1.1	FASTQ Format	262
15.2.1.2	FASTQ Format – Illumina Version	262
15.2.1.3	Illumina FASTQ to Sanger FASTQ	263
15.2.1.4	ABI SOLiD- and Roche 454-Specific Formats	264
15.2.1.5	Illumina SCARF or QSEQ to FASTQ	264
15.2.1.6	Quality Evaluation	265
15.2.1.7	Handling Adapter Sequences – Linkers and Barcodes	265
15.2.1.8	Quality Trimming	266
15.2.2	Mapping Reads to a Reference Genome	266
15.2.2.1	Making Alignments Using MOSAIK	267
15.2.2.2	Making Alignments Using BWA	268
15.2.3	Variant Calling	269
15.2.3.1	SAM Format	270
15.2.3.2	Variant Calling with SAMtools	273
15.2.3.3	Variant Calling with GATK	274
15.3	Notes	275
	References	275
16	DGS (Ditag Genome Scanning) – A Restriction-Based Paired-End Sequencing Approach for Genome Structural Analysis	277
	<i>Jun Chen, Yeong C. Kim, and San Ming Wang</i>	
16.1	Introduction	277
16.2	Methods and Protocols	278
16.2.1	Cloning-Based DGS Protocol	278
16.2.2	Non-Cloning-Based DGS Protocol	281

16.2.3	Computational Mapping Analysis of Experimental Datasets	282
16.3	Applications	283
16.3.1	Analyzing Normal Genome Structure	283
16.3.2	Identifying Somatic Rearrangements in Cancer Genomes	283
16.3.3	A Useful Tool to Study Family Germline Genetic Disorders	284
16.4	Perspectives	284
	References	285
17	Next-Generation Sequencing of Bacterial Artificial Chromosome Clones for Next-Generation Physical Mapping	287
	<i>Robert Bogden, Keith Stormo, Jason Dobry, Amy Mraz, Quanzhou Tao, Michiel van Eijk, Jan van Oeveren, Marcel Prins, Jon Wittendorp, and Mark van Haaren</i>	
17.1	History of the Bacterial Artificial Chromosome Vector Systems	287
17.2	History of Physical Mapping	288
17.3	What is WGP?	289
17.4	Flow of a WGP Project	289
17.5	BAC Pooling Strategies	290
17.6	Methods and Protocols	291
17.6.1	BAC Library and Pooling Strategy	291
17.6.2	Sample Preparation for Illumina Sequencing	292
17.6.3	Illumina Sequencing	293
17.6.4	Deconvolution to Assign the BAC Address to Each Read	293
17.6.5	Contig Building	293
17.7	Applications	294
17.7.1	Results from Real WGP Projects Performed by the Authors	294
17.7.2	Reorganizing Project Funding and Sequencing Budgets	295
17.7.3	Unleash the Power of BAC Clones	296
17.8	Perspectives	296
	References	297
18	HELP-Tagging: Tag-Based Genome-Wide Cytosine Methylation Profiling	299
	<i>Masako Suzuki and John M. Greally</i>	
18.1	Introduction	299
18.2	Genome-Wide DNA Methylation Analysis	299
18.3	What is HELP-Tagging?	300
18.3.1	When is HELP-Tagging the Preferred Cytosine Methylation Assay?	301
18.4	Methods and Protocols	301
18.4.1	Reagents, Materials, and Equipment	301
18.4.2	Buffers and Adapters for HELP-Tagging Library Preparation	302
18.4.3	Precautions	303
18.4.4	DNA Samples for HELP-Tagging Library Preparation	303
18.4.5	HELP-Tagging Library Preparation	304
18.4.6	Illumina Sequencing	307
18.5	Applications	308
18.6	Perspectives	308
	References	309
19	Second-Generation Sequencing Library Preparation: <i>In Vitro</i> Tagmentation via Transposome Insertion	311
	<i>Fraz Syed</i>	
19.1	Introduction	311
19.2	Methods and Protocols	313
19.2.1	Materials	313
19.2.2	Methods	314
19.3	Perspectives	321
	References	321

Part Two	Next-Generation Tag-Based Sequencing	323
20	Moving Towards Third-Generation Sequencing Technologies	325
	<i>Karolina Janitz and Michal Janitz</i>	
20.1	Introduction	325
20.2	Differences Between NGS and Sanger Sequencing	326
20.3	Preparation of Templates for Sequencing	326
20.3.1	Next Generation: Single-Molecule Templates	327
20.4	Real-Time Sequencing	327
20.5	Nanopore Sequencing	328
20.6	Ion Torrent Electronic Sequencing	329
20.7	Genome Enrichment	331
20.8	Advantages of NGS	331
20.9	Problem of Short Reads	333
20.10	Perspectives	335
	References	335
21	Beyond Tags to Full-Length Transcripts	337
	<i>Mohammed Mohiuddin, Stephen Hutchison, and Thomas Jarvie</i>	
21.1	Introduction	337
21.2	Generation of Full-Length Transcriptomes	338
21.2.1	mRNA Fragmentation and Preparation for Sequencing	339
21.2.1.1	Step 1: mRNA Preparation	339
21.2.1.2	Step 2: Fragmentation	339
21.2.1.3	Step 3: First-Strand Synthesis	340
21.2.1.4	Step 4: Second-Strand Synthesis and Library Preparation	340
21.2.2	Sequencing of Full-Length Transcriptomes	341
21.3	Methods	342
21.3.1	Assembly	342
21.3.2	Mapping	343
21.4	Applications	344
21.4.1	Model Organisms	344
21.4.2	Fusion Transcript Detection	347
21.4.3	Digital Gene Expression	349
21.4.4	Allele-Specific Expression	349
21.5	Perspectives	350
	References	351
22	Helicos Single-Molecule Sequencing for Accurate Tag-Based RNA Quantitation	353
	<i>John F. Thompson, Tal Raz, and Patrice M. Milos</i>	
22.1	Introduction	353
22.2	Methods and Protocols	355
22.2.1	Reagents, Required Primers, and Thermocycler Programs	355
22.2.2	Standard Precautions	357
22.2.3	DGE Sample Preparation	357
22.3	Applications	362
22.4	Perspectives	364
	References	365
23	Total RNA-seq: Complete Analysis of the Transcriptome Using Illumina Sequencing-by-Synthesis Sequencing	367
	<i>Shujun Luo, Geoffrey P. Smith, Irina Khrebtukova, and Gary P. Schroth</i>	
23.1	Introduction	367
23.2	Total RNA-Seq	368
23.3	Methods and Protocols	369
23.3.1	Key Materials	369

- 23.3.2 Method 370
- 23.4 Total RNA-Seq Data Collection and Interpretation 378
- 23.5 Applications 380
 - References 381

Part Three Bioinformatics for Tag-Based Technologies 383

- 24 Computational Infrastructure and Basic Data Analysis for Next-Generation Sequencing 385**
 - David Sexton*
 - 24.1 Introduction 385
 - 24.2 Background 386
 - 24.3 Getting Started with the Next-Generation Manufacturers 387
 - 24.4 Infrastructure and Data Analysis 388
 - 24.4.1 Computational Considerations 388
 - 24.4.2 Data Dynamics 389
 - 24.4.3 Software and Postanalysis 390
 - 24.4.4 Staffing Requirements 391
 - 24.5 Applications 392
 - 24.6 Perspectives 392
- 25 CLC Bio Integrated Platform for Handling and Analysis of Tag Sequencing Data 393**
 - Roald Forsberg, Søren Mønsted, and Anne-Mette Hein*
 - 25.1 Introduction 393
 - 25.2 Main Components and Features 394
 - 25.2.1 Data Flow and Data Back-End 394
 - 25.2.2 CLC Genomics Workbench 394
 - 25.2.3 CLC Genomics Server 395
 - 25.2.4 APIs 396
 - 25.2.5 Acceleration of Analysis 396
 - 25.3 Applications 396
 - 25.3.1 First-Level Analysis 396
 - 25.3.1.1 Import 396
 - 25.3.1.2 Demultiplexing 397
 - 25.3.1.3 Trim and Quality Control 397
 - 25.3.2 Second and Third Levels – Application-Specific Steps 398
 - 25.3.2.1 RNA-Seq 398
 - 25.3.2.2 SmallRNA-Seq 399
 - 25.3.2.3 Tag-Seq 400
 - 25.3.2.4 ChIP-Seq 401
 - 25.3.3 Fourth Level – Expression Analysis 403
 - 25.4 Perspectives 404
 - References 405
- 26 Multidimensional Context of Sequence Tags: Biological Data Integration 407**
 - Korbinian Grote and Thomas Werner*
 - 26.1 Introduction 407
 - 26.2 Methods and Strategies 408
 - 26.2.1 Annotation Links Sequence Tags (Reads) to Biology 408
 - 26.2.2 Application of the Methods and Strategies 410
 - 26.2.3 Only Positive Results are Conclusive 411
 - 26.2.4 Automatic Workflow: ChIP-seq of Peroxisome Proliferator-Activated Receptor- γ 412
 - 26.3 Perspectives 414
 - References 415

- 27 Experimental Design and Quality Control of Next-Generation Sequencing Experiments 417**
Peter A.C. 't Hoen, Matthew S. Hestand, Judith M. Boer, Yuching Lai, Maarten van Iterson, Michiel van Galen, Henk P. Buermans, and Johan T. den Dunnen
- 27.1 Introduction 417
 - 27.2 Choice of Platform 417
 - 27.2.1 Read Length and Number of Reads 418
 - 27.2.2 Single-End versus Paired-End Sequencing 419
 - 27.2.3 Platform-Specific Advantages and Disadvantages 419
 - 27.3 Sequencing Depth 420
 - 27.3.1 Expression Profiling 420
 - 27.3.2 ChIP-Seq: Relation Enrichment Factor and Sequencing Depth 421
 - 27.3.3 Barcoding 422
 - 27.4 Replicates, Randomization, and Statistical Testing 422
 - 27.4.1 Technical and Biological Replicates 422
 - 27.4.2 Technical Variability 423
 - 27.4.3 Biological Replicates Increase Accuracy 424
 - 27.4.4 Sample Size 424
 - 27.4.5 Importance of Randomizing Samples 424
 - 27.5 Experimental Controls 425
 - 27.5.1 Spike-Ins 425
 - 27.5.2 Negative Controls in ChIP-Seq Experiments 426
 - 27.6 General Quality Assessment 427
 - 27.6.1 Nucleotide Frequency Characteristics 428
 - 27.6.2 Percentage Duplicate Reads 428
 - 27.7 Platform-Specific Quality Scores 428
 - 27.7.1 Sanger, Roche, Illumina, and SOLiD Quality Scores 429
 - 27.7.2 Conversion and Visualization of Quality Scores 429
 - 27.8 Quality Checks After Alignment 430
 - 27.8.1 Percentage of Reads Aligned and Percentage in Repeat Regions 430
 - 27.8.2 DeepSAGE: Percentage 21–22Mers 430
 - 27.8.3 RNA-Seq: Percentage Tags in Annotated Transcripts 430
 - 27.8.4 miRNA Profiling: Percentage in Annotated miRNAs 430
 - 27.8.5 ChIP-Seq: Enrichment 430
 - 27.8.6 Correlation Measures 431
 - 27.9 What Can Go Wrong 431
 - 27.9.1 Sample Swaps 431
 - 27.9.2 Contamination 431
 - 27.10 Perspectives 432
 - References 432
- 28 UTGB Toolkit for Personalized Genome Browsers 435**
Taro L. Saito, Jun Yoshimura, Budrul Ahsan, Atsushi Sasaki, Reginaldo Kurosh, and Shinichi Morishita
- 28.1 Introduction 435
 - 28.2 Overview of the UTGB Toolkit 436
 - 28.2.1 Availability of the UTGB Toolkit 438
 - 28.3 Methods 438
 - 28.3.1 Installation of the UTGB Toolkit 438
 - 28.3.1.1 Prerequisites 438
 - 28.3.1.2 Easy Installer 438
 - 28.3.1.3 Mac OS X and Linux 438
 - 28.3.1.4 Windows 439
 - 28.3.2 Running the UTGB Toolkit 439
 - 28.3.3 Viewing Help Messages 439
 - 28.3.4 Creating a new UTGB Project 440

28.3.5	Building a Genome Browser	441
28.3.6	Launching a Portable Web Server	441
28.3.7	Configuring Track Views	441
28.3.8	Adding a New Track	442
28.3.8.1	FastaTrack	442
28.3.8.2	ReadTrack	442
28.3.8.3	WigTrack	443
28.3.8.4	Adding Keyword Search	443
28.3.9	Switching Views	443
28.3.10	Publishing Your Genome Browser	443
28.3.11	Manual Installation of UTGB Toolkit (Optional)	444
28.3.11.1	Windows	444
28.3.11.2	Mac OS X and Linux	444
28.3.12	Developing Your Own Tracks	444
28.4	Applications	444
28.4.1	Portable Web Server for Quickly Browsing Local Resources	444
28.4.2	Portable Database Engine	445
28.4.3	Web Application Development Framework	446
28.4.3.1	Server-Side Programming Support	446
28.4.3.2	Web Action	446
28.4.3.3	Database Connection	446
28.4.3.4	Object-Database Mapping	446
28.5	Perspectives	447
	References	447
29	Beyond the Pipelines: Cloud Computing Facilitates Management, Distribution, Security, and Analysis of High-Speed Sequencer Data	449
	<i>Boris Umylny and Richard S.J. Weisburd</i>	
29.1	Introduction	449
29.2	Data Management	450
29.2.1	Data Quantity	450
29.2.2	HSSs	451
29.2.3	Data Analysis	452
29.2.4	Data Size	453
29.3	Distribution	454
29.3.1	Collaboration	454
29.3.2	Distribution of Data, Annotations, and Analysis Tools	455
29.4	Analysis	456
29.4.1	Integrating Data Repositories and Analytics	456
29.4.2	Integrating HSS Discovery Pipelines with Annotation Data	458
29.4.3	Integrating HSS and Traditional Analysis Algorithms	459
29.4.4	Cloud-Based Infrastructure	461
29.5	Security	462
29.6	Healthcare Data and Privacy Issues	464
29.7	Sample Evaluation of a Vendor Solution	465
29.8	Perspectives	465
	References	467
30	Computational Methods for the Identification of MicroRNAs from Small RNA Sequencing Data	469
	<i>Eugene Berezikov</i>	
30.1	Introduction	469
30.2	Implementing the miR-Intess Pipeline	470
30.2.1	Preprocessing of Small RNA Sequencing Data	470
30.2.2	Mapping of Small RNA Reads to the Genome	471
30.2.3	Annotation	471
30.2.4	Identification of Hairpin Structures	472

- 30.2.5 miRNA Signatures 472
- 30.2.6 miRNA Expression Profiles 474
- 30.3 Applications 474
- References 474

Glossary 477

Link Collection for Next-Generation Sequencing 565

Index 575

