Contents

Contributors ix
About the editor xi
Acknowledgments xiii

1. Past, current, and future of transcriptomics

Xinmin Li, Ilya Belalov, and Anton Buzdin

1.1 Introduction 1
1.2 Historical development 2
1.3 Methodological advancements 3
1.4 Microarrays 3
1.5 RNA sequencing (RNA-seq) 5
1.6 Single-cell RNA sequencing (scRNA-seq) 6
1.7 Spatial transcriptomics 6
1.8 Noncoding RNAs in gene regulation 6
1.9 MicroRNAs (miRNAs) 7
1.10 Long noncoding RNAs (lncRNAs) 7
1.11 Regulatory networks 8
1.12 Single-cell and spatial transcriptomics in understanding cellular heterogeneity 8
1.13 Single-cell transcriptomics 8
1.14 Spatial transcriptomics 9
1.15 Insights and applications 9
1.16 Challenges and future directions 9
1.17 Emerging trends and future directions in transcriptomics 10
1.18 RNA sequencing (RNA-seq) and beyond 10
1.19 Dual RNA-seq for host–pathogen interactions 10
1.20 Transcriptomics in food microbiology and agriculture 10
1.21 Multi-omics integration 11
1.22 Single-cell transcriptomics 11
1.23 Challenges and future prospects 11
1.24 Current challenges and controversies in transcriptomics 11
1.25 Technical limitations and data complexity 12
1.26 Interpretation of dual RNA-seq data 12
1.27 Environmental influences on transcriptomic profiles 12
1.28 Transcriptomics in food microbiology and safety 12
1.29 Toxicogenomics and risk assessment 13
1.30 Functional annotation and novel transcripts 13
1.31 Conclusion 13
References 13

2. Pitfalls of transcriptomics and selection of the most appropriate transcriptomic technique

Xinmin Li, Ilya Belalov, and Anton Buzdin

2.1 Introduction 15
2.2 Pitfalls in specific transcriptomic techniques 17
2.3 Choosing the right tool for the job 22
2.4 Case studies and examples 24
2.5 Addressing pitfalls and challenges: Insights from recent research 25
2.6 Future directions in transcriptomics 27
2.7 Practical considerations and conclusion 30
References 32
Further reading 34

3. Bulk RNA sequencing in wet lab

Anton Buzdin, Ilya Belalov, Maria Suntsova, and Xinmin Li

3.1 Introduction to bulk RNA sequencing 35
3.2 Fundamental principles of bulk RNA sequencing 38
3.3 Key technologies and methodologies in bulk RNA sequencing 39
3.4 Sample collection and storage 40
3.5 RNA extraction and purification 41
3.6 Quality control measures 42
3.7 Library preparation for bulk RNA sequencing 43

3.8 Early barcoding and cost efficiency 44
3.9 Direct RNA isolation strategies 45
3.10 Sequencing platforms and technologies 46
3.11 RNA sequencing protocols 49
3.12 Overview of advanced RNA sequencing protocols 50
3.13 Reagents and equipment for bulk RNA sequencing 52
3.14 Technical and analytical quality control in RNA sequencing 53
3.15 Bioinformatics and data analysis 55
3.16 Data preprocessing and normalization methods for bulk RNA sequencing 56
3.17 Sequencing depth and coverage in bulk RNA sequencing 57
3.18 Differential expression analysis in RNA sequencing 59
3.19 Functional genomics and pathway analysis in RNA sequencing 60
3.20 Emerging applications of bulk RNA sequencing in disease research, immunology, and oncology 62
3.21 Challenges and limitations of current RNA sequencing methodologies 63
3.22 Future directions and technologies in RNA sequencing 64
3.23 Best practices for bulk RNA sequencing in the wet lab 65
3.24 Ethical considerations in transcriptomics research 67
3.25 Conclusion 69
References 69

4. Single-cell RNA sequencing in wet lab

Anton Buzdin, Xinmin Li, Maria Suntsova, and Ilya Belalov

4.1 Overview of what scRNA-seq is and why it is indispensable in research 73
4.2 Types of scRNA-seq protocols 74
4.3 Reagents and materials 76
4.4 scRNA-seq platforms: Comparison of chromium 10X, Drop-seq, and Smart-seq 77
4.5 Sample preparation and handling 78
4.6 Single-cell isolation techniques 79
4.7 Library preparation 80
4.8 Sequencing technologies 81
4.9 Techniques to ensure the integrity and quality of scRNA-seq data 82
4.10 Data analysis methods to validate the quality of sequencing results 83
4.11 Recommended sequencing depths for different types of analyses 84
4.12 Overview of computational tools and methods for analyzing scRNA-seq data 85
4.13 Integrating scRNA-seq data with other data types 86
4.14 Research and clinical applications of scRNA-seq 87
4.15 Innovations and future trends in scRNA-seq technology 89
References 90
Further reading 92

5. Spatial transcriptomics

Ilya Belalov, Ye Wang, Anton Buzdin, and Xinmin Li

5.1 Introduction 93
5.2 Spatial transcriptomics technologies 93
5.3 Imaging-based technologies 94
5.4 Sequencing-based technologies 95
5.5 Applications for basic, translational, and clinical research 101
5.6 Considerations for a first ST experiment 106
5.7 Concluding remarks 107
References 107

6. High-throughput and multiplex PCR in wet laboratory to measure gene expression

Anton Buzdin, Ilya Belalov, and Galina Zakharova

6.1 Introduction 113
6.2 Importance of high-throughput and multiplex PCR 114
6.3 Protocols for high-throughput and multiplex PCR 114
6.4 Reagents and requirements for material 116
6.5 Material requirements 117
6.6 Technical and analytical quality control (QC) 119
6.7 Analytical QC 121
6.8 Dynamic intervals and sensitivity 123
6.9 Sensitivity and specificity 125
6.10 Applications and sequencing depth recommendations 127
6.11 Recommended sequencing depths 129
6.12 Key studies and contributions 130

6.13 Further reading 132
6.14 Future directions and innovations 134
6.15 Challenges and limitations 135
6.16 Conclusion 137
References 137

7. Processing primary gene expression data: Normalization, harmonization and data quality control
Nicolas Borisov, Maksim Sorokin, and Anton Buzdin

7.1 Background 139
7.2 Principles of harmonization algorithms 143
7.3 Evaluation of the quality of harmonization 145
7.4 Validation of Shambhala-1/2 protocol on bulk mRNA-seq profiles 150
7.5 Differential clustering of human normal and cancer expression profiles 152
7.6 Correlation, regression, and sign-change analysis of cancer drug balanced efficiency scores after application of different methods of harmonization 156
7.7 Retention of biological properties after uniformly shaped harmonization 161
7.8 Validation of Shambhala-1/2 protocol on sc-mRNA-seq profiles 163
7.9 Discussion 168
Abbreviations 171
References 171

8. Data check and differential gene analysis
Maksim Sorokin and Anton Buzdin

8.1 Primary RNA expression data analysis 181
8.2 Finding differentially expressed genes and gene sets 185
8.3 Concluding remarks 189
References 190

9. Molecular pathway analysis using transcriptomic data
Nicolas Borisov, Maksim Sorokin, Igor Kovalchuk, Marianna Zolotovskaia, and Anton Buzdin

9.1 Background 193
9.2 Topology-based methods for pathway activation assessment 195
9.3 Methods for database preparation for pathway activation assessment 200
9.4 Personalized ranking of cancer drugs based on pathway activation levels 205
9.5 Concluding remarks 207
Abbreviations 208
References 208

10. Omics wise analysis of RNA splicing
Alexander Modestov, Anton Buzdin, and Vladimir Prassolov

10.1 Introduction 213
10.2 Omics technologies for RNA splicing analysis 216
10.3 RNA splicing products and regulators as biomarkers and targets of therapy 220
10.4 Challenges and future prospects 222
References 223

11. Transcriptome-wide analysis of protein synthesis: Ribosome profiling and beyond
Sergey E. Dmitriev, Daniil Luppov, Leonid M. Kats, Aleksandra S. Anisimova, and Ilya M. Terenin

11.1 Introduction 231
11.2 Ribosome profiling quantifies gene expression at the translational level 233
11.3 Ribosome profiling deciphers cryptic coding potential of the genome 239
11.4 Analyses of ribosome positional distribution and transcriptome-wide metagene profile reveal new phenomena 248
11.5 Molecular mechanisms revealed by variants of ribosome profiling and relative methods 251
11.6 Technical challenges, limitations, and artifacts in ribosome profiling studies 263
11.7 Ribosome profiling of cultured mammalian cells: An example protocol 271
Acknowledgments 279
References 279

12. Transcriptomic biomarkers in biomedicine
Anton Buzdin, Alf Giese, Xinmin Li, and Ye Wang

12.1 Theoretical background 299
12.2 Molecular pathway biomarkers in oncology 303

12.3 Other applications of pathway biomarker analysis in biomedicine 314
12.4 Conclusion 316
References 316

13. Translational transcriptomics for personalized oncology

Anton Buzdin, Alexander Sorokin, Marianna Zolotovskaia, Maksim Sorokin, Victor Tkachev, Alf Giese, Marina Sekacheva, Elena Poddubskaya, Julian Rozenberg, and Thanaa Mohammad

13.1 Transcriptomics in clinical oncology 331
13.2 Detection of oncogenic fusion events 332
13.3 Detection of pathogenic splicing or exon skipping events 359
13.4 Gene signatures 360
13.5 Algorithmic approach in personalized oncology 361
13.6 Molecular pathway approach for personalized prediction of cancer drug efficacy 368
13.7 Concluding remarks 379
References 380

14. Transcriptomics for modern biotechnology

Anton Buzdin, Denis Kuzmin, Andrew Garazha, and Ilya Belaley

14.1 Introduction 397
14.2 High-throughput transcriptomic pipelines 400
14.3 Oncobox pathway activation levels quantization 403
14.4 Connectivity map assay 405
14.5 Databases and software for transcriptomic analysis 408
14.6 Industrial and startup landscape 411
14.7 Case studies and recent advances 417
14.8 Conclusion 421
References 424

15. Transcriptomics and quantitative proteomics: Competition or symbiosis?

Anton Buzdin, Sergey Moshkovskii, Xinmin Li, and Ilya Belaley

15.1 Introduction 429
15.2 Rationale for direct protein level measurement 430
15.3 Current progress in quantitative proteomics 431
15.4 Quality metrics comparison 434
15.5 Gene products interrogated 437
15.6 Integration of multi-omics data 439
15.7 Future perspectives 443
15.8 Conclusion 445
References 446

16. Quality assessment of differentially expressed gene signatures

Alexey Stupnikov and Anton Buzdin

16.1 Introduction 449
16.2 Conclusions 459
Acknowledgments 459
References 459

17. Detection of fusion transcripts by RNA-sequencing data

Ivan Musatov, Maksim Sorokin, and Anton Buzdin

17.1 Introduction 463
17.2 Basic software methods for detecting fusion genes 464
17.3 Algorithms based on alignment. STAR and STAR-Fusion 467
17.4 LongGF 476
17.5 Pros and cons of existing tools for bioinformatic fusion detection 479
17.6 Prospects for the development of tools and approaches to search for fusion transcripts 484
17.7 Conclusion 484
References 484

Index 489