

# BÖLÜM 17

## RNA Dizileme (Transkriptom)

Ayşe Gül KASAPOĞLU<sup>1</sup>

Emre İLHAN<sup>2</sup>

Murat TURAN<sup>3</sup>

### Giriş

#### Transkriptom

Moleküler biyolojinin merkezi dogması, DNA'da depolanan genetik bilginin akışı, DNA'nın RNA'ya dönüştürülmesi ve nihayetinde proteinler olarak yazılımlanmasıdır (1,2). Çevresel faktörler tarafından düzenlenen bu genetik bilginin tam ifadesi bir organizmanın fenotipini karakterize eder. Genlerin bir setinin tamamlayıcı RNA moleküllerine transkripsiyonu, bir hücrenin kimliğini belirler ve hücre içerisinde biyolojik aktiviteleri düzenler. Transkriptom olarak tanımlanan tüm bu RNA molekülleri genomun fonksiyonel elemanlarını, gelişimi ve hastalıkları anlamak için gereklidir.

Transkriptom, yüksek derecede karmaşıklığa sahip olup kodlanan ve kodlanmayan RNA türlerinin tümünü kapsar. Tarihsel olarak RNA molekülleri moleküler biyolojinin merkezi dogmasında gizlenmiş, genler ve proteinler arasındaki basit bir ara molekül olmuştur. Bu nedenle haberci (Messenger) RNA (mRNA) molekülleri genetik kodla proteinlere kodlandıkları için en yoğun çalışılan RNA türleridir. Protein kodlayan RNA'lara ilaveten fonksiyonel

<sup>1</sup> Arş. Gör., Erzurum Teknik Üniversitesi, Fen Fakültesi, Moleküler Biyoloji ve Genetik Bölümü, aysegul.kasapoglu@erzurum.edu.tr, ORCID iD: 0000-0002-6447-4921

<sup>2</sup> Doç. Dr., Erzurum Teknik Üniversitesi, Fen Fakültesi, Moleküler Biyoloji ve Genetik Bölümü, emre.ilhan@erzurum.edu.tr, ORCID iD: 0000-0002-8404-7900

<sup>3</sup> Dr. Öğr. Üyesi, Erzurum Teknik Üniversitesi, Fen Fakültesi, Moleküler Biyoloji ve Genetik Bölümü, m.turan@erzurum.edu.tr, ORCID iD: 0000-0003-2900-1755

yanıt mekanizmalarının belirlenmesi gibi RNA biyolojisinin birçok yönünü analiz etmek mümkündür. RNA'nın genom içerisindeki rolü hala araştırılmakla birlikte birçok alanda RNA-seq deneyleri halihazırda kullanılmaktadır. Bütün bu deneylerde kullanıcılar Illumina kısa okuma RNA dizileme yöntemini tercih etse de önümüzdeki zamanlarda uzun okumalı dizileme yöntemleri daha fazla tercih edilecek gibi görünmektedir. Bunun için uzun okumalı dizileme teknolojilerinde iyileştirmeler yapılması gerekmektedir. Uzun okumalı dizileme teknolojileri mRNA'nın paralog dizilemesinde oldukça avantajlı konumdayken maliyetlerinin azalması ve güvenilir hale gelmesi zaman alacaktır. Bütün bunları göz önünde bulundurarak RNA-seq'in gelecek yıllarda nasıl gelişebileceğine dair tahminler tam olarak yapılamamaktadır.

## Kaynaklar

1. Crick F. Central dogma of Molecular Biology. *Nature*. 1970; 227: 561-563.
2. Crick FH. On protein synthesis. *Symposia of the Society for Experimental Biology*. 1958; 12: 138-163.
3. Mattick JS, Makunin IV. Non-coding RNA. *Human Molecular Genetics*. 2006; 15 (1): R17–R29. doi: 10.1093/hmg/ddl046
4. Stefani G, Slack FJ. Small non-coding RNAs in animal development. *Nature Reviews Molecular Cell Biology*. 2008; 9: 219-230.
5. Okazaki Y, Furuno M, Kasukawa T, et al. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature*. 2002; 420: 563-573.
6. Guttman M, Amit I, Garber M, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*. 2009; 458: 223-227.
7. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: Insights into functions. *Nature Reviews Genetics*. 2009; 10: 155-159.
8. Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: Functional surprises from the RNA world. *Genes & Development*. 2009; 23: 1494-1504.
9. Schena M, Shalon D, Davis RW, et al. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*. 1995; 270: 467-470.
10. Casneuf T, Van de Peer Y, Huber W. In situ analysis of cross-hybridisation on microarrays and the inference of expression correlation. *BMC Bioinformatics*. 2007; 8: 461.
11. Shendure J. The beginning of the end for microarrays? *Nature Methods*. 2008; 5: 585-587. doi: 10.1038/nmeth0708-585
12. Kukurba KR, Montgomery SB. RNA sequencing and analysis. *Cold Spring Harbor Protocols*. 2015; 2015(11): 951-969. doi: 10.1101/pdb.top084970
13. Shiraki T, Kondo S, Katayama S, et al. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proceedings of the National Academy of Sciences*. 2003; 100: 15776-15781.
14. Wang Z, Gerstein M, Snyder M. RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*. 2009; 10: 57-63.
15. Emrich SJ, Barbazuk WB, Li L, et al. Gene discovery and annotation using LCM-454 transcriptome sequencing. *Genome Research*. 2007; 17: 69-73.

16. Lister R, O'Malley RC, Tonti-Filippini J, et al. Highly integrated single- base resolution maps of the epigenome in *Arabidopsis*. *Cell*. 2008; 133: 523-536.
17. Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. *Nature Reviews Genetics*. 2019; 20(11): 631-656.
18. Illumina. NGS Library Preparation- 3 Key Technologies. (23.12.2022 tarihinde <https://www.illumina.com/techniques/sequencing/ngs-library-prep.html> adresinden ulaşılmıştır).
19. Garalde DR, Snell, EA, Jachimowicz D, et al. Highly parallel direct RNA sequencing on an array of nanopores. *Nature Methods*. 2018; 15: 201-206.
20. Smith AM, Jain M, Mulroney L, et al. Reading canonical and modified nucleotides in 16S ribosomal RNA using nanopore direct RNA sequencing. *bioRxiv*. 2017; 132274. doi: 10.1101/132274
21. Byrne A, Beaudin AE, Olsen HE, et al. Nanopore long- read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. *Nature Communications*. 2017; 8: 16027.
22. Leinonen R, Sugawara H, Shumway M, et al. The sequence read archive. *Nucleic Acids Research*. 2011; 39: D19–D21.
23. Kono N, Arakawa K. Nanopore sequencing: Review of potential applications in functional genomics. *Development, growth & Differentiation*. 2019; 61(5): 316-326.
24. Hong M, Tao S, Zhang L, et al. (2020). RNA sequencing: new technologies and applications in cancer research. *Journal of Hematology & Oncology*. 2020; 13(1): 1-16.
25. Piovesan A, Caracausi M, Antonaros F, et al. GeneBase 1.1: a tool to summarize data from NCBI Gene datasets and its application to an update of human gene statistics. *Database*. 2016; 2016: baw153.
26. Frankish A, Diekhans M, Ferreira AM, et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Research*. 2019; 47: D766-D773.
27. Ramsköld D, Luo S, Wang YC, et al. Full- length mRNA- Seq from single- cell levels of RNA and individual circulating tumor cells. *Nature Biotechnology*. 2012; 30: 777-782.
28. Bolisetty MT, Rajadinakaran G, Graveley BR. Determining exon connectivity in complex mRNAs by nanopore sequencing. *Genome Biology*. 2015; 16: 204.
29. Ardui S, Ameer A, Vermeesch JR, et al. Single molecule real- time (SMRT) sequencing comes of age: applications and utilities for medical. *Nucleic Acids Research*. 2018; 46: 2159-2168.
30. Jain M, Olsen HE, Paten B, et al. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology*. 2016; 17: 239.
31. Workman RE, Tang AD, Tang PS, et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nature Methods*. 2019; 16: 1297-1305. doi: 10.1038/s41592-019-0617-2
32. Weirather JL, de Cesare M, Wang Y, et al. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Res*. 2017; 6: 100.
33. Tomita H, Vawter MP, Walsh DM, et al. Effect of agonal and postmortem factors on gene expression profile: Quality control in microarray analyses of postmortem human brain. *Biological Psychiatry*. 2004; 55: 346-352.
34. Thompson KL, Pine PS, Rosenzweig BA, et al. Characterization of the effect of sample quality on high density oligonucleotide microarray data using progressively degraded rat liver RNA. *BMC Biotechnology*. 2007; 7: 57.
35. Rudloff U, Bhanot U, Gerald W, et al. Biobanking of human pancreas cancer tissue: Impact of ex-vivo procurement times on RNA quality. *Annals of Surgical Oncology*. 2010; 17: 2229-2236.

36. Griffith RD, Simmons BJ, Bray FN, et al. 1064 nm Q-switched Nd: YAG laser for the treatment of Argyria: a systematic review. *Journal of the European Academy of Dermatology and Venereology*. 2015; 29(11): 2100-2103.
37. Conesa A, Madrigal P, Tarazona S, et al. A survey of best practices for RNA-seq data analysis. *Genome Biology*. 2016; 17(1): 1-19.
38. Morin RD, Zhao YJ, Prabhu AL, et al. Preparation and analysis of Micro-RNA libraries using the Illumina massively parallel sequencing technology. *Methods in Molecular Biology*. 2010; 650: 173-199.
39. Parkhomchuk D, Borodina T, Amstislavskiy V, et al. Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Research*. 2009; 37: e123.
40. Vivancos AP, Guell M, Dohm JC, et al. Strand-specific deep sequencing of the transcriptome. *Genome Research*. 2010; 20: 989-999.
41. Mills JD, Kawahara Y, Janitz M. Strand-specific RNA-Seq provides greater resolution of transcriptome profiling. *Current Genomics*. 2013; 14: 173-181.
42. Borodina T, Adjaye J, Sultan MA. Strand-specific library preparation protocol for RNA sequencing. *Methods in Enzymology*. 2011; 500: 79-98. doi: 10.1016/B978-0-12-385118-5.00005-0
43. Jayaprakash AD, Jabado O, Brown BD, et al. Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Research*. 2011; 39: e141. doi: 10.1093/nar/gkr693
44. Sun G, Wu X, Wang J, et al. A bias-reducing strategy in profiling small RNAs using Solexa. *RNA*. 2011; 17: 2256-2262. doi: 10.1261/rna.028621.111
45. Levin JZ, Yassour M, Adiconis X, et al. Comprehensive comparative analysis of strand specific RNA sequencing methods. *Nature Methods*. 2010; 7: 709-715. doi:10.1038/nmeth.1491.
46. Hrdlickova R, Toloue M, Tian B. RNA-Seq methods for transcriptome analysis. *Wiley Interdisciplinary Reviews: RNA*. 2017; 8(1): e1364.
47. Emmert-Buck MR, Bonner RF, Smith PD, et al. Laser capture microdissection. *Science*. 1996; 274: 998-1001.
48. Kube DM, Savci-Heijink CD, Lamblin AF, et al. Optimization of laser capture microdissection and RNA amplification for gene expression profiling of prostate cancer. *BMC Molecular Biology*. 2007; 8: 25.
49. Cantor H, Simpson E, Sato VL, et al. Characterization of subpopulations of T lymphocytes. I. Separation and functional studies of peripheral T-cells binding different amounts of fluorescent anti-Thy 1.2 (theta) antibody using a fluorescence-activated cell sorter (FACS). *Cellular Immunology*. 1975; 15: 180-196.
50. Huang S. Non-genetic heterogeneity of cells in development: More than just noise. *Development*. 2009; 136: 3853-3862.
51. Kawasaki ES. Microarrays and the gene expression profile of a single cell. *Annals of the New York Academy of Sciences*. 2004; 1020: 92-100.
52. Islam S, Kjallquist U, Moliner A, et al. Highly multiplexed and strand-specific single-cell RNA 5' end sequencing. *Nature Protocols*. 2012; 7: 813-828.
53. Hashimshony T, Wagner F, Sher N, et al. CEL-Seq: Single-cell RNA-Seq by multiplexed linear amplification. *Cell Reports*. 2012; 2: 666-673.
54. Chapman AR, He Z, Lu S, et al. Single cell transcriptome amplification with MALBAC. *PLoS One*. 2015; 10: e0120889. doi: 10.1371/journal.pone.0120889
55. Kivioja T, Vaharautio A, Karlsson K, et al. Counting absolute numbers of molecules using unique molecular identifiers. *Nature Methods*. 2012; 9: 72-74. doi: 10.1038/nmeth.1778
56. Islam S, Zeisel A, Joost S, et al. Quantitative single-cell RNA-seq with unique molecular identifiers. *Nature Methods*. 2014; 11: 163-166. doi: 10.1038/nmeth.2772

57. Grun D, Kester L, van Oudenaarden A. Validation of noise models for single-cell transcriptomics. *Nature Methods*. 2014; 11: 637–640. doi: 10.1038/nmeth.2930
58. Fan HC, Fu GK, Fodor SP. Expression profiling. Combinatorial labeling of single cells for gene expression cytometry. *Science*. 2015; 347: 1258367. doi: 10.1126/science.1258367
59. Avital G, Hashimshony T, Yanai I. Seeing is believing: new methods for in situ single-cell transcriptomics. *Genome Biology*. 2014;15: 110. doi: 10.1186/gb4169
60. Lovatt D, Ruble BK, Lee J, et al. Transcriptome in vivo analysis (TIVA) of spatially defined single cells in live tissue. *Nature Methods*. 2014; 11: 190-196. doi: 10.1038/nmeth.2804
61. Lee JH, Daugharthy ER, Scheiman J, et al. Highly multiplexed subcellular RNA sequencing in situ. *Science*. 2014; 343: 1360-1363. doi: 10.1126/science.1250212
62. Sahraeian SME, Mohiyuddin M, Sebra R, et al. Gaining comprehensive biological insight into the transcriptome by performing a broad- spectrum RNA- seq analysis. *Nature Communications*. 2017; 8: 59.
63. Seyednasrollah F, Laiho A, Elo LL. Comparison of software packages for detecting differential expression in RNA- seq studies. *Briefings in Bioinformatics*. 2013; 16: 59–70.
64. Williams CR, Baccarella A, Parrish JZ, et al. Empirical assessment of analysis workflows for differential expression analysis of human samples using RNA- seq. *BMC Bioinformatics*. 2017; 18: 38.
65. Kim D, Pertea G, Trapnell C, et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*. 2013; 14: R36.
66. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA- seq aligner. *Bioinformatics*. 2013; 29: 15–21.
67. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*. 2015; 12: 357-360.
68. Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*. 2010; 28: 511-515.
69. Bray NL, Pimentel H, Melsted P, et al. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*. 2016; 34: 4-8.
70. Patro R, Mount SM, Kingsford C. Sailfish enables alignment- free isoform quantification from RNA-seq reads using lightweight algorithms. *Nature Biotechnology*. 2014; 32: 462-464.
71. Patro R, Duggal G, Love MI, et al. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*. 2017; 14: 417-419.
72. Wu DC, Yao J, Ho, KS, et al. Limitations of alignment-free tools in total RNA-seq quantification. *BMC Genomics*. 2018; 19: 510.
73. Robert C, Watson M. Errors in RNA-Seq quantification affect genes of relevance to human disease. *Genome Biology*. 2015; 16: 177.
74. Yang C, Wu PY, Tong L, et al. The impact of RNA-seq aligners on gene expression estimation. *BCB '15: Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics*. 2016; 9: 462-471.
75. Fonseca NA, Marioni JC, Brazma A. RNA- Seq gene profiling a systematic empirical comparison. *Plos One* 2014; 9: e107026.
76. Teng M, Love, MI, Davis CA, et al. A benchmark for RNA- seq quantification pipelines. *Genome Biology*. 2016; 17: 74.
77. Quinn TP, Crowley TM, Richardson MF. Benchmarking differential expression analysis tools for RNA-Seq: normalization- based versus log-ratio transformation-based methods. *BMC Bioinformatics*. 2018; 19: 274.
78. Li X, Brock GN, Rouchka EC, et al. A comparison of per sample global scaling and per gene normalization methods for differential expression analysis of RNA-seq data. *Plos One*, 2017; 12: e0176185.

79. Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*. 2012; 7: 562-578.
80. Turro E, Su SH, Gonçalves A, et al. Haplotype and isoform specific expression estimation using multi-mapping RNA-seq reads. *Genome Biology*. 2011; 12: R13.
81. Anders S, Pyl PT, Huber W. HTSeq- a Python framework to work with high- throughput sequencing data. *Bioinformatics*. 2015; 31: 166-169.
82. Anders S, McCarthy DJ, Chen Y, et al. Count-based differential expression analysis of RNA sequencing data using R and bioconductor. *Nature Protocols*. 2013; 8: 1765-1786.
83. Oshlack A, Wakefield MJ. Transcript length bias in RNA-seq data confounds systems biology. *Biology Direct*. 2009; 4: 14.
84. Roberts A, Trapnell C, Donaghey J, et al. Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biology*. 2011; 12: R22.
85. Bourgon R, Gentleman R, Huber W. Independent filtering increases detection power for high-throughput experiments. *Proceedings of the National Academy of Sciences*. 2010; 107: 9456-9551.
86. Risso D, Schwartz K, Sherlock G, et al. GC-content normalization for RNA-seq data. *BMC Bioinformatics*. 2011; 12: 480.
87. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory in Biosciences*. 2012; 131: 281-285.
88. Mortazavi A, Williams BA, Mccue K, et al. Mapping and quantifying mammalian transcripts by RNA-Seq. *Nature Methods*. 2008; 5: 621-628.
89. Bullard JH, Purdom E, Hansen KD, et al. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics*. 2010; 11: 94-107.
90. Dillies MA, Rau A, Aubert J, et al. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in Bioinformatics*. 2013; 14: 671-683.
91. Robinson MD, Mccarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010; 26: 139-140.
92. Evans C, Hardin J, Stoebel DM. (2018). Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. *Briefings in Bioinformatics*. 2018; 19(5): 776-792.
93. Chen K, Hu Z, Xia Z, et al. The overlooked fact: fundamental need for spike- in control for virtually all genome- wide analyses. *Molecular and Cellular Biology*. 2016; 36: 662-667.
94. Hardwick SA, Deveson I, Mercer TR. Reference standards for next-generation sequencing. *Nature Reviews Genetics*. 2017; 18: 473-484.
95. Pine PS, Munro SA, Parsons JR, et al. Evaluation of the External RNA Controls Consortium (ERCC) reference material using a modified Latin square design. *BMC Biotechnology*. 2016; 16: 54.
96. Paul L, Kubala P, Horner G, et al. SIRVs: spike- in RNA variants as external isoform controls in RNA-sequencing. *bioRxiv*. 2016; 080747. doi: 10.1101/080747
97. Hardwick SA, Chen WY, Wong T, et al. Spliced synthetic genes as internal controls in RNA sequencing experiments. *Nature Methods*. 2016; 13: 792-798.
98. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA- seq data with DESeq2. *Genome Biology*. 2014; 15: 550.
99. Law CW, Chen Y, Shi W, et al. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*. 2014; 15: R29.

100. Frazee A, Pertea G, Jaffe AE, et al. Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nature Biotechnology*. 2015; 33: 243-246.
101. Bullard JH, Purdom E, Hansen KD, et al. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics*. 2010; 11: 94.
102. Tarazona S, Garcia-Alcalde F, Dopazo J, et al. Differential expression in RNA-seq: A matter of depth. *Genome Research*. 2011; 21: 2213-2223.
103. Glaus P, Honkela A, Rattray M. Identifying differentially expressed transcripts from RNA-seq data with biological variation. *Bioinformatics*. 2012; 28: 1721-1728.
104. Auer PL, Doerge RW. Statistical design and analysis of RNA sequencing data. *Genetics*. 2010; 185: 405-416.
105. Liu Y, Zhou J, White KP. RNA-seq differential expression studies: More sequence or more replication? *Bioinformatics*. 2014; 30: 301-304.