Molecular Biology Reports (2022) 49:2325–2333
https://doi.org/10.1007/s11033-021-06963-0

REVIEW

RNA sequencing and its applications in cancer and rare diseases

Selvi Ergin[1] · Nasim Kherad[1] · Meryem Alagoz[1]

Received: 11 September 2021 / Accepted: 16 November 2021 / Published online: 6 January 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract
With the invention of RNA sequencing over a decade ago, diagnosis and identification of the gene-related diseases entered a new phase that enabled more accurate analysis of the diseases that are difficult to approach and analyze. RNA sequencing has availed in-depth study of transcriptomes in different species and provided better understanding of rare diseases and taxonomical classifications of various eukaryotic organisms. Development of single-cell, short-read, long-read and direct RNA sequencing using both blood and biopsy specimens of the organism together with recent advancement in computational analysis programs has made the medical professional's ability in identifying the origin and cause of genetic disorders indispensable. Altogether, such advantages have evolved the treatment design since RNA sequencing can detect the resistant genes against the existing therapies and help medical professions to take a further step in improving methods of treatments towards higher effectiveness and less side effects. Therefore, it is of essence to all researchers and scientists to have deeper insight in all available methods of RNA sequencing while taking a step-in therapy design.

Keywords RNA sequencing · Gene-related diseases · Transcriptomes · Eukaryotic organisms · Short-read · Long-read

Introduction

The basis of molecular biology began with genes located in DNA transcribed to RNA for protein synthesis; the emergence of the double-helix structure of DNA in 1953 showed the essence of life as a result of gene interaction [1, 2]. The whole machinery defines the organism's characteristics and maintains the biological functions of the cells and the organism as one. Therefore, RNA analysis is essential in understanding the genomic processes and the diseases' origin. The RNAs, collectively known as transcriptomes, are complex genomic structures with coding and non-coding regions and are intermediaries between genes and proteins. Thus, detailed study on transcriptome is essential to understand the genomic function and to identify molecular compositions of cells. In addition, more comprehensive knowledge on transcriptome can help us understand the cause as well as development of diseases.

Therefore a thorough study of the transcriptome is necessary for understanding genomic function, identifying molecular compositions of cells, and understanding the cause and development of diseases [3]

Among RNA species, messenger RNA (mRNA) is the most valuable one for further study as it carries the genomic data from the organism's DNA [4]. However, analysis of protein-coding RNA requires a precise technique that can distinguish the coding-protein RNA from the non-coding RNAs (ncRNAs). The complexity of the genome arises from the following; Coding genes comprise almost 2% of the whole human genome, and a majority of the coding genes undergo transcription [5]. Additionally, a single genomic locus is likely to exhibit different isoforms resulting in different splicing patterns with possibly various transcriptional start sites [6]. Moreover, unpredictable monoallelic (maternal or paternal allele) expression of genes adds an extra layer of complexity in transcriptomic analysis [7]. In-vivo and in-vitro analysis of homogenous cells populations has shown heterogeneity of the cells due to intrinsic and extrinsic factors such as microenvironment [8]. However, research shows the cells in the same microenvironment can manifest different transcript levels due to factors such as the cell cycle [9].

Under the category of ncRNAs, ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs) as functional elements in mRNA translation, small nuclear RNAs (snRNAs) is RNA splicing, small nucleolar RNAs (snoRNAs)

✉ Meryem Alagoz
alameryem@hotmail.co.uk

[1] Department of Molecular Biology and Genetics, Biruni University, Istanbul, Turkey

in rRNAs modifications [10], microRNAs (miRNAs) and piwi-interacting RNAs (piRNAs) in post-transcriptional regulation of gene expression [11], and long non-coding RNAs (lncRNAs) in chromatin remodelling, transcriptional and post-transcriptional regulation [12]. Designing genome analysis techniques that can accurately and efficiently profile the whole genome and distinguish between the coding and non-coding ones was the scientists' target for decades. Over the past few decades, researchers developed various methods to have an in-depth analysis of RNAs and a more accurate understanding of gene expression. Low-throughput methods such as quantitative polymerase chain reaction (qPCR) which introduced as powerful techniques for the purpose. However, it could not apply to measuring multiple transcripts. And despite the introduction of hybridization-based microarray in 1995 that provided a better solution for the study of gene expression [13, 14], limitations of the method such as cross-hybridization with extremely similar sequences and lack of accuracy in the quantification of lowly- and highly expressed genes [15, 16] led scientists to develop sequence-based techniques to reduce the inaccuracy in the study of transcriptomes (transcriptomics) technologies using complementary DNA (cDNA). The aim of studying transcriptome is to catalogue the whole transcript (coding and non-coding RNAs), determine the splicing pattern and the changes that occur in the post-transcriptional stage, and identify the changes in expression level of each transcript by quantifying the changes based on different intrinsic and extrinsic factors [3]. Although techniques such as Sanger sequencing of cDNA using expressed sequence tag (EST) [17], serial analysis of gene expression (SAGE) and cap analysis of gene expression (CAGE) [18], have improved RNA analysis, their insensitivity in discovering novel genes and high cost of Sanger sequencing makes the techniques inefficient [19].

Next-generation sequencing (NGS) that are High-throughput sequencing can perform sequencing faster with lower cost and higher accuracy. Additionally, it is useful for identifying undefined gene expression sequences in an intense time manner [20]. Further development of long-read RNA sequencing, known as third-generation sequencing, can be used to generate full-length cDNA transcripts with a minimum number of false-positive splice sites and capturing great diversity of transcript isoforms [21]. The introduction of RNA-seq, from bulk- to single-cell RNA sequencing, has given the opportunity to process and map transcriptome.

Although the development of the RNA-seq method goes back more than a decade [22, 23], it has revolutionized the interpretation of eukaryotic transcriptomes [24, 25] by analysis of differential gene expression (DGE) using next-generation sequencing (NGS) with the standard workflow; RNA extraction, followed by mRNA enrichment or ribosomal RNA depletion, cDNA synthesis and preparation of an adaptor-ligated sequencing library. The advantage of the technique is the in-depth ability to perform 10–30 million reads in each sample on usually Illumina short-read sequencing instruments [26]. In addition, the introduction of Long-read RNA-seq, also known as third-generation sequencing, and direct RNA-seq (dRNseq) have made the transcriptomics more thorough [27, 28] without requiring prior information on the RNA sequence [29, 30]. The introduction of single-cell RNA sequencing in 2009 helped scientists map and generate libraries for individual cells [31, 32]. In 2015, Drop-seq, for RNA analysis of a large population of individuals cells at once, and InDrop, for labelling and mapping single cell, have given more diverse ways of transcriptome analysis [32]. Single-cell combinational indexing RNA sequencing (Sci-RNA-seq) in 2017 is a two-step combinatorial barcoding method designed to profile single-cell and single-nucleus transcriptomes with no single-cell isolation step necessary [33]. Split-pool ligation-based transcriptome sequencing (SPLiTseq) in 2018 was designed to interpret the cellular origin of RNA using combinational barcoding [34].

The method introduces advantages compared to the previously discussed methods by providing a detailed understanding of the transcriptome through the quantitative measuring of the gene expression, splicing, maternal or paternal allele expression and altogether, helps to interpret the cause of diseases efficiently with lower cost.

After the laboratory-based workflow, computational analysis, most importantly data processing and analysis, is carried out using various computational tools. Data processing can be performed for both organisms with and without reference genomes. The organisms with a reference genome, short RNA sequencing reads are mapped using the reference genome. On the other hand, for the organisms with no reference genome, de novo transcriptome assembly is applied [35, 36].

This review provides past and current research studies on RNA-seq, and its types focus on the advantages and disadvantages of the technique. Furthermore, it presents the use of the method in cancer as well as rare diseases. Additionally, it introduces the future possibilities of RNseq and its application for understanding disease origin and development in more detail. Finally, the paper gives a brief description of RNAseq application for different types of cancer, rare diseases, and COVID-19 (Coronavirus Disease 2019), that have been challenging medical professionals in finding the most effective way for diagnosis and treatment with least side effects and we hope to shed light on utilizing the technique for more useful and accurate protocol to minimize the error and enhance the therapies and eventually, the diseases' prognosis.

RNA sequencing methods

RNA sequencing techniques can be categorized based on the library preparation methods and the applied approach into short-read sequencing, long-read cDNA sequencing, and long-read direct RNA sequencing. Although short-read and long-read cDNA techniques follow almost many steps, in the same manner, the quantity of sample and computational analysis of the techniques at the beginning and the end of library preparation is different. While short-read sequencing of cDNA provides Short Read Archive (SRA) that consists of almost all sequenced mRNA data [37], long-read cDNA sequencing has helped scientists to develop transcript data with their diverse isoforms [38]. The following information presents current published knowledge on short-read cDNA sequencing, long-read cDNA, and direct RNA sequencing.

Short-read CDNA sequencing

This method has replaced microarray in RNAs gene expression with less cost and more straightforward application with a higher quality of data through the transcriptome [39]. The commonly used platform under this category is via transcript's reversible terminator sequence and synthesis techniques [40, 41]. Like all other techniques, the technique is carried out on platforms such as IonTorrent and Illumina and performs RNA sequencing analysis through an indirect method using cDNA. And the method includes RNA extraction, mRNA enrichment, mRNA fragmentation, cDNA synthesis, cDNA fragmentation, cDNA amplification, sequencing, and data analysis [42].The base pair banding of mRNA fragments for the technique is 150- to 200 bps for library purification and preparation, and therefore, the prepared cDNA is mainly between 200 and 400 bps [43]. A short-read sequencing library is prepared with an average of 20–30 million reads for each sample. After the complete sequencing, the library is purified by computational processing to identify the reads aligned with the targeted individual transcripts.

This method helps to report an association of intra-platform with inter-platform [44, 45]. Nevertheless, limitations due to possible occurring errors during sample preparation and computational analysis may cause false reports in the identification and quantification of diverse forms of isoforms that are manifested from a gene [46], especially the transcripts with a large number of base pairs such as the ones found in humans [47]. Therefore, it is understandable that short-read RNA sequencing is not fully efficient to perform a complete analysis of long transcripts [48]. In addition to the limitations of RNA size, multi-mapped reads are not accurate. Long-read sequencing has lifted the limitations of size by tagging full-length cDNA and the use of unique molecular identifiers that are copied along with cDNA prior to library preparation (UMIs) [49, 50].

Long-read CDNA sequencing

As mentioned previously, short-read sequencing requires the assembly of short RNA fragment reads, which affects the accuracy of the genome mapping process and the whole sequence cannot be identified and analyzed. However, long-read sequencing can identify large-size RNA and process the full length, making genome mapping possible for mammalian cells containing 1–2 kb of transcripts and may surpass 100 kb [51–53]. The method is performed on a number of platforms that were developed in the past few years, and ones are Single-Molecule Real-Time (SMRT) technology from Pacific Biosciences (PacBio sequencing) and protein nanopore sequencing technology from Oxford Nanopore Technologies (ONT).

The standard protocol includes conversion of high-quality RNA to full-length cDNA by template-switching reverse transcriptase [54], and the cDNA undergoes amplification by polymerase chain reaction (PCR) to prepare the SMRT library [54]. While the ONT platform follows the same protocol as PacBio [55], reverse transcriptase was shown to affect library preparation and the length of transcript read on ONT [56]. In contrast to the advantages, long-read cDNA sequencing requires a great amount of time for the large size of the genome to be processed [57], and therefore, further studies are necessary to optimize the time.

Long-read direct RNA sequencing

Unlike short-read and long-read cDNA sequencing, long-read RNA sequencing, also known as dRNA-seq (DRS), does not require cDNA generation and therefore can eliminate the errors that occur during cDNA amplification and avoid RNA-RNA chimaeras produced by cDNA [58]. Although the limitation of reading length is not the challenge with the technique, the fragmentation of the input read is still challenging [59, 60]. The technique is carried out on nanopore sequencing technology developed by ONT [43, 61]. The process includes two ligation steps. The first ligation step includes ligation of duplex adaptor to polyA tail of RNA, followed by reverse-transcription followed by the second ligation step, which is the attachment of the motor protein-attached sequencing adaptor. Finally, the products go through library preparation [62]. The other advantage of DRS over the other two lies in the ability of the technique

to identify the RNA base modifications, and thus can shed light on the epigenetics of the species [62, 63].

RNA sequencing in viral diseases, cancer, and rare diseases

RNA sequencing has provided an effective approach in detecting different types of cancers and rare diseases and, thus, has shed light on developing more effective treatments. DRS has been applied for genomic studies of viral transcriptomes, and it uses cDNA to analyse and interpret viral RNA [64–66]. Previous studies applied the technique to investigate human poly(A) RNA and DNA-based viruses [67]. A recent study has shown full-length sequencing of HCoV-229E virus that belongs to the coronavirus family and encompasses the known largest RNA genome. In this study, the technique used defective interfering RNAs (DI-RNAs) for in vitro analysis of transcript using full-length cDNA [68]. In this study, in patients who manifested resistance during therapy, RNA-seq detected human gemcitabine-resistant pancreatic cancer cells (PANC1) as potential therapeutic targets [69].

In addition to the discussed RNA sequencing methods, in situ RNA sequencing was developed to perform RNA sequencing inside the cell without cell lysis and RNA extraction [70]. The study on breast cancer applied the technique to analyse short RNA fragments of ACTB gene and HER2 (abundant growth-promoting protein outside breast cells) RNA in preserved cells and tissues and helped to detect tissue heterogeneity at a molecular level [70]. Despite all new inventions and advancements in medicine, cancer remains elusive and is considered one of the most life-threatening malignant diseases. With the development of RNA sequencing as one of the high-throughput methods of transcriptome analysis, interpretation of diseases and their genetic causes at the molecular level has been conceivable. Single-cell RNA sequencing, known as scRNA-seq, has been used to analyse single malignant cell's heterogeneity to present the cause of cancers [69, 71, 72], such as pancreatic ductal adenocarcinoma [69]. RNA-seq can find out the uses of tumour mutational burden (TMB), whose study is noteworthy as a possible immune checkpoint biomarker and helps in treatment and cancer prognosis [73]. By detecting a mutation in MET proto-oncogene and isocitrate dehydrogenase 1 (IDH1) gene using RNA-seq, the possibility of designing a better therapy for lung adenocarcinoma and chondrosarcoma has been made possible [74, 75]. Therefore, the technique has facilitated target therapy by detecting the causative gene or the mutation of target genes in various types of cancer, such as acute myeloid leukaemia (AML) [76, 77]. In head and neck cancer [78] and oligodendroglioma [79], single-cell RNA sequencing has helped to elucidate the difference between malignant and benign cells using the data collected for copy number variations (CNV). Besides applications of RNA-seq in treatment design for cancers, the tool can be used as a diagnostic tool in blood-based sarcoma [80]. Although this review has covered limited past and present studies on RNA-seq in cancer diagnosis and target therapy, it is abundantly clear that the tool in identifying the genetic and epigenetic cause of cancer, assisting in better therapy design by detecting the resistant genes, and elucidating the mutations in the genes as cancer biomarkers for better therapy.

The advantages of RNA-seq extend to in better understanding of rare diseases. Over 7000 rare Mendelian disorders have been identified so far. However, the genetic basis of more than half of all Mendelian diseases reported remains elusive, despite being monogenic [81]. Furthermore, these diseases can show variable phenotypes even in cases where the causal disease gene is identified, even in patients such as siblings [82, 83], which presents diagnostic and patient management challenges [84]. RNA-seq offers the ability to calculate allele-specific expressions that are likely to expose the existence of a broad heterozygous regulatory, splicing, nonsense variant or epimutation to help identify candidate rare disease genes and variants [85–90]. Table 1 introduces some of the rare diseases that are investigated using RNA-seq.

Advantages of RNA-seq extend in better understanding of rare diseases. Over 7000 rare Mendelian disorders have been identified so far. The genetic basis of more than half of all Mendelian diseases reported remains elusive, despite being monogenic [93]. These diseases can show variable phenotypes even in cases where the causal disease gene is identified, even in patients such as siblings [94, 95] with the same genetic mutation, which presents in diagnostic and patient management challenges [96]. RNA-seq offers the ability to calculate allele-specific expression that are likely to expose the existence of a broad heterozygous regulatory, splicing or nonsense variant or epimutation to help identify candidate rare disease genes and variants [97–100]. Table 1 introduces some of the rare diseases that were investigated using RNA-seq.

Conclusion and future perspectives

Advancement in RNA-seq has been one of the major revolutions in the study and interpretation of transcriptome in the past few years. With ongoing innovation and development in bioinformatics, data analysis software and platform technologies, cataloguing full-length transcript and library preparation for all organisms, whether single-cell organisms, such as yeast to mammalians, many questions elude scientists carry out further investigations on various physiological and genetic abnormalities can be answered.

Table 1 Applications of RNA sequencing in identification and diagnosis of rare diseases

RNA-seq application	Disease	Genetic abnormality	Gene's function
Measureing allele-specific expression with whole-blood RNA-seq	Idiopathic cardiomyopathy	Mutation in EFHD2 gene	EFHD2 encodes Ca^{2+} protein that maintains B cell death-cell programming, activation of immune cells, immune cell motility
Identification of casual genes with whole-blood RNA-seq	enoyl CoA reductase protein-associated neurodegeneration (MEPAN) [91]	Biallelic heterozygous pathogenic variant in MECR gene	Regulates the motor skills and is likely to be involved in peroxisome-proliferator-activator-receptor (PPAR)-dependent signalling [92]
Identification of variants in regulatory upstream regions of genes in monogenetic neuromuscular disorders	Congenital Muscular Dystrophy (CMD) [93, 94]	Heterozygous variant in GDP-Mannose Pyrophosphorylase B (GMPPB) gene [93]	Regulates protein, fructose, and mannose metabolism and impairment on the gene causes defective o-glycosylation of α-dystroglycan [93]
Diagnosis of Mendelian rare diseases by detecting splice-affecting variant	collagen VI dystrophy [95, 96]	Intron inclusion in COL6A1 gene [95]	Encodes collagen VI that causes muscle weakness and deformities of joints [97]
Diagnosis of Mendelian rare diseases	Duchenne Muscular Dystrophy (DMD) [98]	Heterozygous variant in DMD gene [94, 95]	Encodes dystrophin protein that forms dystrophin-glycoprotein complex in extracellular matrix [99]

Moreover, library preparation has kept the information accessible to those who are researching transcript-related studies. Furthermore, the researcher can use this tool in comparing the tissues and cells in normal and abnormal conditions to track and reveal the causatives of different diseases and identify metablic abnormalities or alterations that happen in molecular and cellular levels and identify metabolic abnormalities or alterations that happen in molecular and cellular levels.

The current outbreak of COVİD-19 and the emergence of variants in short-term time have been a challenge for the researchers in finding a better tool in interpreting the full-length RNA of the SARS-CoV-2 to develop a more efficient treatment and durable vaccine. And RNA sequencing with the advantage of reading a large-size transcript has provided an insight into developing a platform that can help in a detailed analysis of SARS-CoV-2 RNA to reveal the cause of genetic variation and resistance towards the currently used treatments. Besides, the diagnostic tools are critical in cancer and rare diseases and with ongoing improvement in RNA sequencing techniques and existing diagnostic tools for some diseases, it is expected to see great advancements in developing standard diagnostic tools that benefit the biomarkers of disease that are being detected by RNA sequencing. Not to mention that the collection of all data from different organisms' transcriptomes can improve the field of taxonomy by aligning the sequenced transcripts and measuring the level of similarities among the organisms. Therefore, it is expected that the unknown and undefined forms of isoforms can be determined and eventually help the unidentified genes' function and full potentials to be uncovered, and questions in molecular and cellular evolution and diversity of many pathogenic viruses will be answered.

Although this review has covered limited present and past studies and achievements on applications and advantages of RNA-seq, it is hoped that the readers of the review will benefit from the collected information and shed light on future applications of RNA sequencing in better understanding of genetically diversified human diseases.

Funding Not applicable.

Data availability Not applicable.

Declarations

Conflict of interest The author declares that they have no conflict of interest.

Code availability Not applicable.

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent for publication Authors have full consent of the publication.

References

1. Crick F (1970) Central dogma of molecular biology. Nature 227:561–563
2. Crick FH (1958) On protein synthesis. Symp Soc Exp Biol 12:138–163
3. Wang B, Kumar V, Olson A, Ware D (2019) Reviving the transcriptome studies: an insight into the emergence of single-molecule transcriptome sequencing. Front Genet 10:384. https://doi.org/10.3389/fgene.2019.00384
4. Saliba AE, Westermann AJ, Gorski SA, Vogel J (2014) Single-cell RNA-seq: advances and future challenges. Nucleic Acids Res 42:8845–8860. https://doi.org/10.1093/nar/gku555
5. Hangauer MJ, Vaughn IW, McManus MT (2013) Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. PLoS Genet 9:e1003569. https://doi.org/10.1371/journal.pgen.1003569
6. Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublomme JT, Raychowdhury R, Schwartz S, Yosef N, Malboeuf C, Lu D et al (2013) Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature 498:236–40. https://doi.org/10.1038/nature12172
7. Deng Q, Ramskold D, Reinius B, Sandberg R (2014) Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian ncells. Science 343:193–6. https://doi.org/10.1126/science.1245316
8. Bengtsson M, Stahlberg A, Rorsman P, Kubista M (2005) Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. Genome Res 15:1388–92. https://doi.org/10.1101/gr.3820805
9. Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S (2008) Transcriptome wide noise controls lineage choice in mammalian progenitor cells. Nature 453:544–7. https://doi.org/10.1038/nature06965
10. Mattick JS, Makunin IV (2006) Non-coding RNA. Hum Mol Genet. 15 Spec No 1:R17–R29
11. Stefani G, Slack FJ (2008) Small non-coding RNAs in animal development. Nat Rev Mol Cell Biol 9:219–230
12. Wilusz JE, Sunwoo H, Spector DL (2009) Long noncoding RNAs: Functiona surprises from the RNA world. Genes Dev 23:1494–1504
13. Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467–470
14. Okoniewski MJ, Miller CJ (2006) Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations. BMC Bioinformatics 7:276
15. Casneuf T, Van de Peer Y, Huber W (2007) In situ analysis of cross-hybridisation on microarrays and the inference of expression correlation. BMC Bioinformatics 8:461
16. Shendure J (2008) The beginning of the end for microarrays? Nat Methods 5:585–587
17. Itoh K, Matsubara K, Okubo K (1994) Identification of an active gene by using large-scale cDNA sequencing. Gene 140:295–296
18. Shiraki T, Kondo S, Katayama S, Waki K, Kasukawa T, Kawaji H, Kodzius R, Watahiki A, Nakamura M, Arakawa T et al (2003) Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proc Natl Acad Sci 100:15776–15781
19. Kukurba KR, Montgomery SB (2015) RNA Sequencing and Analysis. Cold Spring Harb Protoc 11:951–969. https://doi.org/10.1101/pdb.top084970
20. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res 18:1509–1517
21. Emrich SJ, Barbazuk WB, Li L, Schnable PS (2007) Gene discovery and annotation using LCM-454 transcriptome sequencing. Genome Res 17:69–73
22. Lister R et al (2008) Highly integrated single- base resolution maps of the epigenome in arabidopsis. Cell 133:523–536
23. Schadt EE, Turner S, Kasarskis A (2010) A window into third-generation sequencing. Hum Mol Genet 19:R227–R240
24. Morin R et al (2008) Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. Biotechniques 45:81–94
25. Nagalakshmi U et al (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320:1344–1349
26. Lai Y (2010) Differential expression analysis of digital gene expression data: RNA-tag filtering, comparison of t-type tests and their genome-wide co-expression-based adjustments. Int J Bioinform Res Appl 6(4):353–365. https://doi.org/10.1504/IJBRA.2010.035999
27. Garalde DR et al (2018) Highly parallel direct RNA sequencing on an array of nanopores. Nat Methods 15:201–206
28. Byrne A et al (2017) Nanopore long- read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. Nat Commun 8:16027
29. Wang Z, Gerstein M, Snyder M (2009) RNA-seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63
30. Zhao S, Fung-Leung W-P, Bittner A, Ngo K, Liu X (2014) Comparison of RNA-seq and microarray in transcriptome profiling of activated t cells. PLoS One 9:e78644
31. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N et al (2009) mRNASeq whole-transcriptome analysis of a single cell. Nat Methods 6:377–382
32 AlJanahi AA, Danielsen M, Dunbar CE (2018) An Introduction to the Analysis of Single-Cell RNA-Sequencing Data. Molecular Therapy Methods Clin Dev 10:189–196. https://doi.org/10.1016/j.omtm.2018.07.003
33. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Regev AK, McCarroll SAA (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 161(5):1202–1214. https://doi.org/10.1016/j.cell.2015.05.002
34. Cao J, Packer J, Ramani V, Cusanovich D, Huynh C, Daza R et al (2017) Comprehensive single-cell transcriptional profiling of a multicellular organism. Science 357:661–667
35. Rosenberg A, Roco C, Muscat R, Kuchina A, Sample P, Yao Z et al (2018) Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science 360:176–182
36. Engström PG, Steijger T, Sipos B, Grant GR, Kahles A, The RGASP Consortium, Rätsch G, Goldman N, Hubbard TJ, Harrow J et al (2013) Systematic evaluation of spliced alignment programs for RNA-seq data. Nat Methods 10:1185–1191
37. Li WV, Li JJ (2018) Modelling and analysis of RNA-seq data: a review from a statistical perspective. Quant Biol 6(3):195–209. https://doi.org/10.1007/s40484-018-0144-7
38. Leinonen R, Sugawara H, Shumway M (2011) The sequence read archive. Nucleic Acids Res 39:D19–D21
39. Panda K, Slotkin RK (2020) Long-read cDNA sequencing enables a "Gene-Like" transcript annotation of transposable elements. Plant Cell 32(9):2687–2698. https://doi.org/10.1105/tpc.20.00115

40. Xiang CC, Chen Y (2000) cDNA microarray technology and its applications. Biotechnol Adv 18(1):35–46. https://doi.org/10.1016/s0734-9750(99)00035-x
41. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IM, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, Aniebo IC, Bailey DM, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA, Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Catenazzi CEM, Chang S, Neil Cooley R, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Scott Furey W, George D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, Hoschler K, Hurwitz S, Ivanov DV, Johnson MQ, James T, Huw Jones TA, Kang GD, Kerelska TH, Kersey AD, Khrebtukova I, Kindwall AP, Kingsbury Z, Kokko-Gonzales PI, Kumar A, Laurent MA, Lawley CT, Lee SE, Lee X, Liao AK, Loch JA, Lok M, Luo S, Mammen RM, Martin JW, McCauley PG, McNitt P, Mehta P, Moon KW, Mullens JW, Newington T, Ning Z, Ling Ng B, Novo SM, O'Neill MJ, Osborne MA, Osnowski A, Ostadan O, Paraschos LL, Pickering L, Pike AC, Pike AC, Chris Pinkard D, Pliskin DP, Podhasky J, Quijano VJ, Raczy C, Rae VH, Rawlings SR, Chiva Rodriguez A, Roe PM, Rogers J, Rogert Bacigalupo MC, Romanov N, Romieu A, Roth RK, Rourke NJ, Ruediger ST, Rusman E, Sanches-Kuiper RM, Schenker MR, Seoane JM, Shaw RJ, Shiver MK, Short SW, Sizto NL, Sluis JP, Smith MA, Ernest Sohna Sohna J, Spence EJ, Stevens K, Sutton N, Szajkowski L, Tregidgo CL, Turcatti G, Vandevondele S, Verhovsky Y, Virk SM, Wakelin S, Walcott GC, Wang J, Worsley GJ, Yan J, Yau L, Zuerlein M, Rogers J, Mullikin JC, Hurles ME, McCooke NJ, West JS, Oaks FL, Lundberg PL, Klenerman D, Durbin R, Smith AJ (2008) Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456(7218):53–59. https://doi.org/10.1038/nature07517
42 Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next generation sequencing technologies. Nat Rev Genet 17(6):333–351. https://doi.org/10.1038/nrg.2016.49
43. Cartolano M, Huettel B, Hartwig B, Reinhardt R, Schneeberger K (2016) cDNA library enrichment of full length transcripts for SMRT long read sequencing. PLoS One. https://doi.org/10.1371/journal.pone.0157779
44. Martin Hölzer and Manja Marz. Software Dedicated to Virus Sequence Analysis "Bioinformatics Daniel R Garalde, Elizabeth A Snell, Daniel Jachimowicz, Botond Sipos, Joseph H Lloyd, Mark Bruce, Nadia Pantic, Tigist Admassu, Phillip James, Anthony Warland, Michael Jordan, Jonah Ciccone, Sabrina Serra, Jemma Keenan, Samuel Martin, Luke McNeill, E Jayne Wallace, Lakmal Jayasinghe, Chris Wright, Javier Blasco, Stephen Young, Denise Brocklebank, Sissel Juul, James Clarke, Andrew J Heron, and Daniel J Turner. (2018) Highly parallel direct RNA sequencing on an array of nanopores. Nat Methods, 15: 201–206, ISSN 1548–7105. https://doi.org/10.1038/nmeth.4577. Goes Viral". Adv Virus Res.
45. .Su Z et al (2014) A comprehensive assessment of RNA- seq accuracy, reproducibility and information content by the sequencing quality control consortium. Nat Biotechnol 32:903–914
46. Li S et al (2014) Multi- platform assessment of transcriptome profiling using RNA- seq in the ABRF next- generation sequencing study. Nat Biotechnol 32:915–925
47. Djebali S et al (2012) Landscape of transcription in humancells. Nature 489:101–108
48. Frankish A et al (2019) GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res 47:D766–D773
49. Tilgner H et al (2018) Microfluidic isoform sequencing shows widespread splicing coordination in the human transcriptome. Genome Res 28:231–242
50. Fu GK, Hu J, Wang P-H, Fodor SP (2011) Counting individual DNA molecules by the stochastic attachment of diverse labels. Proc Natl Acad Sci USA 108:9026–9031
51. Islam S et al (2014) Quantitative single- cell RNA- seq with unique molecular identifiers. Nat Methods 11:163–166
52. Kovaka S, Zimin AV, Pertea GM et al (2019) Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol 20:278. https://doi.org/10.1186/s13059-019-1910-1
53. Amarasinghe SL, Su S, Dong X et al (2020) Opportunities and challenges in long-read sequencing data analysis. Genome Biol 21:30. https://doi.org/10.1186/s13059-020-1935-5
54. Thomas S, Underwood JG, Tseng E, Holloway AK (2014) Long-read sequencing of chicken transcripts and identification of new transcript isoforms. PLOS One 9:e94650
55. Ramsköld D et al (2012) Full- length mRNA- Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol 30:777–782
56. Oikonomopoulos S, Wang YC, Djambazian H, Badescu D, Ragoussis J (2016) Benchmarking of the Oxford Nanopore MinION sequencing for quantitative and qualitative assessment of cDNA populations. Sci Rep 6:31602
57. Prazsák I et al (2018) Long- read sequencing uncovers a complex transcriptome topology in varicella zoster virus. BMC Genomics 19:873
58. Workman RE et al (2018) Nanopore native RNA sequencing of a human poly(A) transcriptome. Preprint at bioRxiv. https://doi.org/10.1101/459529
59. Karst SM, Dueholm MS, McIlroy SJ, Kirkegaard RH, Nielsen PH, Albertsen M (2018) Retrieval of a million high-quality, full-length microbial 16S and 18S rRNA gene sequences without primer bias. Nat Biotechnol 36:190–195. https://doi.org/10.1038/nbt.4045
60. Mikheyev AS, Tin MMY (2014) A first look at the Oxford Nanopore MinION sequencer. Mol Ecol Resour 14(6):1097–1102
61. Chua EW, Ng PY, MinION (2016) A novel tool for predicting drug hypersensitivity? Front Pharmacol 7:156. https://doi.org/10.3389/fphar.2016.00156
62. Jain M, Olsen HE, Paten B, Akeson M (2016) The oxford nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol 17:239
63. .Weirather JL et al (2017) Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. F1000Res 6:100
64. Wongsurawat T, Jenjaroenpun P, Wassenaar TM, Taylor D (2018) Decoding the epitranscriptional landscape from native RNA sequences. Preprint at bioRxiv. https://doi.org/10.1101/487819
65. Moldován N, Tombácz D, Sz˝ucs A, Csabai Z, Snyder M, Boldogkoi Z (2018) Multi-platform sequencing approach reveals a novel transcriptome profile in pseudorabies virus. Front Microbiol 8:2708
66. Tombácz D, Csabai Z, Sz˝ucs A, Balázs Z, Moldován N, Sharon D, Snyder M, Boldogkoi Z (2017) Long-read isoform sequencing reveals ˝ a hidden complexity of the transcriptional landscape of herpes simplex virus type 1. Front Microbiol 8:1079. https://doi.org/10.3389/fmicb.2017.01079

67. Moldován N, Balázs Z, Tombácz D, Csabai Z, Sz˝ucs A, Snyder M, Boldogkoi Z (2017) Multi-platform analysis reveals a complex ˝ transcriptome architecture of a circovirus. Virus Res 237:37–46. https://doi.org/10.1016/j.virusres.2017.05.010
68. Depledge DP, Puthankalam SK, Sadaoka T, Beady D, Mori Y, Placantonakis D, Mohr I, Wilson A (2018) Native RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. bioRxiv 373522
69. Viehweger A, Krautwurst S, Lamkiewicz K, Madhugiri R, Ziebuhr J, Hölzer M, Marz M (2019) (2019) Direct RNA nanopore sequencing of full-length coronavirus genomes provides novel insights into structural variants and enables modification analysis. Genome Res 29(9):1545–1554. https://doi.org/10.1101/gr.247064.118
70. Sharma A, Cao EY, Kumar V, Zhang X, Leong HS, Wong AML et al (2018) (2018) Longitudinal single-cell RNA sequencing of patient-derived primary cells reveals drug-induced infidelity in stem cell hierarchy. Nat Commun 9:4931
71. Cieślik M, Chinnaiyan AM (2018) Cancer transcriptome profiling at the juncture of clinical translation. Nat Rev Genet 19:93–109
72. Suva ML, Tirosh I (2019) Single-cell RNA sequencing in cancer: lessons learned and emerging challenges. Mol Cell 75:7–12
73. Huang XT, Li X, Qin PZ, Zhu Y, Xu SN, Chen JP (2018) Technical advances in single-cell RNA sequencing and applications in normal and malignant hematopoiesis. Front Oncol
74. Wang L, Ge J, Lan Y, Shi Y, Luo Y, Tan Y et al (2020) (2020) Tumor mutational burden is associated with poor outcomes in diffuse glioma. BMC Cancer 20:213
75. Seo JS, Ju YS, Lee WC, Shin JY, Lee JK, Bleazard T et al (2012) The transcriptional landscape and mutational profile of lung adenocarcinoma. Genome Res 22:2109–2119
76. Nakagawa M, Nakatani F, Matsunaga H, Seki T, Endo M, Ogawara Y et al (2019) Selective inhibition of mutant IDH1 by DS-1001b ameliorates aberrant histone modifications and impairs tumor activity in chondrosarcoma. Oncogene 38:6835–6849
77. Yu J, Jiang PYZ, Sun H, Zhang X, Jiang Z, Li Y et al (2020) Advances in targeted therapy for acute myeloid leukemia. Biomark Res 8:17
78. Qi Z, Barrett T, Parikh AS, Tirosh I, Puram SV (2019) Single-cell sequencing and its applications in head and neck cancer. Oral Oncol. 99:104441. https://doi.org/10.1016/j.oraloncology.2019.104441
79. Puram SV, Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S et al (2017) Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. Cell 171(1611–24):e24
80. Tirosh I, Venteicher AS, Hebert C, Escalante LE, Patel AP, Yizhak K, Fisher JM, Rodman C, Mount C, Filbin MG, Neftel C, Desai N, Nyman J, Izar B, Luo CC, Francis JM, Patel AA, Onozato ML, Riggi N, Livak KJ, Gennert D, Satija R, Nahed BV, Curry WT, Martuza RL, Mylvaganam R, Iafrate AJ, Frosch MP, Golub TR, Rivera MN, Getz G, Rozenblatt-Rosen O, Cahill DP, Monje M, Bernstein BE, Louis DN, Regev A, Suvà ML (2016) Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. Nature 539(7628):309–313. https://doi.org/10.1038/nature20123
81. Heinhuis KM, In 't Veld SGJG, Dwarshuis G, van den Broek D, Sol N, Best MG, Coevorden FV, Haas RL, Beijnen JH, van Houdt WJ, Würdinger T, Steeghs N (2020) RNA-sequencing of tumor-educated platelets, a novel biomarker for blood-based sarcoma diagnostics. Cancers 12(6):1372. https://doi.org/10.3390/cancers12061372
82. Boycott KM, Rath A, Chong JX, Hartley T, Alkuraya FS, Baynam G, Lochmüller H (2017) International cooperation to enable the diagnosis of all rare genetic diseases. Am J Human Genet 100(5):695–705. https://doi.org/10.1016/j.ajhg.2017.04.003
83. Rudnik-Schöneborn S, Barisić N, Eggermann K, Ortiz Brüchle N, Grđan P, Zerres K (2016) Distally pronounced infantile spinal muscular atrophy with severe axonal and demyelinating neuropathy associated with the S230L mutation of SMN1. Neuromuscul Disord 26(2):132–135. https://doi.org/10.1016/j.nmd.2015.12.003
84. Davidson BA, Hassan S, Garcia EJ, Tayebi N, Sidransky E (2018) Exploring genetic modifiers of Gaucher disease: the next horizon. Hum Mutat 39(12):1739–1751. https://doi.org/10.1002/humu.23611
85. Missaglia S, Tasca E, Angelini C, Moro L, Tavian D (2015) Novel missense mutations in PNPLA2 causing late onset and clinical heterogeneity of neutral lipid storage disease with myopathy in three siblings. Mol Genet Metab 115(2–3):110–117. https://doi.org/10.1016/j.ymgme.2015.05.001
86. Kremer L, Bader D, Mertes C et al (2017) Genetic diagnosis of Mendelian disorders via RNA sequencing. Nat Commun 8:15824. https://doi.org/10.1038/ncomms15824
87. Albers C, Paul D, Schulze H et al (2012) Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome. Nat Genet 44:435–439. https://doi.org/10.1038/ng.1083
88. Barbosa M, Joshi RS, Garg P, Martin-Trujillo A, Patel N, Jadhav B, Watson CT, Gibson W, Chetnik K, Tessereau C, Mei H, De Rubeis S, Reichert J, Lopes F, Vissers LELM, Kleefstra T, Grice DE, Edelmann L, Soares G, Maciel P, Brunner HG, Buxbaum JD, Gelb BD, Sharp AJ (2018) Identification of rare de novo epigenetic variations in congenital disorders. Nat Commun. 9(1):2064. https://doi.org/10.1038/s41467-018-04540-x
89. Frésard L, Smail C, Ferraro NM, Teran NA, Li X, Smith KS, Bonner D, Kernohan KD, Marwaha S, Zappala Z, Balliu B, Davis JR, Liu B, Prybol CJ, Kohler JN, Zastrow DB, Reuter CM, Fisk DG, Grove ME, Davidson JM, Hartley T, Joshi R, Strober BJ, Utiramerur S, Undiagnosed Diseases Network, Care4Rare Canada Consortium, Lind L, Ingelsson E, Battle A, Bejerano G, Bernstein JA, Ashley EA, Boycott KM, Merker JD, Wheeler MT, Montgomery SB (2019) Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. Nat Med 25(6):911–919. https://doi.org/10.1038/s41591-019-0457-8
90. Kroczek C et al (2010) Swiprosin-1/EFhd2 controls B cell receptor signaling through the assembly of the B cell receptor, Syk, and phospholipase C gamma2 in membrane rafts. J Immunol 184:3665–3676
91. Dütting S, Brachs S, Fraternal MD (2011) twins: swiprosin-1/EFhd2 and Swiprosin-2/EFhd1, two homologous EF-hand containing calcium binding adaptor proteins with distinct functions. Cell Commun Signal 9:2
92. Heimer G et al (2016) MECR mutations cause childhood-onset dystonia and optic atrophy, a mitochondrial fatty acid synthesis disorder. Am J Human Genet 99:1229–1244
93. Gonorazky HD, Naumenko S, Ramani AK, Nelakuditi V, Mashouri P, Wang P, Kao D, Ohri K, Viththiyapaskaran S, Tarnopolsky MA, Mathews KD, Moore SA, Osorio AN, Villanova D, Kemaladewi DU, Cohn RD, Brudno M, Dowling JJ. (2019) Expanding the boundaries of RNA sequencing as a diagnostic tool for rare mendelian disease. Am J Hum Genet 104(3):466–483. https://doi.org/10.1016/j.ajhg.2019.01.012. Epub 2019 Feb 28. Erratum in: Am J Hum Genet. 2019 May 2;104(5):1007
94. Belaya K, Rodríguez Cruz PM, Liu WW, Maxwell S, McGowan S, Farrugia ME, Petty R, Walls TJ, Sedghi M, Basiri K, Yue WW, Sarkozy A, Bertoli M, Pitt M, Kennett R, Schaefer A, Bushby K, Parton M, Lochmüller H, Palace J, Muntoni F, Beeson D (2015) Mutations in GMPPB cause congenital myasthenic

syndrome and bridge myasthenic disorders with dystroglycanopathies. Brain 138(Pt 9):2493–2504. https://doi.org/10.1093/brain/awv185

95. Carss KJ, Stevens E, Foley AR et al (2013) Mutations in GDP-mannose pyrophosphorylase B cause congenital and limb-girdle muscular dystrophies associated with hypoglycosylation of α-dystroglycan. Am J Hum Genet 93(1):29–41. https://doi.org/10.1016/j.ajhg.2013.05.009
96. Cummings BB, Marshall JL, Tukiainen T, Lek M, Donkervoort S, Foley AR, Bolduc V, Waddell LB, Sandaradura SA, O'Grady GL, Estrella E, Reddy HM, Zhao F, Weisburd B, Karczewski KJ, O'Donnell-Luria AH, Birnbaum D, Sarkozy A, Hu Y, Gonorazky H, Claeys K, Joshi H, Bournazos A, Oates EC, Ghaoui R, Davis MR, Laing NG, Topf A; Genotype-Tissue Expression Consortium, Kang PB, Beggs AH, North KN, Straub V, Dowling JJ, Muntoni F, Clarke NF, Cooper ST, Bönnemann CG, MacArthur DG. (2017) Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. Sci Transl Med 9(386):eaal5209. https://doi.org/10.1126/scitranslmed.aal5209
97. .Foley AR, Mohassel P, Donkervoort S, Bolduc V, Bönnemann CG. (2004i) Collagen VI-Related Dystrophies. 2004 Jun 25 [updated 2021 Mar 11]. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Mirzaa G, Amemiya A (eds) GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993–2021
98. Baker NL, Mörgelin M, Peat R, Goemans N, North KN, Bateman JF, Lamandé SR (2005) Dominant collagen VI mutations are a common cause of Ullrich congenital muscular dystrophy. Hum Mol Genet 14(2):279–93. https://doi.org/10.1093/hmg/ddi025
99 Nowak KJ, Davies KE (2004) Duchenne muscular dystrophy and dystrophin: pathogenesis and opportunities for treatment. EMBO Rep 5(9):872–876. https://doi.org/10.1038/sj.embor.7400221
100. White SJ, Aartsma-Rus A, Flanigan KM, Weiss RB, Kneppers AL, Lalic T, Janson AA, Ginjaar HB, Breuning MH, den Dunnen JT (2006) Duplications in the DMD gene. Hum Mutat 27(9):938–945. https://doi.org/10.1002/humu.20367

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.