Rep Radiother Oncol. 2024 December; 10(1):e142195. https://doi.org/10.5812/rro-142195.

Published online 2024 January 13. **Review Article**

Role of Transcriptomics in Precision Oncology

Ruby Srivastava [1,*]

[1]CSIR-Centre for Cellular and Molecular Biology, Hyderabad, India

[*]*Corresponding author*: CSIR-Centre for Cellular and Molecular Biology, Hyderabad, India. Email: amitruby1@gmail.com

Received 2023 October 22; **Revised** 2023 November 19; **Accepted** 2023 December 27.

Abstract

Transcriptome profiling is one of the most widely used approaches in the field of multiomics research. It plays a crucial role in the prognostic, diagnostic, and predictive treatment of cancer patients. Novel next-generation sequencing (NGS) technologies permit the identification of cancer biomarkers, gene signatures, and their abnormal expression, affecting oncogenic and molecular targets and novel biomarkers for cancer therapies. Multiomics studies have changed the overall understanding of cancer and opened a precise perspective for tumor diagnostics and therapy. The use of these approaches has strengthened our understanding of disease pathophysiology and classifications at the molecular level, including specific interference with drug mechanisms of action. Still, it has limited added value in the clinical setting. The omics data on precision medicine include the application of data from genes, transcripts, and proteins for diagnosis, monitoring of diseases, risk factor determination, counseling, and development of novel therapeutics. Bioinformatics applications have expanded statistics-based analysis toward deriving molecular pathways and process models for characterizing phenotypes and drug action mechanisms. In this review, we will discuss transcriptomics and interference analysis that allows the identification of predictive biomarkers at the molecular level to test drug response and analyze the molecular process interface of disease progression-relevant pathophysiology and mechanism of action to propose predictive biomarkers.

Keywords: Transcriptomics, Oncology, Bioinformatics, Drugs, Diagnosis

1. Context

Cancer is a disease that depends on both oncogenic mutations and non-mutated genes for survival, and so named as oncogene and non-oncogene addictions. The reason for improper treatment and death of patients is poor diagnosis and prognosis of the disease, which is defined by the 10 hallmarks of cancers (1), due to which the treatment process is affected. Cancer is developed due to genetic and epigenetic changes, which accumulate within the cell through which the abnormal biological features (1, 2) specific to cancer cells occur. The multiomics provides the basic understanding of precision oncology (3). The major progress in the research of mutations driving cancer cells (4) has happened since the sequencing of the human genome. Mutants play a key role in understanding gene function because mutations hinder cellular processes. The specific mutations responsible for malignant transformation and hereditary cancer syndromes (5) can be identified by genomics. In DNA sequencing, the mutation assays present certain limitations in personalized care (6). If we consider the proteomics, only a small percentage of the human genome is expressed. Also, gene expression is a very complicated and multistage process, which is controlled by various regulating mechanisms, such as DNA methylation (7), DNA-binding proteins (8), or small interfering RNA (siRNA) (9). The determination of a karyotype and genomic hybridization at the molecular level affirms the heterogeneity of cancer cells. DNA sequencing of the cells within the same tumor may differ remarkably to define different subpopulations involving clinical diversity (10, 11). Moreover, during tumor development and treatment, it is observed that the plasticity of cancer genomes exhibits itself in the cell-to-cell flexibility of DNA sequence, which can be used for large-scale clinical assistance (12). So, it can be concluded that the cancer genome may provide worthy information on the DNA sequence and its structure, but cancer genomics is not sufficient to report the real phenotype of the cell. Thus, other approaches (such as proteomics, transcriptomics, and metabolomics) are needed to find proper molecular diagnostic targets and mark specific therapies for cancer patients. Proteomics

is much closer to the determination of molecular mechanisms for the phenotype of cells. Proteins are the key factors in all cellular processes, whereas mutations change their expressions and/or activities in many ways (13, 14). The analysis of the cancer proteome gives a landscape of post-translational modifications and interactions between cellular mechanisms and their locations (15). Microarrays are commonly used for protein recognition. Microarrays use monoclonal antibodies or other binders to identify individual proteins. Protein microarrays have wide applications in molecular diagnostics, specifically in the discovery (16) of cancer biomarkers. Yet the main barrier for proteomic studies at (17) wider levels is the different physical and chemical properties of proteins, as well as the need to use highly specific antibody panels.

The transcriptome consists of "classical" RNAs (messenger RNA [mRNA], ribosomal RNA [rRNA], and transfer RNA [tRNA]) and multiple subtypes of noncoding RNA (microRNA [miRNA] and long noncoding RNA [ncRNA]), which have novel regulatory functions in cell biology (18, 19). Microarray technology or recent transcriptome sequencing (RNA sequencing [RNA-Seq]) methods are used for profiling gene expression. Transcriptomics includes precise details about base pairs (bp) and the ability to detect novel RNAs that cannot be detected on microarrays. Fluorescence in situ hybridization (FISH) and reverse transcriptase-polymerase chain reaction (RT-PCR) are used to detect gene rearrangements in routine, but these techniques are limited, with only testing for 1 gene at a time. RNA-Seq has been performed successfully in recent years (20-23). Applications of RNA-Seq include the detection of novel clinically relevant gene fusions in cancer at the early stage of development. The detection of novel gene fusions leads to novel opportunities in the treatment of advanced-level cancer and therapeutics with kinase inhibitors (18). Roberts et al. recently identified a novel treatment hypothesis for patients with pediatric B-cell acute lymphoblastic type of leukemia in clinical trials and identified kinase fusions involving genes for target therapies (24, 25). Gene expression signatures of clinical relevance can be used to classify cancer types into molecular subsets (26-28). In other studies, microarray-based transcriptome profiling on primary breast cancer samples was divided into 5 molecular subsets to study the biological and clinical relevance of the disease (29).

The various modules of RNA-Seq study the classical elements of the transcriptome, including mRNA, rRNA, and tRNA. Multiple subtypes of RNA with novel regulatory functions in cell biology have been discovered. The major part of the transcriptome consists of ncRNAs, miRNAs (30, 31), siRNAs (32), and long ncRNAs. These novel RNAs play multiple roles in cell biology, including regulation of transcription, post-transcriptional events, gene silencing, translation, and protein-level function, apart from the classical function for mRNAs that encode proteins (33). The role of miRNAs is genomic alterations, including mutation, deletion, amplification, and epigenetic modifications (34). MicroRNAs can function as oncogenes or tumor suppressor genes (35). Small interfering RNAs arbitrate a highly specific gene-silencing mechanism, which is conserved to mammalian biology (33) and used as a potential strategy for gene-silencing therapies in biomedical research (36). Long ncRNAs are pervasive in cancer, have varied regulatory functions, and have been intrinsically identified recently (37). Circular RNA (circRNA) is also a human transcriptome feature and is ubiquitous in many other metazoans. The multiple functions of circRNA include serving as protein scaffolds or miRNA sponges and being translated into polypeptides (38). With a longer half-life and more resistance to ribonuclease (RNase) R than linear RNAs (39), circRNAs are used as potential candidates for diagnostic biomarkers and therapeutic targets (38).

Transcriptome connects the cellular phenotype and genetic characteristics of tumor biology. It contains all the information encoded in RNA transcribed from DNA. While the genome is relatively stable, the transcriptome reacts actively to physiological or pathological conditions. Thus, we can say that the transcriptome-based studies have opened a new era for understanding the cancer mechanisms.

In the next section, various computational bioinformatics tools that are used to perform the transcriptomic analysis will be discussed.

2. Computational Methodology

Over the years, transcriptomics with improved sensitivity, accuracy, and reproducibility with reduced biases have been widely used to understand the cause and effect of biological processes through differential gene-/transcript-/exon-expression analysis. The main steps in RNA-Seq data analysis include experimental design with quality control, read alignment, quantification of gene and transcript levels, visualization, differential expression gene (DEG) analysis, alternative splicing, functional analysis, and gene fusion detection. Being sequence-based, RNA-Seq is useful for identifying genetic variants for expression quantitative trait loci (eQTL) analysis, and even clinical diagnosis. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis, protein-protein interactions, integrated

multiomics data analysis, and drug-gene interaction predictions can suggest personalized treatment for cancer patients. The original sequencing data are taken either by the experimental results or by other public repositories, such as Gene Expression Omnibus (GEO), The Cancer Genome Atlas (TCGA), International Cancer Genome Consortium (ICGC), GEO (40) and European Genome-phenome Archive (EGA) (41), to identify transcriptome-level gene expression datasets with accessible clinical details. The format conversion tools and the extraction of necessary information can be obtained by software packages that have constructed workflows, for example, Anduril (42) and Galaxy (43). The advanced features of these servers include building complex pipelines with large datasets that need automated correlation. The strength of these web servers is the user-friendly environment and easier visualization of data with highly advanced tools.

The raw RNA-Seq reads were extracted from the Sequence Read Archive (SRA) files and converted into FASTQ files. For longer reads of the fragment sequence, adapters may also be present, and trimming these will improve the number of reads mapped. The FASTQC tool was used to generate the quality report for the sequence, and MultiQC was used to collect the generated reports. Trimmomatic (44) and Cutadapt (45) are used to remove the adapter and quality trimming. Phred quality score $>$ 20 and read length $>$ 50 bp (46) are selected for the downstream analysis of reads. The statistical comparisons among the trimming algorithms (read mapping rate and the surviving reads) were performed using the Kruskal-Wallis test, followed by the dunn test package (47) in R (48). For *Homo sapiens*, GRCh38 (hg38) from Ensembl65 was used as the reference genome. The read alignment was carried out by Tophat2 (49), STAR (50) and Hisat2 (51). Transcriptome alignment methods were also tested by Bowtie2 (52) and STAR against the Ensembl (v82) transcriptome. BAM files were sorted by read name and genome position using SAMtools (53), and unmapped reads were discarded. The results from the BAM files were visualized using Integrative Genomics Viewer (IGV) (54). Quantification depends on both the reference genome (the FASTA file) and its associated annotations (the GTF file). Counting methods, such as Cuffinks (55), eXpress (56), HTSeq (57), RSEM (58), and Stringtie (59), were used for mapping alignments. Various algorithms were used for gene expression values using the normalization techniques. For pseudoalignment, 3 commonly used pseudoaligners, Kallisto (60), Sailfish (61), and Salmon (62) were used. Differential expression analyses were carried out to study the gene expressions. The methods used were divided into 3 categories as follows (46): (1) Methods that assume a negative binomial distribution of data: BaySeq (63), Cuffdiff (64), DESeq2 (65), EBseq (66), and edgeR (67); (2) methods that assume a log-normal distribution as Ballgown (68) and the Trend and Voom limma (69) variants; and (3) non-parametric methods such as NOISeq (70) and SAMseq (71) (R package).

The criteria for DEGs were adjusted with P-value (adjP) $\leq$ 0.1 and fold change (FC) values $\geq$ 1.5 and $\leq$ -1.5 for the upregulated and downregulated genes. The analysis was carried out by the principal component analysis (PCA) plot, heatmap of the sample-to-sample distance matrix (with clustering) based on the normalized counts, dispersion estimates, histogram of P-values for the genes, and MA plot. Principal component analysis tools, such as https://gccri.bishop-lab.uthscsa.edu/correlation-analyzer/ (72), were used to carry out the analysis (Figure 1).

The GO and KEGG analyses for the different species were performed by various computational tools, such as GEne SeT AnaLysis Toolkit (Webgestalt) web server (73), the "clusterProfiler" package (74), and "msigdbr" package (75). Differential expression genes for each species were selected separately for the differentially connected genes, and gene modules and P-values were obtained by hypergeometric analysis. The functional analyses, including GO, pathway, and chromosome enrichment analysis of the candidate genes, can also performed by DAVID (76), a high-throughput and integrated data-mining environment. The pathway maps represent proteins and genes, which are accountable for the reaction networks. The molecular interactions were manually drawn from the KEGG database. KEGG pathways were used to obtain significantly enriched terms with a P-value $<$ 0.05. Significant terms were taken for adjP $\leq$ 0.1. The corrected P-values were used for the false discovery rate (FDR). Weighted gene co-expression network (WGCNA) (77) provides the modules or clusters of highly correlated genes.

The Drug Gene Interaction Database (DGIdb) (78, 79) was used to predict the interaction between DEGs and drugs so that the oncologist can make treatment decisions regarding the therapeutic targets with the approved drugs and immunotherapies.

Various tools and computational methodologies are used for gene screening and prognostic signature development. The LASSO (80) Cox regression was used to obtain the prognosis-related TRGs using the "glmnet" package, and the gene signature was identified based on the LASSO Cox regression coefficients. Multivariable analysis was carried out to select the suitable reason with the range of prognostic factors. Combining the gene signature and clinicopathological parameters, a nomogram was built to predict the survival probability

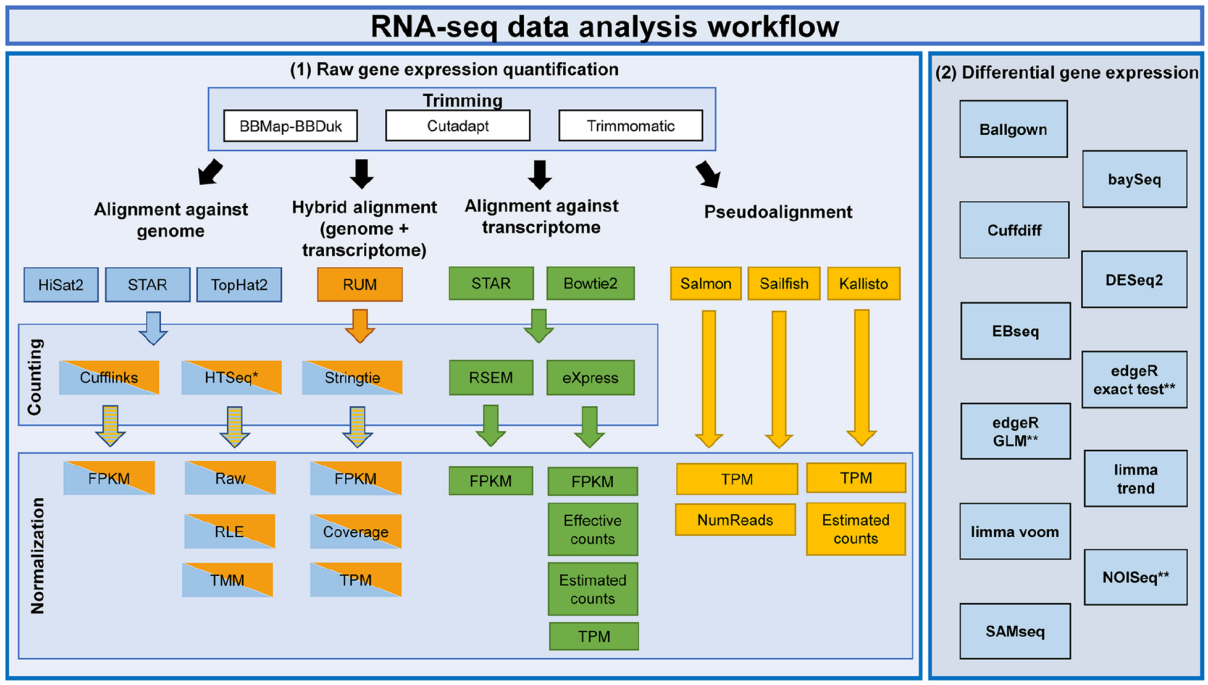

Figure 1. The workflow for RNA sequencing (RNA-Seq) analysis. Left panel (1) represents the raw gene expression quantification workflow. Each box contains algorithms and methods for trimming, alignment, counting, normalization, and pseudoalignment levels for RNA-Seq. The right panel (2) represents differential gene expression quantification algorithms (*HTSeq is performed in union and intersection-strict modes; **EdgeR exact test, edgeR GLM, and NOISeq have estimated internally 3 separated normalization techniques (46)).

of different time points using the "rms" package (81). The internal validation of the risk score was carried out using the bootstrap resampling method. The calibrated nomogram was accessed by the calibration curves (82). The performance of the gene signature was estimated using X-tile software (82). Multivariate analysis was carried out to estimate the overall survival (OS). The nonparametric tests such as *t*-test, Mann-Whitney-Wilcoxon test, and Kruskal-Wallis test are used for statistical analysis. These methods do not require a distribution to meet the needed assumptions for analysis, especially if the data is not normally distributed (48).

3. Transcriptomics in Clinical Trials

Transcriptomics applications include classifications of cancers, identification of early detection of cancer biomarkers, formation of cancer prognostic and predictive panels, intratumoral heterogeneity (ITH) and tumor microenvironment (TME) related research, and RNA-based therapeutics. The WINTHER trial was the first such study, which included integrated transcriptional-genomics studies for the treatment of solid tumors with matched therapeutics results (83). The WINTHER database is a valuable resource that provides information on the "targeting" gene expression and efficient registered and clinically tested drugs. The treatment process was recommended by the clinical management committee, and evaluation was carried out by exploratory matching score, calculated in an eventually blinded manner. An Eastern Cooperative Oncology Group (ECOG) performance with 2 or fewer previous therapies showed a status of 0. The high matching score was independently related to the long progression-free survival. Though the WINTHER trial showed that transcriptomic analysis could be successfully used for the treatment in selected patients, the analysis was complex and required bioinformatics expertise. Transcriptomic analysis based on matched therapies to molecular alterations was also carried out in the clinical trials on pediatric patients (83, 84). In one such study, 20 patients with refractory pediatric sarcoma were selected to study the gene expression profiling and identification of overexpressed genes and deregulated pathways that could be therapeutically targeted (83). The targeted therapy was received by 9 patients with higher overall ($P = 0.0014$) and progression-free ($P = 0.0011$) survival rates. Recently, updated results of the INFORM study

with 1300 patients (enrolled at 72 centers) showed that 525 included in the analysis reported "very high" and "high" priority actionable targets for 8% and 14.8% of patients, respectively (85). Another Pediatric Sequencing (PIPseq) Program studied the clinical utility of molecular analysis in pediatric tumors (86). Molecular alterations (variant calls, copy number variations [CNV], fusions, and overexpressed genes) and clinically relevant alterations were initially reviewed.

4. Challenges and Limitations

The main challenges in transcriptomics are the handling of tissue samples and the application of advanced computational methodologies. The experimental data of RNA-Seq formalin-fixed, paraffin-embedded tissue samples include degraded, fragmented (87), or defoiled (88) RNA, which was followed by bioinformatics. Degradation causes errors during sample preparation, leading to the presence of sequence data from a different sample. In addition, tumor samples can be contaminated by normal cells that surround the tumor. Experimental methods (cell sorting or laser capture micro-dissection) (87) and bioinformatics algorithms are used to remove the contamination effects (88). Furthermore, methodological artifacts are endured in transcriptome analysis that requires precise evaluation. Another challenge is the application of advanced computational methodologies. The complex analysis of transcriptomic data requires a high-level bioinformatic approach, needing orderly expansion of efficient algorithms. The implementation of transcriptomic analysis in clinical functions can be more complex compared to genomics analysis. The reproducibility issues also need to be addressed. As the RNA profile is used to compare tumor tissue with normal tissue from the same organ, in some cases, the peripheral blood or buccal swab samples are compared, due to which there are variations in the results. Studies have shown that transcriptomics in clinical trials develops diverse and complex algorithms to specify the functionality of molecular alterations (84).

5. Conclusions

Novel next-generation sequencing (NGS) techniques have improved our understanding of the tumor biology of individual patients and provided an abundance of translational discovery opportunities, as tumor genomic profiling approaches provide average signatures and a snapshot of the tumor state (biopsy) but do not frequently show the complete tumor biology, all tumor parts, or the inherent heterogenic individual cell populations. However, understanding the tumor complexity and heterogeneity, as well as the dynamic expression of the genome, requires advanced computational methodologies. The future of precision medicine lies in integrated multiomics data analysis for better treatment decisions. Transcriptomics analysis has identified appropriate targeted therapies and validated some results in clinical trials, predicting that transcriptomics will increase the targeted therapy and can lead to favorable outcomes. The future prospects of transcriptomic data are promising but remain limited to the cost, time implications, and inconsistent analysis. Clinical trials are still going on to explore the relevance of transcriptomics analysis in cancer biology.

Footnotes

Authors' Contribution: Ruby Srivastava is the only author of the article, and the study was solely carried out by the author.

Conflict of Interests: There is no conflict of interest for this manuscript.

Funding/Support: DST WOSA (SR/WOS-A/CS-69/2018).

References

1. You JS, Jones PA. Cancer genetics and epigenetics: Two sides of the same coin? *Cancer Cell*. 2012;**22**(1):9–20. [PubMed ID: 22789535]. [PubMed Central ID: PMC3396881]. https://doi.org/10.1016/j.ccr.2012.06.008.
2. Takeshima H, Ushijima T. Accumulation of genetic and epigenetic alterations in normal cells and cancer risk. *NPJ Precis Oncol*. 2019;**3**:7. [PubMed ID: 30854468]. [PubMed Central ID: PMC6403339]. https://doi.org/10.1038/s41698-019-0079-0.
3. Buzdin A, Sorokin M, Garazha A, Glusker A, Aleshin A, Poddubskaya E, et al. RNA sequencing for research and diagnostics in clinical oncology. *Semin Cancer Biol*. 2020;**60**:311–23. [PubMed ID: 31412295]. https://doi.org/10.1016/j.semcancer.2019.07.010.
4. Nussinov R, Jang H, Tsai CJ, Cheng F. Review: Precision medicine and driver mutations: Computational methods, functional assays and conformational principles for interpreting cancer drivers. *PLoS Comput Biol*. 2019;**15**(3). e1006658. [PubMed ID: 30921324]. [PubMed Central ID: PMC6438456]. https://doi.org/10.1371/journal.pcbi.1006658.
5. Thomas E, Mohammed S. Advances in Genetic Testing for Hereditary Cancer Syndromes. *Recent Results Cancer Res*. 2016;**205**:1–15. [PubMed ID: 27075345]. https://doi.org/10.1007/978-3-319-29998-3_1.
6. Esplin ED, Oei L, Snyder MP. Personalized sequencing and the future of medicine: discovery, diagnosis and defeat of disease. *Pharmacogenomics*. 2014;**15**(14):1771–90. [PubMed ID: 25493570]. [PubMed Central ID: PMC4336568]. https://doi.org/10.2217/pgs.14.117.
7. Jin Z, Liu Y. DNA methylation in human diseases. *Genes Dis*. 2018;**5**(1):1–8. [PubMed ID: 30258928]. [PubMed Central ID: PMC6147084]. https://doi.org/10.1016/j.gendis.2018.01.002.

8. Smith NC, Matthews JM. Mechanisms of DNA-binding specificity and functional gene regulation by transcription factors. *Curr Opin Struct Biol*. 2016;**38**:68–74. [PubMed ID: 27295424]. https://doi.org/10.1016/j.sbi.2016.05.006.
9. Dana H, Chalbatani GM, Mahmoodzadeh H, Karimloo R, Rezaiean O, Moradzadeh A, et al. Molecular Mechanisms and Biological Functions of siRNA. *Int J Biomed Sci*. 2017;**13**(2):48–57. [PubMed ID: 28824341]. [PubMed Central ID: PMC5542916].
10. Bedard PL, Hansen AR, Ratain MJ, Siu LL. Tumour heterogeneity in the clinic. *Nature*. 2013;**501**(7467):355–64. [PubMed ID: 24048068]. [PubMed Central ID: PMC5224525]. https://doi.org/10.1038/nature12627.
11. Gottlieb B, Babrzadeh F, Oros KK, Alvarado C, Wang C, Gharizadeh B, et al. New insights into the role of intra-tumor genetic heterogeneity in carcinogenesis: identification of complex single gene variance within tumors. *J Cancer Metastatis Treat*. 2018;**4**(7):37. https://doi.org/10.20517/2394-4722.2018.26.
12. Lee JK, Choi YL, Kwon M, Park PJ. Mechanisms and Consequences of Cancer Genome Instability: Lessons from Genome Sequencing Studies. *Annu Rev Pathol*. 2016;**11**:283–312. [PubMed ID: 26907526]. https://doi.org/10.1146/annurev-pathol-012615-044446.
13. Engin HB, Kreisberg JF, Carter H. Structure-Based Analysis Reveals Cancer Missense Mutations Target Protein Interaction Interfaces. *PLoS One*. 2016;**11**(4). e0152929. [PubMed ID: 27043210]. [PubMed Central ID: PMC4820104]. https://doi.org/10.1371/journal.pone.0152929.
14. Vuong H, Cheng F, Lin CC, Zhao Z. Functional consequences of somatic mutations in cancer using protein pocket-based prioritization approach. *Genome Med*. 2014;**6**(10):81. [PubMed ID: 25360158]. [PubMed Central ID: PMC4213513]. https://doi.org/10.1186/s13073-014-0081-7.
15. Jimenez CR, Zhang H, Kinsinger CR, Nice EC. The cancer proteomic landscape and the HUPO Cancer Proteome Project. *Clin Proteomics*. 2018;**15**:4. [PubMed ID: 29416445]. [PubMed Central ID: PMC5785860]. https://doi.org/10.1186/s12014-018-9180-6.
16. Kiernan UA. Biomarker rediscovery in diagnostics. *Expert Opin Med Diagn*. 2008;**2**(12):1391–400. [PubMed ID: 23496785]. https://doi.org/10.1517/17530050802566488.
17. Bhawe KM, Aghi MK. Microarray Analysis in Glioblastomas. *Methods Mol Biol*. 2016;**1375**:195–206. [PubMed ID: 26113463]. [PubMed Central ID: PMC5056625]. https://doi.org/10.1007/7651_2015_245.
18. Cech TR, Steitz JA. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell*. 2014;**157**(1):77–94. [PubMed ID: 24679528]. https://doi.org/10.1016/j.cell.2014.03.008.
19. Taft RJ, Pang KC, Mercer TR, Dinger M, Mattick JS. Non-coding RNAs: regulators of disease. *J Pathol*. 2010;**220**(2):126–39. [PubMed ID: 19882673]. https://doi.org/10.1002/path.2638.
20. Maher CA, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X, et al. Transcriptome sequencing to detect gene fusions in cancer. *Nature*. 2009;**458**(7234):97–101. [PubMed ID: 19136943]. [PubMed Central ID: PMC2725402]. https://doi.org/10.1038/nature07638.
21. Maher CA, Palanisamy N, Brenner JC, Cao X, Kalyana-Sundaram S, Luo S, et al. Chimeric transcript discovery by paired-end transcriptome sequencing. *Proc Natl Acad Sci U S A*. 2009;**106**(30):12353–8. [PubMed ID: 19592507]. [PubMed Central ID: PMC2708976]. https://doi.org/10.1073/pnas.0904720106.
22. Stransky N, Cerami E, Schalm S, Kim JL, Lengauer C. The landscape of kinase fusions in cancer. *Nat Commun*. 2014;**5**:4846. [PubMed ID: 25204415]. [PubMed Central ID: PMC4175590]. https://doi.org/10.1038/ncomms5846.
23. Klijn C, Durinck S, Stawiski EW, Haverty PM, Jiang Z, Liu H, et al. A comprehensive transcriptional portrait of human cancer cell lines. *Nat Biotechnol*. 2015;**33**(3):306–12. [PubMed ID: 25485619]. https://doi.org/10.1038/nbt.3080.
24. Shaw AT, Hsu PP, Awad MM, Engelman JA. Tyrosine kinase gene rearrangements in epithelial malignancies. *Nat Rev Cancer*. 2013;**13**(11):772–87. [PubMed ID: 24132104]. [PubMed Central ID: PMC3902129]. https://doi.org/10.1038/nrc3612.
25. Roberts KG, Morin RD, Zhang J, Hirst M, Zhao Y, Su X, et al. Genetic alterations activating kinase and cytokine receptor signaling in high-risk acute lymphoblastic leukemia. *Cancer Cell*. 2012;**22**(2):153–66. [PubMed ID: 22897847]. [PubMed Central ID: PMC3422513]. https://doi.org/10.1016/j.ccr.2012.06.005.
26. Roberts KG, Li Y, Payne-Turner D, Harvey RC, Yang YL, Pei D, et al. Targetable kinase-activating lesions in Ph-like acute lymphoblastic leukemia. *N Engl J Med*. 2014;**371**(11):1005–15. [PubMed ID: 25207766]. [PubMed Central ID: PMC4191900]. https://doi.org/10.1056/NEJMoa1403088.
27. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 2000;**403**(6769):503–11. [PubMed ID: 10676951]. https://doi.org/10.1038/35000501.
28. Barton S, Hawkes EA, Wotherspoon A, Cunningham D. Are we ready to stratify treatment for diffuse large B-cell lymphoma using molecular hallmarks? *Oncologist*. 2012;**17**(12):1562–73. [PubMed ID: 23086691]. [PubMed Central ID: PMC3528389]. https://doi.org/10.1634/theoncologist.2012-0218.
29. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature*. 2000;**406**(6797):747–52. [PubMed ID: 10963602]. https://doi.org/10.1038/35021093.
30. Calin GA, Croce CM. MicroRNA signatures in human cancers. *Nat Rev Cancer*. 2006;**6**(11):857–66. [PubMed ID: 17060945]. https://doi.org/10.1038/nrc1997.
31. Pritchard CC, Cheng HH, Tewari M. MicroRNA profiling: approaches and considerations. *Nat Rev Genet*. 2012;**13**(5):358–69. [PubMed ID: 22510765]. [PubMed Central ID: PMC4517822]. https://doi.org/10.1038/nrg3198.
32. Dykxhoorn DM, Lieberman J. The silent revolution: RNA interference as basic biology, research tool, and therapeutic. *Annu Rev Med*. 2005;**56**:401–23. [PubMed ID: 15660519]. https://doi.org/10.1146/annurev.med.56.082103.104606.
33. Quinodoz S, Guttman M. Long noncoding RNAs: an emerging link between gene regulation and nuclear organization. *Trends Cell Biol*. 2014;**24**(11):651–63. [PubMed ID: 25441720]. [PubMed Central ID: PMC4254690]. https://doi.org/10.1016/j.tcb.2014.08.009.
34. Berindan-Neagoe I, Monroig Pdel C, Pasculli B, Calin GA. MicroRNAome genome: a treasure for cancer diagnosis and therapy. *CA Cancer J Clin*. 2014;**64**(5):311–36. [PubMed ID: 25104502]. [PubMed Central ID: PMC4461198]. https://doi.org/10.3322/caac.21244.
35. Croce CM. Causes and consequences of microRNA dysregulation in cancer. *Nat Rev Genet*. 2009;**10**(10):704–14. [PubMed ID: 19763153]. [PubMed Central ID: PMC3467096]. https://doi.org/10.1038/nrg2634.
36. Manjunath N, Dykxhoorn DM. Advances in synthetic siRNA delivery. *Discov Med*. 2010;**9**(48):418–30. [PubMed ID: 20515610].
37. Chu C, Spitale RC, Chang HY. Technologies to probe functions and mechanisms of long noncoding RNAs. *Nat Struct Mol Biol*. 2015;**22**(1):29–35. [PubMed ID: 25565030]. https://doi.org/10.1038/nsmb.2921.
38. Jeck WR, Sharpless NE. Detecting and characterizing circular RNAs. *Nat Biotechnol*. 2014;**32**(5):453–61. [PubMed ID: 24811520]. [PubMed Central ID: PMC4121655]. https://doi.org/10.1038/nbt.2890.
39. Zhou WY, Cai ZR, Liu J, Wang DS, Ju HQ, Xu RH. Circular RNA: metabolism, functions and interactions with proteins. *Mol Cancer*. 2020;**19**(1):172. [PubMed ID: 33317550]. [PubMed Central ID: PMC7734744]. https://doi.org/10.1186/s12943-020-01286-3.
40. *Gene Expression Omnibus*. 2023. Available from: https://www.ncbi.nlm.nih.gov/geo/.
41. *European Genome-phenome Archive*. 2023. Available from: https://ega-archive.org/.
42. Ovaska K, Laakso M, Haapa-Paananen S, Louhimo R, Chen P,

Aittomaki V, et al. Large-scale data integration framework provides a comprehensive view on glioblastoma multiforme. *Genome Med*. 2010;**2**(9):65. [PubMed ID: 20822536]. [PubMed Central ID: PMC3092116]. https://doi.org/10.1186/gm186.
43. Goecks J, Nekrutenko A, Taylor J, Galaxy T. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol*. 2010;**11**(8):R86. [PubMed ID: 20738864]. [PubMed Central ID: PMC2945788]. https://doi.org/10.1186/gb-2010-11-8-r86.
44. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;**30**(15):2114–20. [PubMed ID: 24695404]. [PubMed Central ID: PMC4103590]. https://doi.org/10.1093/bioinformatics/btu170.
45. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 2011;**17**(1):10. https://doi.org/10.14806/ej.17.1.200.
46. Corchete LA, Rojas EA, Alonso-Lopez D, De Las Rivas J, Gutierrez NC, Burguillo FJ. Systematic comparison and assessment of RNA-seq procedures for gene expression quantitative analysis. *Sci Rep*. 2020;**10**(1):19737. [PubMed ID: 33184454]. [PubMed Central ID: PMC7665074]. https://doi.org/10.1038/s41598-020-76881-x.
47. Dinno A. *Dunn.test, Dunn's Test of Multiple Comparisons Using Rank Sums*. 2017. Available from: https://cran.r-project.org/web/packages/dunn.test/index.html.
48. R Core Team. *R, A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2019. Available from: https://www.r-project.org/.
49. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013;**14**(4):R36. [PubMed ID: 23618408]. [PubMed Central ID: PMC4053844]. https://doi.org/10.1186/gb-2013-14-4-r36.
50. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;**29**(1):15–21. [PubMed ID: 23104886]. [PubMed Central ID: PMC3530905]. https://doi.org/10.1093/bioinformatics/bts635.
51. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;**12**(4):357–60. [PubMed ID: 25751142]. [PubMed Central ID: PMC4655817]. https://doi.org/10.1038/nmeth.3317.
52. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;**9**(4):357–9. [PubMed ID: 22388286]. [PubMed Central ID: PMC3322381]. https://doi.org/10.1038/nmeth.1923.
53. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;**25**(16):2078–9. [PubMed ID: 19505943]. [PubMed Central ID: PMC2723002]. https://doi.org/10.1093/bioinformatics/btp352.
54. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;**29**(1):24–6. [PubMed ID: 21221095]. [PubMed Central ID: PMC3346182]. https://doi.org/10.1038/nbt.1754.
55. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;**28**(5):511–5. [PubMed ID: 20436464]. [PubMed Central ID: PMC3146043]. https://doi.org/10.1038/nbt.1621.
56. Roberts A, Pachter L. Streaming fragment assignment for real-time analysis of sequencing experiments. *Nat Methods*. 2013;**10**(1):71–3. [PubMed ID: 23160280]. [PubMed Central ID: PMC3880119]. https://doi.org/10.1038/nmeth.2251.
57. Anders S, Pyl PT, Huber W. HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;**31**(2):166–9. [PubMed ID: 25260700]. [PubMed Central ID: PMC4287950]. https://doi.org/10.1093/bioinformatics/btu638.
58. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011;**12**:323. [PubMed ID: 21816040]. [PubMed Central ID: PMC3163565]. https://doi.org/10.1186/1471-2105-12-323.
59. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015;**33**(3):290–5. [PubMed ID: 25690850]. [PubMed Central ID: PMC4643835]. https://doi.org/10.1038/nbt.3122.
60. Bray NL, Pimentel H, Melsted P, Pachter L. Erratum: Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. 2016;**34**(8):888. [PubMed ID: 27504780]. https://doi.org/10.1038/nbt0816-888d.
61. Patro R, Mount SM, Kingsford C. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nat Biotechnol*. 2014;**32**(5):462–4. [PubMed ID: 24752080]. [PubMed Central ID: PMC4077321]. https://doi.org/10.1038/nbt.2862.
62. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017;**14**(4):417–9. [PubMed ID: 28263959]. [PubMed Central ID: PMC5600148]. https://doi.org/10.1038/nmeth.4197.
63. Hardcastle TJ, Kelly KA. baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics*. 2010;**11**:422. [PubMed ID: 20698981]. [PubMed Central ID: PMC2928208]. https://doi.org/10.1186/1471-2105-11-422.
64. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol*. 2013;**31**(1):46–53. [PubMed ID: 23222703]. [PubMed Central ID: PMC3869392]. https://doi.org/10.1038/nbt.2450.
65. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;**15**(12):550. [PubMed ID: 25516281]. [PubMed Central ID: PMC4302049]. https://doi.org/10.1186/s13059-014-0550-8.
66. Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman AI, Smits BM, et al. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics*. 2013;**29**(8):1035–43. [PubMed ID: 23428641]. [PubMed Central ID: PMC3624807]. https://doi.org/10.1093/bioinformatics/btt087.
67. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;**26**(1):139–40. [PubMed ID: 19910308]. [PubMed Central ID: PMC2796818]. https://doi.org/10.1093/bioinformatics/btp616.
68. Frazee AC, Pertea G, Jaffe AE, Langmead B, Salzberg SL, Leek JT. Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nature Biotechnology*. 2015;**33**(3):243–6. https://doi.org/10.1038/nbt.3172.
69. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;**43**(7). e47. [PubMed ID: 25605792]. [PubMed Central ID: PMC4402510]. https://doi.org/10.1093/nar/gkv007.
70. Tarazona S, Garcia-Alcalde F, Dopazo J, Ferrer A, Conesa A. Differential expression in RNA-seq: a matter of depth. *Genome Res*. 2011;**21**(12):2213–23. [PubMed ID: 21903743]. [PubMed Central ID: PMC3227109]. https://doi.org/10.1101/gr.124321.111.
71. Li J, Tibshirani R. Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. *Stat Methods Med Res*. 2013;**22**(5):519–36. [PubMed ID: 22127579]. [PubMed Central ID: PMC4605138]. https://doi.org/10.1177/0962280211428386.
72. Miller HE, Bishop AJR. Correlation AnalyzeR: functional predictions from gene co-expression correlations. *BMC Bioinformatics*. 2021;**22**(1):206. [PubMed ID: 33879054]. [PubMed Central ID: PMC8056587]. https://doi.org/10.1186/s12859-021-04130-7.
73. Zhang B, Kirov S, Snoddy J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res*. 2005;**33**(Web Server issue):W741–8. [PubMed ID: 15980575]. [PubMed

Central ID: PMC1160236]. https://doi.org/10.1093/nar/gki475.
74. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)*. 2021;**2**(3):100141. [PubMed ID: 34557778]. [PubMed Central ID: PMC8454663]. https://doi.org/10.1016/j.xinn.2021.100141.
75. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;**102**(43):15545–50. [PubMed ID: 16199517]. [PubMed Central ID: PMC1239896]. https://doi.org/10.1073/pnas.0506580102.
76. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;**4**(1):44–57. [PubMed ID: 19131956]. https://doi.org/10.1038/nprot.2008.211.
77. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;**9**:559. [PubMed ID: 19114008]. [PubMed Central ID: PMC2631488]. https://doi.org/10.1186/1471-2105-9-559.
78. Freshour SL, Kiwala S, Cotto KC, Coffman AC, McMichael JF, Song JJ, et al. Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. *Nucleic Acids Res*. 2021;**49**(D1):D1144–51. [PubMed ID: 33237278]. [PubMed Central ID: PMC7778926]. https://doi.org/10.1093/nar/gkaa1084.
79. Griffith M, Griffith OL, Coffman AC, Weible JV, McMichael JF, Spies NC, et al. DGIdb: mining the druggable genome. *Nat Methods*. 2013;**10**(12):1209–10. [PubMed ID: 24122041]. [PubMed Central ID: PMC3851581]. https://doi.org/10.1038/nmeth.2689.
80. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw*. 2010;**33**(1):1–22. [PubMed ID: 20808728]. [PubMed Central ID: PMC2929880].
81. Harrell FE. *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. **608**. Springer; 2001.
82. Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clin Cancer Res*. 2004;**10**(21):7252–9. [PubMed ID: 15534099]. https://doi.org/10.1158/1078-0432.CCR-04-0713.
83. Weidenbusch B, Richter GHS, Kesper MS, Guggemoos M, Gall K, Prexler C, et al. Transcriptome based individualized therapy of refractory pediatric sarcomas: feasibility, tolerability and efficacy. *Oncotarget*. 2018;**9**(29):20747–60. [PubMed ID: 29755686]. [PubMed Central ID: PMC5945512]. https://doi.org/10.18632/oncotarget.25087.
84. Oberg JA, Glade Bender JL, Sulis ML, Pendrick D, Sireci AN, Hsiao SJ, et al. Implementation of next generation sequencing into pediatric hematology-oncology practice: moving beyond actionable alterations. *Genome Med*. 2016;**8**(1):133. [PubMed ID: 28007021]. [PubMed Central ID: PMC5180407]. https://doi.org/10.1186/s13073-016-0389-6.
85. van Tilburg CM, Pfaff E, Pajtler KW, Langenberg KP, Fiesel P, Jones BC, et al. The pediatric precision oncology study INFORM: Clinical outcome and benefit for molecular subgroups. *J Clin Oncol*. 2020;**38**(18 suppl):LBA10503. https://doi.org/10.1200/JCO.2020.38.18_suppl.LBA10503.
86. Nieuwenhuis TO, Yang SY, Verma RX, Pillalamarri V, Arking DE, Rosenberg AZ, et al. Consistent RNA sequencing contamination in GTEx and other data sets. *Nat Commun*. 2020;**11**(1):1933. [PubMed ID: 32321923]. [PubMed Central ID: PMC7176728]. https://doi.org/10.1038/s41467-020-15821-9.
87. Ong CJ, Tan QX, Lim HJ, Shannon NB, Lim WK, Hendrikson J, et al. An Optimised Protocol Harnessing Laser Capture Microdissection for Transcriptomic Analysis on Matched Primary and Metastatic Colorectal Tumours. *Sci Rep*. 2020;**10**(1):682. [PubMed ID: 31959771]. [PubMed Central ID: PMC6971024]. https://doi.org/10.1038/s41598-019-55146-2.
88. Shen Q, Hu J, Jiang N, Hu X, Luo Z, Zhang H. contamDE: differential expression analysis of RNA-seq data for contaminated tumor samples. *Bioinformatics*. 2016;**32**(5):705–12. [PubMed ID: 26556386]. https://doi.org/10.1093/bioinformatics/btv657.