Gene Expression Analysis in Cancer Cells Using RNA-Seq Technologies: A Comprehensive Review

Hamid Mohsin Ghyadh[1], **Jinan RIYADH ABBAS Al-Dujaili**[2,3], **Esra jasim Bari**[4], **Farooq L. J. Almohaisen**[5]

Abstract

Cancer is a dynamic and heterogeneous disease arising from complex molecular alterations at multiple levels of regulation. To understand these changes, there is a need for high-resolution transcriptome methods that are able to capture global gene expression variations and their functional effects. RNA sequencing (RNA-Seq) is a game-changing next generation sequencing technology for unbiased transcriptome profiling of cancer. This review describes the applications of RNA-Seq technology in cancer gene expression research and covers the basic principles of this methodology, the analysis process, and the biological and clinical implications. RNA-Seq yields information on differentially expressed genes, alternative splicing, gene fusion and non-coding RNA transcripts that promote tumor growth. Single-cell RNA sequencing has allowed for better characterization of tumor heterogeneity and microenvironment at the cellular level. Functional enrichment and pathway analyses provide a systems biology interpretation of transcriptome alterations, and identify deregulated signaling pathways and molecular networks associated with oncogenesis. RNA-Seq enables precision oncology through biomarker discovery, subtyping of cancers and therapy response prediction. Integration of RNA-Seq and multi-omics data has also contributed to better patient classification and molecular signature identification. However, the complexity of data, technological variability and the consistency of the analytical pipeline hinder its wide clinical deployment. RNA-Seq is a cornerstone of cancer research, providing deep insights into tumor biology and enabling personalized treatment. Advances in sequencing, computational and integrative multi-omics techniques are expected to bring about improved cancer diagnosis, prognosis and therapy.

[1]Department of Medical Laboratory Techniques, Shatra Technical College, Southern Technical University, Iraq, Email Hamid.ghyadh@stu.edu.iq .
[2]College of Agriculture, Tikrit University, Tikrit, Iraq
[3]Ministry Of Agriculture ,Mesopotamia State Company for Seeds, Barley section, Assistant Chief Agricultural Engineer, Iraq, Email: jananriad@st.tu.edu.iq
[4]Department of Medical Laboratory Techniques Shatra Technical College Southern Technical University, Email. asra99014@gnail.com
[5]Department of medical Laboratories techniques, Southern Technical University, Basra , Iraq
Email: farooq.almohaisen@stu.edu.iq

Keywords: Gene Expression, Cancer Cells, RNA-Seq Technologies

Introduction

Cancer is one of the major health problem in the world causing morbidity, mortality and socioeconomic impact in both developed and developing countries . While there have been advances in early diagnosis, targeted drugs, immunotherapy and personalized medicine, cancer still kills millions globally and strains the healthcare infrastructure. Cancer kills about 10 million people globally. The most common cancers are breast, lung, prostate, colorectal and stomach cancers. The global burden of cancer is expected to increase in the coming decades due to factors such as population growth, increased life expectancy, urbanization, and the spread of unhealthy lifestyle behaviors, including tobacco use, physical inactivity, poor diets, and excessive alcohol consumption (Sung et al., 2021).

Carcinogenesis is multifactorial involving lifestyle, genetic predisposition, chronic inflammation, viral infections, environmental pollutants, occupational exposures and hormonal imbalances. The stepwise acquisition of genetic and epigenetic alterations that perturb cellular homeostasis drives cancer development . Alterations disrupt important control processes for cell proliferation, cell death, DNA repair, blood vessel formation, immune evasion and metastasis of cancer cells (Hanahan, 2022). Such molecular alterations are complicated and may cause tumor heterogeneity in the same histological type, which is challenging for cancer detection and treatment. Tumor heterogeneity impacts disease progression, therapy resistance, recurrence and prognosis (Dagogo-Jack & Shaw, 2018).

Tumor cells interact with and are affected by the surrounding immune cells, fibroblasts, stromal elements and signaling molecules, collectively known as the tumor microenvironment . This interaction is a significant factor that contributes to cancer growth and response to treatment. This dynamic ecology favors tumor survival, immune evasion and metastatic propagation and deserves molecular characterization besides histological examination (Hinshaw and Shevde, 2019). So, in the current oncology, molecular profiling is more and more used for the cancer biology studies and the discovery of useful biomarkers for diagnosis, prognosis and treatment choice.

Gene expression analysis is one of the most powerful approaches to study the cancer etiology at the transcriptome level . Investigators are able to identify cancer phenotype-specific dysregulated pathways and molecular markers by analyzing mRNA, non-coding RNAs and signaling pathways. These studies have been critical for the development of precision oncology, which is based on the genomic features of tumors instead of the morphological classification for treatment decisions (Hoadley et al., 2018). RNA sequencing (RNA-Seq) is a disruptive technology that offers a global, high-resolution snapshot of gene expression and opens new avenues for cancer research and clinical translation .

Gene expression analysis has been extensively used in the cancer research field to explore molecular pathways involved in tumor initiation, growth, metastasis and therapy resistance. Unlike conventional morphology-based histopathological methods, gene expression profiling enables the study of dynamic molecular alterations occurring in cancer cells and their microenvironment. Gene transcriptional activity in tumor tissues can be compared with normal cells to reveal oncogenesis pathways (Alizadeh et al., 2015).

Cancer is a heterogeneous group of diseases with different molecular fingerprints even among patients with the same histological type of cancer . Heterogeneity is a major determinant of prognosis, response to therapy and clinical outcome. Molecular subgroups of cancer that are not identifiable by conventional diagnostic methods are identified by gene expression analysis. Molecular subtyping based on transcriptional profile has improved the treatment of breast, leukemia and lung cancer (Perou et al., 2000; Hoadley et al., 2018) with better prognosis and treatment.

Emergence of RNA-Seq Technology

In the last 20 years high-throughput sequencing has revolutionized studies of gene expression. Before the advent of RNA sequencing (RNA-Seq), microarray technology was the dominant method for transcriptome studies. Microarrays were a breakthrough in molecular biology, providing the ability to measure thousands of genes at the same time, but were limited by dependence on probes, background noise, low sensitivity for rare transcripts and inability to detect novel transcripts or unknown splice variants (Wang et al., 2009). These limitations limited the depth and accuracy of transcriptomic analysis in complex diseases such as cancer, where small genetic changes can have profound biological and clinical consequences.

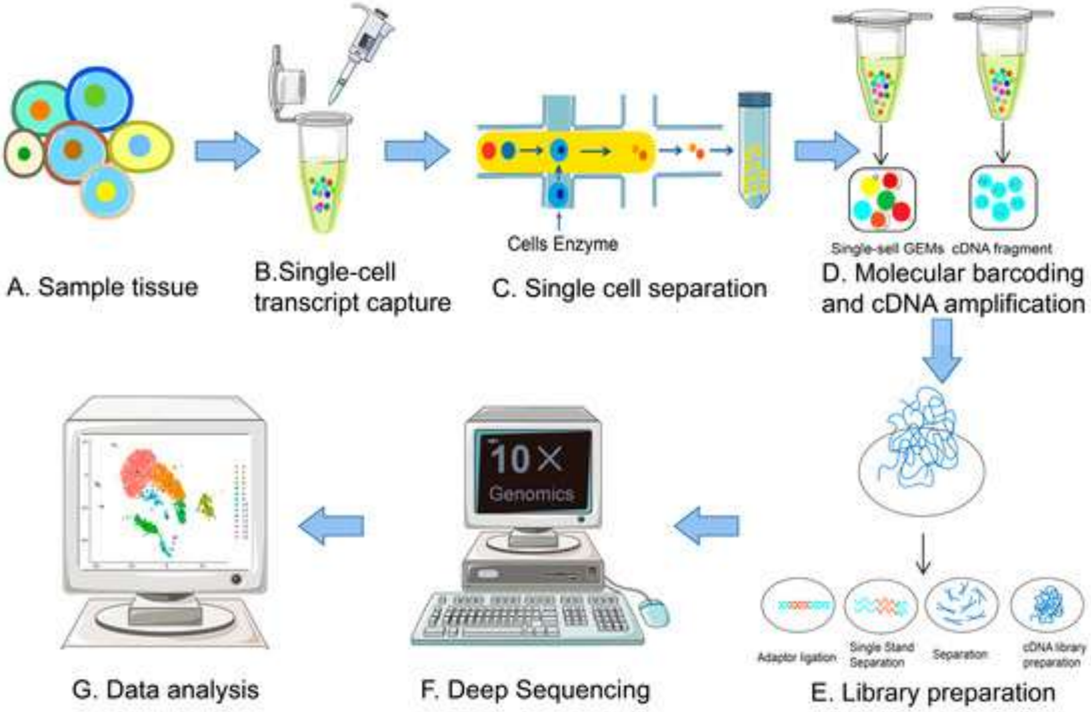
Next-generation sequencing (NGS) and RNA - Seq , a more complete and accurate method of transcriptome analysis , have revolutionized the field of transcriptomics . RNA-Seq is a high-throughput sequencing technology for collecting and quantifying RNA molecules from a biological sample. RNA-Seq, unlike microarrays, does not require predefined probes to detect known and novel transcripts, including mRNA, lncRNA, miRNA, circRNA, and fusion transcripts (Ozsolak & Milos, 2011). This capability has greatly enhanced our understanding of the complexity of the transcriptome in cancer biology.

RNA-Seq is highly sensitive and has a broad dynamic range, allowing it to accurately quantify genes with high expression levels as well as genes with low expression levels. Rare transcripts, alternative splicing events, small changes in expression might be biomarkers or therapeutic targets for cancer. RNA-Seq at the single-base resolution is a major breakthrough which enables researchers to identify SNVs, RNA editing events, allele-specific expression and transcript isoforms better than the previous technologies (Stark et al.,., 2019).

RNA-Seq has had a major contribution to cancer research with the refinement of tumor molecular classification, understanding the mechanisms of drug resistance and characterization of tumor microenvironment interactions . Also, single-cell transcriptomes can be analyzed instead of bulk tissue samples, thus providing a higher resolution by single-cell RNA sequencing (scRNA-Seq). Tumor heterogeneity, cancer stem cell, immunological infiltration and metastatic progression have been better understood (Tang et al., 2019).

The falling cost of sequencing and increasing bioinformatics capabilities are taking RNA-Seq from research to clinical cancer . It improves diagnosis, prognosis and choice of treatment in precision medicine schemes. RNA-Seq is currently one of the most powerful tools in cancer genomics and transcriptomics.

Figure(1) An overview of the RNA-sequencing procedures.



Significance of RNA-Seq in Cancer Research

RNA sequencing (RNA-Seq) is a major tool in cancer research because it enables precise and sensitive transcriptomic profiling of tumors. Uncontrolled proliferation, evasion of apoptosis, angiogenesis, metastasis and immune escape in cancer result from dysregulation of transcriptional processes. RNA-Seq can detect these complex transcriptional aberrations with an unprecedented precision needed to understand the behavior of the tumor (Wang et al., 2009).

RNA-Seq is an excellent addition to oncology in the detection of DEGs between malignant and normal tissues. These genes are involved in tumor initiation and progression and may be used as diagnostic biomarkers or therapeutic targets. This approach allows to identify oncogenes overexpressed and tumor suppressor genes suppressed, thus helping researchers to elucidate mechanisms of disease and to find targeted therapeutic genes (Hanahan, 2022). RNA-Seq also reveals cancer-specific molecular markers that define tumor subtypes with distinct prognostic and therapeutic implications.

Another benefit of RNA-Seq is that it can detect transcriptome events other than changes in gene expression. Alternative splicing, gene fusions, allele-specific expression, RNA editing and dysregulation of non-coding RNA are the basis for cancer development and progression. Microarrays miss many such events. RNA-Seq is better suited to analyze fusion genes, such as BCR-ABL in leukemia and EML4-ALK in lung cancer, with major therapeutic implications (Mertens et al., 2015). In addition, the study of microRNA and long non-coding RNA has uncovered regulatory networks in tumor suppression and oncogenic signaling.

RNA-Seq is also increasingly being applied in translational and clinical oncology. Now we can diagnose early, predict prognosis, choose therapy and monitor the response to therapy. Transcriptomic profiling and precision medicine enable tumor-specific therapy based on genetic features, rather than on histology-based categorization. RNA-Seq is therefore both a research tool and a rapidly growing clinical resource transforming oncology and personalized cancer therapy. (Tang et al., 2019).

RNA Sequencing

RNA sequencing (RNA-Seq) is a high throughput next generation sequencing (NGS) based method for transcriptome analysis of a biological sample at a specific point of time. It quantifies and qualitatively analyzes gene expression by sequencing complementary DNA (cDNA) produced from RNA molecules and provides a broad view of transcriptional activity in the cell. RNA-Seq is an unbiased approach to identify known and novel transcripts, and is a powerful tool for transcriptome research in complex disorders like cancer (Wang et al., 2009). Conventional hybridization methods use fixed probes.

RNA-Seq is based on the conversion of RNA to cDNA, fragmentation, library preparation and high-throughput sequencing on complex systems. Sequencing data are computationally aligned to a reference genome or de novo assembled to reconstruct transcript structures and estimate gene expression . This method enables high sensitivity and large dynamic range detection of highly expressed genes and low abundance transcripts that may regulate cancer biology (Ozsolak & Milos, 2011).

RNA-Seq has single nucleotide resolution and can detect small genetic changes, alternative splicing, RNA editing and gene fusions. These molecular processes are important since small changes in regulation of gene expression can lead to uncontrolled cell proliferation, metastasis and drug resistance in cancer. RNA-Seq can also find non-coding RNAs like miRNAs and lncRNAs that have a major role in tumor growth and progression (Stark et al., 2019).

Table (1) Steps in RNA-Seq Workflow and Their Purpose

Step	**Description**	**Importance**
RNA extraction	Isolation of total RNA	Starting material quality
Enrichment	mRNA selection	Focus on target transcripts
cDNA synthesis	RNA → cDNA conversion	Stability for sequencing
Sequencing	High-throughput read generation	Data production
Analysis	Alignment & quantification	Biological interpretation

Major Sequencing Platforms

With the advent of high-throughput sequencing systems, the development of RNA sequencing (RNA-Seq) technologies has transformed the accuracy, depth, and efficiency of transcriptomic studies. Different sequencing systems have different biochemistry and technical concepts and each has advantages and disadvantages that influence their suitability for applications in cancer transcriptomics. These platforms for RNA-Seq data analysis and selection of gene expression analysis tools should be known to oncology researchers.

Illumina's sequencing-by-synthesis method is the most common platform for RNA-Seq. The platform for its high precision, low error rates, scalability and cost effectiveness is the industry leader. Millions of short reads are generated in parallel by Illumina sequencing, which can be applied to large scale transcriptome studies including differential gene expression analysis, biomarker discovery, and cancer subtyping. It provides high throughput for most cancer genomics research and large-scale projects such as the Cancer Genome Atlas (TCGA) (Goodwin et al., 2016).

A further important platform is the Ion Torrent sequencing from Thermo Fisher Scientific. The Ion Torrent method detects hydrogen ions released as DNA polymerizes, which means that it can sequence faster and has less complicated equipment than optical methods. While it has shorter read lengths and higher error rates in homopolymer regions compared to Illumina, it is still widely used in targeted sequencing applications and smaller scale transcriptomic studies, particularly in clinical research where a rapid turnaround time is needed (Reuter et al., 2015).

Long-read sequencing has also greatly improved transcriptome analysis. Long reads spanning transcripts are available from single-molecule real-time (SMRT) sequencing platforms from Pacific Biosciences (PacBio). This capability allows for the discovery of alternate splicing, fusion genes and structural variations in cancer genomes not captured by short-read technologies. However, its use in large-scale investigations has been hampered by higher costs and lower throughput than short-read sequencing.

Oxford Nanopore Technologies' revolutionary real-time, portable nanopore sequencing technology is. Biological nanopores are used to pass molecules of nucleic acids through and the electrical current is measured to determine the nucleotide sequences. Its ultra-long reads and real-time processing make it ideal for structural variation identification and rapid clinical diagnosis. Nanopore sequencing is extensively used to detect complex genomic rearrangements and monitor tumor growth in cancer research.

Table (2) Characteristics of RNA-Seq Sequencing Platforms

Platform	**Read Type**	**Strengths**	**Limitations**	**Main Use**
Illumina	Short reads	High accuracy, low cost	Limited read length	Bulk RNA-Seq
Ion Torrent	Short reads	Fast sequencing	Homopolymer errors	Clinical targeted studies
PacBio	Long reads	Full-length transcripts	High cost	Splicing & fusion detection
Oxford Nanopore	Ultra-long reads	Real-time sequencing	Higher error rate	Structural variants

Limitations and Technical Challenges

RNA sequencing (RNA-Seq) has revolutionized cancer transcriptomics, but its research and therapeutic applications are hampered by methodological, technological and analytical challenges. Such procedural constraints, related to sample handling, sequencing and downstream bioinformatics analysis, may affect the reliability and reproducibility of results.

The main concerns are RNA integrity and sample quality. RNA molecules are easily degraded, especially in clinical cancer specimens with different tissue collection, storage and preservation methods. Degraded RNA can result in altered transcript representation and reduced detection sensitivity, especially for low-abundance transcripts, which may be physiologically relevant in tumor development (Tariq et al., 2021). This is particularly problematic in retrospective clinical studies, where the quality of the samples is not controlled.

The main problems are complexity and lack of standardization of bioinformatics workflows. RNA-Seq data analysis includes read mapping , transcript quantification , normalization and differential expression analysis . The analysis of the same data set using different pipelines and parameter choices can lead to quite different results, which is a challenge for reproducibility and consistency across research. Cancer genomics is also particularly difficult since the biological interpretation can vary with small differences in expression profiles (Conesa et al., 2016).

Technical limitations: depth and coverage of sequencing. Insufficient sequencing depth may miss rare transcripts or low frequency subclonal populations that may contribute to tumor heterogeneity and drug resistance. However, deep sequencing increases the cost of experiment without any biological knowledge . Experimental design is a major optimization problem for large-scale cancer studies (Soneson & Delorenzi, 2013)

Batch effects remain an issue in RNA-Seq studies . Differences in library preparation, sequencing platforms, reagent lots, or laboratory conditions can create systematic bias unrelated to biological variance. Batch effects can obscure differential expression signals or lead to false positives if not handled by statistical normalisation (Leek et al. 2010).

The complexity of the transcriptome also complicates the interpretation of RNA-Seq data. Biological complexity is hard to make sense of because of alternative splicing, overlapping genes, fusion transcripts and non-coding RNA controls. Genomic instability and transcriptome dysregulation associated with cancer often cause ambiguous read mapping and isoform assignment.

Finally, cost and computational infrastructure are still major barriers for large cohort research and for clinical deployment. The dropping prices of sequencing over the past decade has made high-throughput RNA-Seq a significant investment requiring high-performance computation and bioinformatics skills that may inhibit its application in low-resource settings .

RNA-Seq is one of the most powerful techniques in cancer genomics, but these technical and analytical hurdles need to be overcome to achieve reproducibility, clinical translation and biological results .

Functional Enrichment Analysis

Functional enrichment analysis is an integral part of RNA-Seq data interpretation, particularly in the cancer transcriptomics field where thousands of DEGs are identified. Functional enrichment methods are used to identify overrepresented biological processes, molecular functions and signaling pathways in order to interpret transcriptional changes from raw gene expression data.

Functional enrichment analysis attempts to identify if particular gene sets, for instance, genes involved in the same biological pathway or cellular function, are statistically overrepresented in DEGs. This enables researchers to study coordinated biological systems that drive cancer formation and progression, instead of gene-level changes (Khatri et al., 2012.(

Gene Ontology (GO) analysis is a commonly used functional annotation framework that classifies genes into three categories: biological process, molecular activity and cellular component. The GO analysis is used in cancer research to detect disrupted cell cycle control, apoptosis, DNA repair, immunological response and metabolic reprogramming. These pathways are often dysregulated in tumor cells, and are characteristic of systems-level cancer biology (Ashburner et al., 2000.(

Besides GO analysis, pathway-based enrichment analysis such as KEGG is also performed in many RNA-Seq datasets. KEGG pathway analysis showed significant enrichment of signaling pathways such as PI3K-Akt, MAPK, p53 and Wnt that are critical in tumor initiation, progression and metastasis. The pathway-centric view helps to understand how different gene alterations in biological networks contribute to oncogenesis (Kanehisa et al., 2017.(

Overrepresentation analysis is also important because Gene Set Enrichment Analysis (GSEA) does not only consider pre-selected DEGs but all genes in the dataset. GSEA tests for statistically significant, concordant differences between two biological states (e.g. phenotypes) such as tumors and normal tissue in pre-defined sets of genes. This technique is useful in cancer research, as threshold-based DEG analysis may miss minor, yet coordinated, gene expression changes (Subramanian et al., 2005).

Table (3) Major Functional Enrichment Methods in RNA-Seq Analysis

Method	**Purpose**	**Output**
GO Analysis	Functional categorization	Biological processes, functions

KEGG Pathway Analysis	Pathway mapping	Signaling pathways
GSEA	Gene set comparison	Enriched gene sets
Reactome Analysis	Network pathways	Biological interaction maps

Detection of Alternative Splicing Events

Alternative splicing, a key post-transcriptional regulation process, generates several mRNA isoforms from a single gene, boosting transcriptome and proteomic diversity. In cancer biology, aberrant alternative splicing is increasingly linked to tumor genesis, development, immune evasion, and treatment resistance. RNA sequencing (RNA-Seq) is the most powerful method for genome-wide splicing event detection and characterisation, delivering insights that microarray-based approaches could not.

Mutations in splicing factors, regulatory protein expression, or epigenetic alterations impacting splice site selection can cause dysregulated splicing patterns in cancer cells. These changes can produce oncogenic isoforms or eliminate tumor-suppressive ones. Alternative splicing can produce protein isoforms with improved proliferative potential, apoptosis resistance, or subcellular localization, which contribute to malignant transformation and tumor heterogeneity (Black, 2003).

High-resolution sequencing reads that bridge exon–exon junctions allow RNA-Seq to detect alternative splicing events. This identifies exon skipping, intron retention, mutually exclusive exons, and alternative 5' or 3' splice sites. RNA-Seq data processing techniques may rebuild transcript isoforms and measure their abundance, allowing malignant and normal tissue splicing patterns to be compared (Wang et al., 2008).

Tumor-specific isoforms that might be biomarkers or therapeutic targets are one of the most important effects of alternative splicing in cancer. In breast, lung, and colorectal cancer, particular splice variations are linked to disease development and poor prognosis. These isoforms may affect medication sensitivity, causing treatment resistance and illness recurrence.

Hematological and solid cancers often include mutations in key splicing machinery components such SF3B1, SRSF2, and U2AF1, which are connected to dysregulated splicing. RNA-Seq studies have demonstrated that these mutations can modify splicing patterns, transforming the transcriptome landscape of cancer cells and influencing many biological pathways (Seiler et al., 2018).

Discovery of Fusion Genes

Structural genomic rearrangements link two genes to generate gene fusions resulting in chimeric transcripts with altered or new biological functions. In cancer gene fusions activate signaling pathways , alter tumor suppressor genes or generate constitutively active proteins leading to uncontrolled cell proliferation . RNA sequencing (RNA-Seq) can directly sequence expressed transcripts without genomic breakpoints and has become a key tool for fusion gene discovery and characterization.

In contrast to FISH or karyotyping, the nucleotide-level resolution of RNA-Seq permits genome-wide detection of fusion transcripts including rare and previously uncharacterized events. This is important because fusion genes are often used as diagnostic markers and therapeutic targets in cancer research. In chronic myeloid leukemia the BCR–ABL fusion gene generates a constitutive tyrosine kinase activity that has led to highly successful targeted drugs such as imatinib (Mitelman et al., 2007).

In RNA-Seq based fusion detection, the sequencing reads that span exon-exon junctions from different genes or the discordant read pairs that align to different genomic loci are identified. Bioinformatics tools such as STAR-Fusion, FusionCatcher and Arriba are used to detect and validate RNA-Seq fusion events . These computational approaches have increased the sensitivity and specificity of fusion gene discovery, allowing for the identification of clinically relevant rearrangements in multiple cancer types (Haas et al., 2019).

Gene fusions such as EML4–ALK in non–small cell lung cancer and TMPRSS2–ERG in prostate cancer play a crucial role in carcinogenesis and disease progression. Both diagnostic biomarkers and therapeutic targets for precision oncology, these fusion events. The direct translational impact of RNA-Seq–based fusion identification is exemplified by the benefit of ALK drugs to patients with ALK rearrangements.

Gene fusions are also more common and diverse than expected, particularly in genomically unstable malignancies, as also revealed by RNA-Seq. Such fusion events are rare or specific to tumor subclones emphasizing the need for high depth sequencing and sensitive computational methods. Some of the fusion transcripts may not encode functional proteins, but can regulate gene expression and tumor biology through non-coding mechanisms.

Table (4) Clinically Relevant Gene Fusions in Cancer

Fusion Gene	**Cancer Type**	**Clinical Significance**
BCR–ABL	Leukemia	Targeted therapy (Imatinib)
EML4–ALK	Lung cancer	ALK inhibitors
TMPRSS2–ERG	Prostate cancer	Diagnostic marker
FGFR3–TACC3	Glioblastoma	Therapeutic target

Tumor Heterogeneity and Single-Cell RNA-Seq

Tumor heterogeneity is one of the most fundamental biological hallmarks of cancer and represents a major challenge to achieve effective and durable therapeutic responses. It deals with the issue of coexistence of genetically, epigenetically and phenotypically different populations of cells within the same tumour mass, or between primary and metastatic lesions. This heterogeneity is driven by ongoing genetic mutations, clonal evolution, and selective pressures from the tumor microenvironment and therapeutic interventions. Therefore, tumors are not homogeneous cell populations but highly dynamic ecosystems, which complicates the diagnosis, prognosis and treatment strategies (McGranahan & Swanton, 2017).

Standard bulk RNA sequencing (RNA-Seq) quantifies average gene expression across thousands or millions of cells, obscuring the existence of rare, but biologically and clinically important sub-populations. This averaging effect limits the resolution of intratumoral diversity and conceals important information on resistant clones, cancer stem cells and transient cellular states that drive tumor progression and relapse (Alizadeh et al., 2015).

This limitation has been overcome by the development of single-cell RNA sequencing (scRNA-Seq) providing transcriptomic profiling at the resolution of individual cells. scRNA-Seq is extensively applied in cancer studies to characterize discrete tumor cell populations, reconstruct lineage hierarchies and discover rare cell types, driving disease aggressiveness and drug resistance (Zheng et al., 2017).

One of the major contributions of scRNA-Seq is the detection of cancer stem cell populations and drug resistant clones. These populations are often the cause of recrudescence of the tumor after treatment. These cells are generally rare, and therefore not detectable by bulk RNA-Seq methods. Single-cell gene expression analysis enables the classification of the functional states of individual cells and their contribution to tumor maintenance and progression (Tirosh et al., 2016).

scRNA-Seq has also played a major role in our understanding of the tumor microenvironment, which includes immune cells, fibroblasts, endothelial cells, and other stromal components besides malignant cells. These non-cancerous cells interact actively with tumor cells and play an important role in immune escape, angiogenesis and metastasis. Single-cell analyses have revealed the immune complexity of tumors with exhausted T cells and immunosuppressive cell populations driving resistance to immunotherapy (Kumar et al., 2018).

Also, scRNA-Seq can be utilized to reconstruct tumor evolutionary trajectories by computational approaches, such as pseudotime analysis. Such approaches enable the reconstruction of the temporal ordering of cellular states, providing insights into the evolution of cancer cells in response to therapy and environmental constraints. This dynamic view is crucial for understanding the mechanisms behind resistance development and tumor adaptation (Saelens et al., 2019).

Table (5) Cellular Components of the Tumor Microenvironment Identified by scRNA-Seq

Cell Type	Function	Role in Cancer
Cancer stem cells	Self-renewal	Tumor recurrence
Cytotoxic T cells	Immune killing	Anti-tumor response
Regulatory T cells	Immune suppression	Tumor immune escape
Cancer-associated fibroblasts	ECM remodeling	Tumor progression
Endothelial cells	Angiogenesis	Tumor vascularization

Clinical Applications and Biomarker Discovery

RNA sequencing (RNA-Seq) is used in clinical oncology for the discovery of diagnostic, prognostic and predictive biomarkers. Transcriptomic biomarkers can detect molecular changes in the tumor that may result in earlier detection, classification of the disease and selection of treatment. In contrast to single gene experiments, RNA-Seq measures thousands of transcripts simultaneously, thereby increasing the probability of discovering clinically useful molecular signatures.

The most important clinical application of RNA-Seq is in the subtyping of cancer. Breast, lung and colorectal cancers have multiple molecular subgroups that influence prognosis and treatment. RNA-Seq transcriptome profiling revealed gene expression signatures which are more discriminating of these subtypes than histopathological approaches. Molecular stratification improves patient outcome and guides treatment decisions in personalized medicine (Perou et al., 2000).

RNA-Seq has been useful in the identification of prognostic biomarkers, genes or gene signatures that predict outcome of disease irrespective of treatment. For instance, overexpression of proliferation-related genes or downregulation of immune response pathways can suggest tumor aggressiveness and reduced survival. These transcriptome markers are increasingly incorporated into clinical decision-making models for patient risk assessment and follow-up (Michels et al., 2019).

RNA-Seq is also commonly used to develop predictive biomarkers that predict tumor response to a specific therapy. Expression levels of genes involved in DNA repair, immunological checkpoint control or drug metabolism may predict sensitivity or resistance to chemotherapy, targeted therapy or immunotherapy . This is of crucial importance in precision oncology, as treatment is based on tumor molecular profiles (Simon, 2015).

A major clinical application of RNA-Seq is the analysis of circulating tumor RNA (ctRNA) or extracellular vesicles from blood samples by liquid biopsy. This minimally invasive approach allows for real-time diagnosis of tumor dynamics, treatment response and recurrence. Transcriptome profiling from liquid biopsies is an emerging field with promise as an alternative or complement to tissue biopsies, especially when tumor tissue is limited.

In clinical practice, multi-gene expression panels developed by RNA-Seq were used. These panels aggregate multiple gene expressions into a risk score or classification system to improve diagnosis and prognosis in comparison to single biomarkers. There are commercial breast cancer diagnostics to guide treatment.

RNA-Seq in Precision Oncology and Future Perspectives

RNA sequencing (RNA-Seq) is increasingly used in precision oncology, where cancer treatment is tailored to the genetic profile of each patient as opposed to tumor site or histological categorization. RNA-Seq reveals changes in gene expression that can help physicians and researchers target drugs, predict response to therapies, and track the progression of disease in a personalized manner.

RNA-Seq based stratification of molecular signatures is required for precision oncology. These fingerprints are the outcome of biological processes including proliferation, immune activity and metabolic reprogramming, which may differ even within the same cancer type. This stratification allows more accurate evaluation of the risk to patients and the choice of treatments, resulting in improved clinical outcomes and lower toxicity from inadequate medicines (Vogelstein et al., 2013).

RNA-Seq is also useful to identify targets for targeted therapy. A lot of carcinogenic drivers are defined by transcriptional activity and DNA mutations. The RNA-Seq can be used for therapy by identifying the overexpression of oncogenes, signaling pathway activity, and dysregulation of drug resistance. The expression of immune checkpoint-related genes may indicate the therapeutic effect of immunotherapy, and the activation of growth factor signaling pathways may indicate the suitability of targeted inhibitors (Garraway & Lander, 2013).

RNA-Seq is also widely used for treatment response and monitoring of disease progression . Longitudinal transcriptome profiling may help clinicians monitor tumor gene expression in response to therapy. By monitoring dynamically, resistance can be detected early and treatment strategy altered. Some RNA-Seq based liquid biopsy approaches are under investigation for less invasive real-time monitoring of tumor evolution.

RNA-Seq has the tremendous clinical potential, but its implementation in clinical practice still remains difficult. Need for standardized methods, faster turnaround, lower costs and strong clinical validation of transcriptome biomarkers Clinical decision-making from RNA-Seq data will require collaboration between clinicians, bioinformaticians and molecular biologists and the use of modern computational methods.

Improvements in single-cell sequencing, spatial transcriptomics, and AI in the future should increase the therapeutic value of RNA-Seq. These technologies will better characterize tumor heterogeneity, interactions with the microenvironment and geographic gene expression patterns RNA-Seq and other technologies are supposed to bring a data-driven, personalized approach to cancer care.

Conclusions

In the field of cancer genomics, RNA sequencing, or RNA-Seq, has revolutionized the field by allowing a comprehensive and high-resolution examination of gene expression across the entire transcriptome. RNA-Seq method is a quantitative, and objective method unlike the other technologies. It allows the identification of known and novel transcripts, alternative splicing events, gene fusions and non-coding RNAs, all accounting for the molecular complexity of cancer.

The current study aims to show that RNA-Seq is not only a tool for differential gene expression studies but also an important platform to understand multiple layers of cancer biology. These layers encompass tumor heterogeneity, regulatory RNA networks, and pathway dysregulation. The combination of this technology with single cell sequencing technologies has greatly enhanced the ability to dissect tumor ecosystems at a cellular level. This has enabled the identification of the rare but clinically relevant subpopulations that drive tumor growth and treatment resistance.

RNA-sequencing has also been shown to be very useful in clinical oncology, especially in the areas of biomarker discovery, cancer subtyping and precision medicine. The translational potential of RNA-Seq data in modern cancer treatment is underlined by the growing utilization of transcriptomic markers derived from RNA-Seq data to support diagnosis, predict treatment response and monitor disease progression.

But still, even with these benefits there are a number of obstacles to overcome. These hurdles are the need of uniform analytical pipelines, data complexity, high processing requirements and technical variability. These limitations need to be overcome in order to enhance reproducibility and to guarantee reliable translation of RNA-Seq into routine clinical practice.

The future of cancer biology is promising with the integration of multi-omics, spatial transcriptomics and artificial intelligence with RNA-Seq technology that will bring about a drastic improvement in our understanding of the biology of cancer. The resulting advances are likely to greatly accelerate the move to fully personalized oncology where treatment regimens are tailored to the specific genetic profile of each individual patient’s tumor.

References

1. Alizadeh, A. A., Aranda, V., Bardelli, A., Blanpain, C., Bock, C., Borowski, C., Caldas, C., Califano, A., Doherty, M., Elsner, M., Esteller, M., Fitzgerald, R., Korbel, J. O., Lichter, P., Mason, C. E., Navin, N., Pe’er, D., Polyak,

K., Roberts, C. W. M., ... Zuber, J. (2015). Toward understanding and exploiting tumor heterogeneity. *Nature Medicine, 21*(8), 846–853.

2. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: Tool for the unification of biology. *Nature Genetics, 25*(1), 25–29.
3. Black, D. L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annual Review of Biochemistry, 72*, 291–336.
4. Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. *New England Journal of Medicine, 372*(9), 793–795.
5. Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology, 17*, 13.
6. Dagogo-Jack, I., & Shaw, A. T. (2018). Tumour heterogeneity and resistance to cancer therapies. *Nature Reviews Clinical Oncology, 15*(2), 81–94.
7. Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses. *The FASEB Journal, 22*(2), 338–342.
8. Garraway, L. A., & Lander, E. S. (2013). Lessons from the cancer genome. *Cell, 153*(1), 17–37.
9. Gusenbauer, M., & Haddaway, N. R. (2020). Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Research Synthesis Methods, 11*(2), 181–217.
10. Haas, B. J., Dobin, A., Stransky, N., Li, B., Yang, X., Tickle, T., et al. (2019). STAR-Fusion: Fast and accurate fusion transcript detection from RNA-Seq. *bioRxiv*. https://doi.org/10.1101/120295
11. Haber, D. A., & Velculescu, V. E. (2014). Blood-based analyses of cancer: Circulating tumor cells and circulating tumor DNA. *Cancer Discovery, 4*(6), 650–661.
12. Hanahan, D. (2022). Hallmarks of cancer: New dimensions. *Cancer Discovery, 12*(1), 31–46.
13. Hinshaw, D. C., & Shevde, L. A. (2019). The tumor microenvironment innately modulates cancer progression. *Cancer Research, 79*(18), 4557–4566.
14. Hoadley, K. A., Yau, C., Hinoue, T., Wolf, D. M., Lazar, A. J., Drill, E., Shen, R., Taylor, A. M., Cherniack, A. D., Thorsson, V., Akbani, R., Bowlby, R., Wong, C. K., Wiznerowicz, M., Sanchez-Vega, F., Robertson, A. G., Schneider, B. G., Lawrence, M. S., Noushmehr, H., ... Laird, P. W. (2018). Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell, 173*(2), 291–304.
15. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2017). KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research, 45*(D1), D353–D361.
16. Khatri, P., Sirota, M., & Butte, A. J. (2012). Ten years of pathway analysis: Current approaches and outstanding challenges. *PLoS Computational Biology, 8*(2), e1002375.
17. Kumar, M. P., Duan, Q., McKenzie, I. M., & Chronister, W. D. (2018). Single-cell RNA-seq reveals molecular and functional heterogeneity in cancer. *Nature Genetics, 50*(8), 1212–1223.
18. Leek, J. T., Scharpf, R. B., Bravo, H. C., Simcha, D., Langmead, B., Johnson, W. E., Geman, D., Baggerly, K., & Irizarry, R. A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics, 11*(10), 733–739.
19. Liu, X., et al. (2019). Multi-omics data integration in cancer research. *Nature Reviews Genetics, 20*(6), 317–332.
20. Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & López-Cózar, E. D. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics, 12*(4), 1160–1177.
21. McGranahan, N., & Swanton, C. (2017). Clonal heterogeneity and tumor evolution: Past, present, and the future. *Cell, 168*(4), 613–628.
22. Mertens, F., Johansson, B., Fioretos, T., & Mitelman, F. (2015). The emerging complexity of gene fusions in cancer. *Nature Reviews Cancer, 15*(6), 371–381.
23. Michels, K. B., Hunter, D. J., & Chan, A. T. (2019). Biomarkers for cancer risk prediction. *Nature Reviews Cancer, 19*(6), 375–387.
24. Mitelman, F., Johansson, B., & Mertens, F. (2007). The impact of translocations and gene fusions on cancer causation. *Nature Reviews Cancer, 7*(4), 233–245.
25. Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), e1000097.
26. Ozsolak, F., & Milos, P. M. (2011). RNA sequencing: Advances, challenges and opportunities. *Nature Reviews Genetics, 12*(2), 87–98.
27. Perou, C. M., Sørlie, T., Eisen, M. B., Van De Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, Ø., Pergamenschikov, A., Williams, C., Zhu, S. X., Lønning, P. E., Børresen-Dale, A. L., Brown, P. O., & Botstein, D. (2000). Molecular portraits of human breast tumours. *Nature, 406*(6797), 747–752.
28. Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., et al. (2000). Molecular portraits of human breast tumours. *Nature, 406*(6797), 747–752.
29. Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics, 26*(1), 139–140.

30. Saelens, W., Cannoodt, R., Todorov, H., & Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. *Nature Biotechnology, 37*(5), 547–554.
31. Seiler, M., Peng, S., Agrawal, A. A., Palacino, J., Teng, T., Zhu, P., et al. (2018). Somatic mutational landscape of splicing factor genes and their functional consequences across 33 cancer types. *Cell Reports, 23*(1), 282–296.
32. Shlien, A., & Malkin, D. (2015). Copy number variations and gene fusions in cancer. *Nature Reviews Cancer, 15*(9), 559–573.
33. Simon, R. (2015). Clinical trials for predictive medicine: Biomarker-based therapy selection. *Clinical Cancer Research, 21*(19), 4231–4236.
34. Soneson, C., & Delorenzi, M. (2013). A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics, 14*, 91.
35. Stark, R., Grzelak, M., & Hadfield, J. (2019). RNA sequencing: The teenage years. *Nature Reviews Genetics, 20*(11), 631–656.
36. Stransky, N., Cerami, E., Schalm, S., Kim, J. L., Lengauer, C., & Brennan, C. (2014). The landscape of kinase fusions in cancer. *Nature Communications, 5*, 4846.
37. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences, 102*(43), 15545–15550.
38. Subramanian, I., et al. (2020). Multi-omics data integration approaches in systems biology. *Briefings in Bioinformatics, 21*(6), 1939–1952
39. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians, 71*(3), 209–249.
40. Tang, X., Huang, Y., Lei, J., Luo, H., & Zhu, X. (2019). The single-cell sequencing: New developments and medical applications. *Cell & Bioscience, 9*(1), 53.
41. Tariq, M. A., Kim, H. J., Jejelowo, O., & Pourmand, N. (2021). Whole-transcriptome RNA sequencing data: Challenges and opportunities for biological discovery. *Journal of Biomedical Science, 28*(1), 1–15.*
42. Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., Treacy, D. J., Trombetta, J. J., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science, 352*(6282), 189–196.
43. Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., & Kinzler, K. W. (2013). Cancer genome landscapes. *Science, 339*(6127), 1546–1558.
44. Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., et al. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature, 456*(7221), 470–476.
45. Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics, 10*(1), 57–63.
46. World Health Organization. (2024). Cancer fact sheet. Geneva: World Health Organization.
47. Zhang, A., Sun, H., Wang, P., Han, Y., & Wang, X. (2018). Recent advances in biomarker discovery based on RNA sequencing in cancer. *Cancer Letters, 418*, 1–9.
48. Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., et al. (2017). Massively parallel digital transcriptional profiling of single cells. *Nature Communications, 8*, 14049.