ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may155

From Long Reads to Precision Oncology: Recent Advances in Sequencing Technologies and their Emerging Role in Cancer Diagnostics

Manan Chandra

Publication Date: 2026/05/25

Abstract: Recent advances in long-read sequencing technologies have significantly expanded the diagnostic capabilities of next-generation sequencing (NGS) in oncology. While short-read sequencing has enabled large-scale genomic profiling, it remains limited in resolving structural variants, gene fusions, and complex genomic rearrangements that characterize tumor genomes. Long-read platforms now enable improved detection of these features, along with direct analysis of epigenetic modifications and transcript isoforms. In this review, we present a comprehensive and integrated overview of long-read sequencing technologies within the context of cancer diagnostics, spanning the full NGS workflow from sample preparation and quality control to sequencing, demultiplexing, and clinical interpretation. We highlight key applications in structural variant detection, liquid biopsy, transcriptomics, and precision oncology, while critically examining current limitations and outlining future directions for clinical translation.

How to Cite: Manan Chandra (2026) From Long Reads to Precision Oncology: Recent Advances in Sequencing Technologies and their Emerging Role in Cancer Diagnostics. *International Journal of Innovative Science and Research Technology*, 11(5), 1505-1515. https://doi.org/10.38124/ijisrt/26may155

I. INTRODUCTION

Cancer remains one of the leading causes of morbidity and mortality worldwide, driven by a complex and dynamic landscape of genomic, transcriptomic, and epigenetic alterations (Sung et al., 2021). Tumor genomes are characterized not only by single nucleotide variants but also by large-scale structural rearrangements, copy number alterations, gene fusions, and regulatory changes that collectively influence disease progression and therapeutic response (Vogelstein et al., 2020). Accurate characterization of these features is therefore essential for diagnosis, prognosis, and the implementation of precision oncology.

The development of DNA sequencing technologies has played a central role in enabling such molecular characterization. Early sequencing efforts were based on Sanger sequencing, which provided highly accurate base-level resolution but was inherently low-throughput and limited in scalability (Sanger et al., 1977). While Sanger sequencing was instrumental in identifying key oncogenic mutations, its application in cancer diagnostics was largely restricted to targeted analysis of individual genes.

The advent of high-throughput next-generation sequencing (NGS), particularly through platforms developed by Illumina, transformed the field by enabling massively parallel sequencing of millions of DNA fragments (Metzker, 2010). This shift allowed comprehensive genomic profiling and facilitated the development of clinically relevant assays, including targeted gene panels, whole-exome sequencing, and transcriptome analysis. As a result, short-read sequencing became the foundation of modern cancer diagnostics.

Despite these advances, short-read sequencing technologies remain fundamentally constrained by limited read length, typically ranging from 50 to 300 base pairs. This limitation restricts the ability to resolve complex genomic features such as structural variants, repetitive regions, and long-range haplotypes (Amarasinghe et al., 2020; Logsdon et al., 2020). In cancer genomes, where chromosomal instability and genomic rearrangements are common, these constraints can lead to incomplete or ambiguous characterization of clinically relevant alterations.

Long-read sequencing technologies developed by Pacific Biosciences and Oxford Nanopore Technologies have emerged as powerful tools to address these limitations. By generating reads that span kilobases to megabases, these platforms enable direct detection of structural variants, phasing of alleles, and simultaneous analysis of epigenetic modifications (Logsdon et al., 2020; van Dijk et al., 2023). These capabilities provide a more comprehensive view of tumor biology and offer new opportunities for improving cancer diagnostics.

In this review, we present an integrated overview of recent advances in long-read sequencing technologies within the context of cancer diagnostics. We examine the evolution of sequencing approaches, the end-to-end NGS workflow, and key applications in structural variant detection, transcriptomics, epigenetic profiling, and liquid biopsy. We further discuss current challenges and future directions, with

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may155

a focus on the clinical translation of long-read sequencing in precision oncology.

II. EVOLUTION OF SEQUENCING TECHNOLOGIES IN ONCOLOGY

The evolution of sequencing technologies reflects a progressive effort to overcome technical limitations that have historically constrained cancer genomics. Each generation of sequencing has expanded analytical capabilities, yet has also revealed new challenges, particularly in the context of tumor heterogeneity and genomic complexity.

Early sequencing efforts relied on Sanger sequencing, which provided highly accurate base-level resolution but was inherently low-throughput and labor-intensive. While Sanger sequencing played a foundational role in identifying key oncogenic mutations, its limited scalability restricted its use to targeted analysis of individual genes or small genomic regions. Consequently, early cancer diagnostics were largely confined to detecting known mutations in candidate genes, without the ability to capture broader genomic alterations.

The introduction of short-read next-generation sequencing (NGS) platforms, particularly those developed by Illumina, marked a transformative shift in oncology. By enabling massively parallel sequencing of millions of DNA fragments, these platforms dramatically increased throughput while reducing cost. This facilitated genome-wide analyses and enabled the development of clinically relevant assays such as targeted gene panels, whole-exome sequencing (WES), and transcriptome profiling (Metzker, 2010). As a result, short-read sequencing became the backbone of modern cancer diagnostics, supporting the identification of single nucleotide variants, small insertions and deletions, and gene expression changes.

However, the inherent limitation of short-read sequencing lies in its read length, typically ranging from 50 to 300 base pairs. This constraint prevents accurate reconstruction of complex genomic regions and limits the ability to resolve structural variants, repetitive elements, and long-range haplotypes. In cancer genomes—where chromosomal rearrangements, gene fusions, and copy number alterations are common—these limitations become particularly significant (Amarasinghe et al., 2020). Many clinically relevant alterations are therefore either missed entirely or inferred indirectly, reducing diagnostic precision.

Long-read sequencing technologies represent a critical advancement by addressing these limitations through the generation of reads spanning kilobases to megabases. Platforms developed by Pacific Biosciences and Oxford Nanopore Technologies enable direct observation of structural variants, phasing of alleles, and characterization of complex genomic rearrangements (Logsdon et al., 2020). These capabilities are particularly important in oncology, where tumor genomes often exhibit high levels of structural complexity and heterogeneity.

Initially, long-read sequencing technologies were limited by relatively high error rates and lower throughput compared to short-read platforms. However, recent technological advancements have significantly improved their performance. High-fidelity (HiFi) sequencing has reduced error rates to levels approaching those of short-read technologies, while machine learning–based basecalling has improved accuracy in nanopore sequencing (Wenger et al., 2019). These improvements have enhanced the reliability of long-read data for clinical applications.

In addition, hybrid sequencing strategies that combine short- and long-read data have emerged as a practical approach for balancing accuracy and structural resolution. Short reads provide high base-level accuracy, while long reads enable reconstruction of complex genomic regions (Shafin et al., 2020). Together, these approaches offer a more comprehensive view of tumor genomes than either technology alone, as summarized in Figure 1.

The evolution of sequencing technologies in oncology therefore reflects a transition from targeted, low-throughput analysis toward comprehensive, multi-dimensional characterization of tumor genomes. Long-read sequencing represents a key step in this progression, enabling more complete and accurate detection of genomic alterations that are critical for cancer diagnosis, prognosis, and treatment selection.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may155

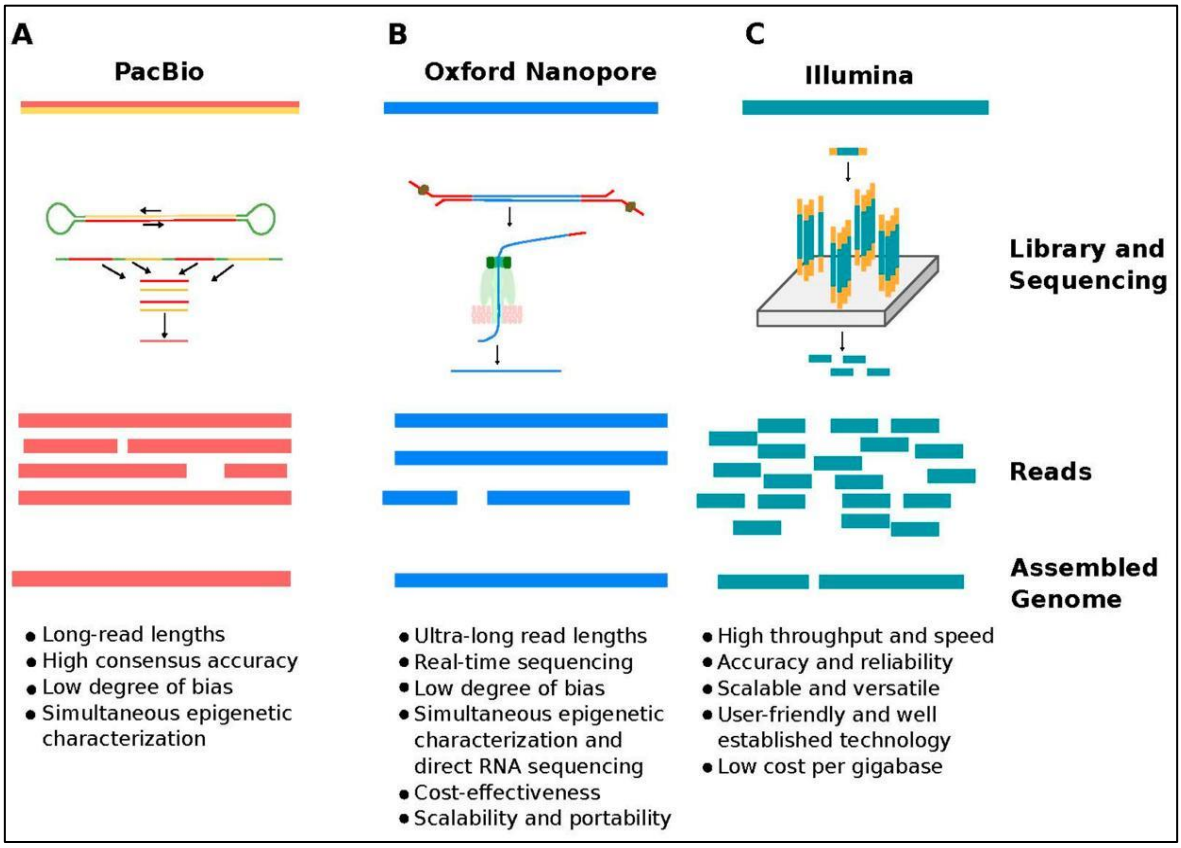

Fig 1 Comparison of Long-Read Sequencing Technologies, Highlighting Differences in Read Length, Structural Variant Detection, Gene Fusion Resolution, and Epigenetic Profiling Capabilities in Cancer Diagnostics.

III. END-TO-END NGS WORKFLOW IN CANCER DIAGNOSTICS

The successful application of sequencing technologies in cancer diagnostics depends not only on sequencing platforms themselves but also on the integrity and optimization of the entire workflow—from sample acquisition to computational analysis. Each step introduces potential sources of bias, error, and variability that can significantly impact downstream interpretation, as illustrated in Figure 2. Long-read sequencing, while offering substantial advantages, also imposes unique requirements on sample quality and processing, necessitating careful consideration at each stage of the pipeline.

➢ *Sample Types and Nucleic Acid Extraction*

Cancer genomic analyses rely on a diverse range of biological samples, including fresh or frozen tumor tissues, formalin-fixed paraffin-embedded (FFPE) specimens, peripheral blood for circulating tumor DNA (ctDNA), and cultured cells. Each sample type presents distinct challenges in terms of nucleic acid yield, integrity, and contamination.

FFPE samples, widely used in clinical practice, often contain degraded and chemically modified DNA due to formalin fixation. Given the challenges associated with degraded or low-input RNA—particularly from FFPE samples—specialized library preparation kits designed for low-input and partially degraded material are increasingly used to maintain transcriptome coverage and data quality (Jansson et al., 2024; Pignatta et al., 2025). In highly fragmented samples, alternative approaches such as single-stranded library preparation can improve recovery and representation of degraded DNA fragments (Gansauge and Meyer, 2020).

These modifications can introduce sequencing artifacts and reduce effective read length, particularly impacting long-read sequencing performance. In contrast, circulating cell-free DNA (cfDNA) is highly fragmented—typically around 150–200 base pairs—and present at low abundance, making it difficult to analyze using platforms optimized for long DNA molecules (Wan et al., 2017; Heitzer et al., 2019).

Long-read sequencing requires high-molecular-weight DNA to achieve its full potential in resolving structural variants and complex genomic regions. As a result, extraction protocols must be carefully optimized to minimize fragmentation and preserve DNA integrity. Mechanical shearing, enzymatic degradation, and suboptimal storage conditions can all compromise sample quality and limit the benefits of long-read approaches.

In practice, these constraints are particularly relevant in oncology, where compromised sample quality can directly limit the diagnostic yield of sequencing assays.

➢ *Quality Control of DNA, RNA, and Libraries*

Quality control (QC) is a critical determinant of sequencing success, particularly in oncology workflows

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may155

where sample quality is highly variable. Fluorometric quantification using the Qubit fluorometer provides accurate measurement of nucleic acid concentration, while fragment size distribution is assessed using capillary electrophoresis systems such as the Agilent TapeStation or Fragment Analyzer platforms.

For long-read sequencing, fragment size distribution is especially important, as read length is directly influenced by input DNA integrity. High-molecular-weight DNA enables longer reads and improved structural resolution, whereas degraded samples yield shorter reads that reduce analytical advantages (Amarasinghe et al., 2020).

Library quality control further ensures proper adapter ligation and fragment representation. Inadequate QC can lead to sequencing bias, uneven coverage, and reduced sensitivity for detecting low-frequency variants. In clinical contexts, where accurate detection of mutations may directly influence treatment decisions, stringent QC thresholds are essential.

Taken together, these quality control steps are not merely technical checkpoints but key determinants of downstream analytical reliability.

- *Library Preparation Strategies in Oncology*

Library preparation is a critical step that determines which regions of the genome are sequenced and how efficiently they are captured. In cancer diagnostics, different library preparation strategies are employed depending on the clinical question (Table 1).

Whole-genome sequencing (WGS) provides the most comprehensive view of the tumor genome, enabling detection of structural variants, copy number alterations, and large-scale rearrangements. Whole-exome sequencing (WES) focuses on coding regions and offers a cost-effective alternative for identifying mutations in protein-coding genes.

Targeted gene panels are widely used in clinical oncology to detect actionable mutations, offering high depth of coverage and sensitivity. Amplicon-based approaches are particularly useful for detecting hotspot mutations at low variant allele frequencies (Goodwin et al., 2016). RNA sequencing enables detection of gene fusions and transcript isoforms, while immune repertoire sequencing (TCR/BCR) provides insights into tumor-immune interactions. Targeted gene panels remain a cornerstone of clinical oncology workflows, particularly for the detection of actionable mutations in oncogenes and tumor suppressor genes. For example, assays such as the TruSight Oncology 500 enable comprehensive profiling of hundreds of cancer-associated genes, including single nucleotide variants, insertions and deletions, copy number alterations, and gene fusions within a single assay (Jennings et al., 2020). These panels are optimized for clinical use, offering high sensitivity, relatively low input requirements, and faster turnaround times compared to whole-genome approaches. However, their targeted nature inherently limits the detection of novel or complex genomic alterations, reinforcing the complementary role of broader sequencing strategies.

In addition to tumor-intrinsic genomic profiling, sequencing approaches targeting immune repertoires have gained importance in oncology. T-cell receptor (TCR) and B-cell receptor (BCR) sequencing enable characterization of clonality and immune diversity, providing insights into tumor–immune interactions and response to immunotherapy (Robins, H.S., et al. 2013). These approaches are particularly relevant in hematologic malignancies and in monitoring treatment response in solid tumors.

More recently, innovations in library preparation have focused on improving scalability and cost efficiency. For instance, bulk RNA barcoding and sequencing (BRB-seq) enables early multiplexing of samples through the incorporation of sample-specific barcodes during reverse transcription, allowing pooled processing of multiple samples in a single workflow (Alpern et al., 2019). When combined with automated liquid handling systems, such approaches can further enhance throughput and reproducibility, reducing both reagent consumption and hands-on variability.

Automation platforms are increasingly integrated into sequencing workflows to improve scalability and reproducibility. Systems such as the Agilent Bravo and Opentrons OT-2 enable standardized execution of nucleic acid extraction and library preparation protocols, reducing hands-on variability and operator-dependent bias. In addition to improving consistency, these platforms facilitate higher throughput processing and are particularly valuable in large-scale studies or clinical laboratories where reproducibility and efficiency are critical (Muir et al., 2020).

Long-read sequencing enhances these strategies by enabling phasing of variants, improved detection of structural variants, and accurate reconstruction of full-length transcripts (Wenger et al., 2019; Shafin et al., 2020). However, library preparation for long-read sequencing often requires higher input DNA quantities and careful handling to preserve fragment length, which can be challenging in clinical settings. The choice of library preparation strategy therefore reflects a balance between clinical need, sample quality, and the level of genomic resolution required.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may155

Table 1 Library Preparation Strategies in Cancer Diagnostics

Library Type	Target	Key Applications in Cancer	Strengths	Limitations
WGS	Entire genome	Structural variants, CNVs, genome-wide profiling	Comprehensive; detects all variant types	High cost; complex analysis
WES	Coding regions	Mutation detection in genes	Cost-effective; established	Misses non-coding; limited SV
Gene Panels	Selected genes	Actionable mutations	High sensitivity; fast	Limited scope
Amplicon	Specific loci	Hotspot mutations; MRD	Ultra-sensitive	PCR bias
RNA-seq	Transcriptome	Fusions; isoforms	Functional insight	RNA instability
TCR/BCR	Immune receptors	Immuno-oncology	Immune profiling	Complex interpretation

- *Sequencing Technologies and Platform-Specific Considerations*

Modern sequencing platforms differ significantly in their underlying chemistry, read length, accuracy, and throughput (Metzker, 2010), as summarized in Table 2. Short-read platforms provide high accuracy and throughput but are limited in resolving complex genomic regions. Long-read platforms, including those developed by Pacific Biosciences and Oxford Nanopore Technologies, offer the ability to generate reads spanning kilobases to megabases (Logsdon et al., 2020).

Recent technological advancements have significantly improved the performance of long-read sequencing. High-fidelity (HiFi) sequencing has reduced error rates, while nanopore sequencing enables real-time data acquisition and direct detection of base modifications. Ultra-long reads allow resolution of repetitive regions and structural variants that are inaccessible to short-read methods.

Hybrid sequencing approaches that combine short- and long-read data have emerged as a practical solution, leveraging the high accuracy of short reads and the structural resolution of long reads (Wenger et al., 2019). These approaches are increasingly used in cancer genomics to achieve comprehensive and accurate characterization of tumor genomes.

As a result, platform selection is increasingly guided not only by technical specifications but by the specific clinical questions being addressed.

- *Demultiplexing, Bioinformatics, and Clinical Interpretation*

Following sequencing, demultiplexing assigns reads to individual samples based on barcode sequences (Goodwin et al., 2016). Errors in this step can lead to cross-sample contamination and misinterpretation of results, particularly in multiplexed clinical workflows.

Downstream bioinformatics analysis includes alignment, variant calling, structural variant detection, fusion identification, and methylation analysis. Long-read sequencing introduces additional complexity due to its distinct error profiles and data characteristics, requiring specialized algorithms and computational pipelines (Amarasinghe et al., 2020).

Interpretation of sequencing results is a critical step in clinical diagnostics (Heitzer et al., 2019). The identification of actionable mutations, gene fusions, and epigenetic alterations must be contextualized within clinical frameworks to inform treatment decisions. Increasingly, machine learning approaches are being applied to integrate multi-omic data and improve diagnostic accuracy (Shafin et al., 2020).

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may155

Ultimately, the effectiveness of NGS in cancer diagnostics depends on the seamless integration of laboratory workflows, computational analysis, and clinical interpretation. Long-read sequencing enhances each of these stages but also introduces new challenges that must be addressed to realize its full clinical potential.

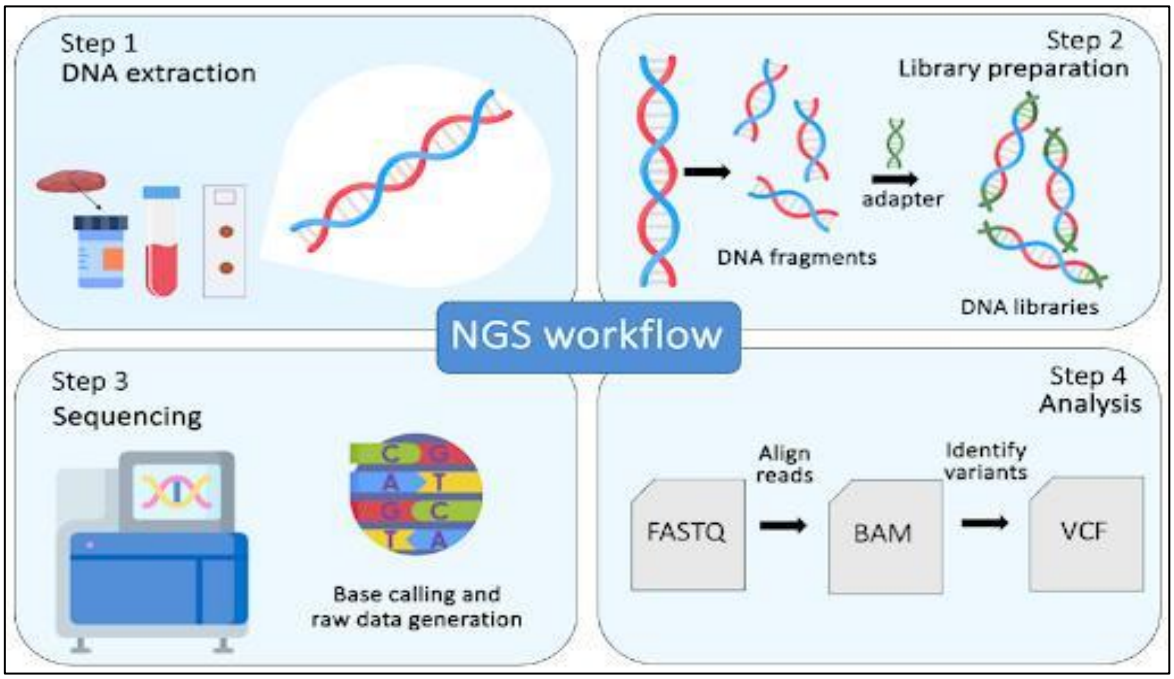

Fig 2 End-to-End Workflow of Next-Generation Sequencing in Cancer Diagnostics, Illustrating Key Steps from Nucleic Acid Extraction to Sequencing, Demultiplexing, Bioinformatic Analysis, and Clinical Interpretation.

Table 2 Comparison of Sequencing Platforms in Cancer Diagnostics

Feature	**Short-Read (Illumina)**	**Long-Read (PacBio / ONT)**
Read length	50–300 bp	10 kb – >1 Mb
Accuracy	Very high	High / Moderate
Throughput	Very high	Moderate
Structural variants	Limited	Excellent
Gene fusions	Partial	Full-length
Epigenetics	Indirect	Direct
Repetitive regions	Poor	High
Phasing	Limited	Strong
Sample requirement	Low input OK	High integrity
Clinical maturity	Established	Emerging

IV. CORE APPLICATIONS OF LONG-READ SEQUENCING IN CANCER DIAGNOSTICS

Long-read sequencing technologies have fundamentally expanded the scope of cancer diagnostics by enabling direct interrogation of genomic and epigenomic features that are poorly resolved by short-read sequencing. Tumor genomes are characterized by structural complexity, clonal heterogeneity, and dynamic regulatory changes, all of which require sequencing approaches capable of capturing long-range genomic context. By generating reads that span kilobases to megabases, long-read platforms provide a more

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may155

complete representation of tumor architecture and enable more accurate detection of clinically relevant alterations, as summarized in Figure 3.

➢ *Structural Variants and Copy Number Alterations*

Structural variants (SVs), including insertions, deletions, inversions, duplications, and translocations, represent a major class of genomic alterations in cancer (Logsdon et al., 2020; Chaisson et al., 2019). These variants can disrupt tumor suppressor genes, activate oncogenes, or alter regulatory landscapes. However, their detection using short-read sequencing is inherently limited, as short fragments often fail to span variant breakpoints, requiring indirect inference based on discordant read pairs or coverage shifts (Amarasinghe et al., 2020).

Long-read sequencing overcomes these limitations by directly spanning SV breakpoints, allowing precise mapping of complex rearrangements (Logsdon et al., 2020; O'Neill et al., 2024). This is particularly important in cancers exhibiting chromothripsis, where chromosomes undergo extensive fragmentation and reassembly, or in the formation of extrachromosomal DNA (ecDNA), which can drive oncogene amplification and therapeutic resistance (O'Neill et al., 2024). By capturing these events within individual reads, long-read sequencing reduces ambiguity and improves confidence in variant detection.

In addition, long-read sequencing enhances copy number variation (CNV) analysis by enabling more uniform coverage across repetitive regions and facilitating phasing of duplicated segments (Chaisson et al., 2019). This allows more accurate identification of clinically relevant alterations such as amplification of oncogenes and deletion of tumor suppressors, both of which are critical for diagnosis, prognosis, and treatment selection.

➢ *Fusions and Transcript Isoforms*

Beyond large-scale genomic rearrangements, alterations at the transcript level provide an additional layer of complexity that is critical for understanding tumor behavior. Gene fusions are among the most clinically actionable genomic alterations in oncology and serve as diagnostic markers and therapeutic targets in multiple cancer types (Dorney et al., 2023). While short-read RNA sequencing can detect fusion events, it often fails to reconstruct full-length fusion transcripts, particularly in cases involving complex breakpoints or multiple isoforms (Amarasinghe et al., 2020).

Long-read RNA sequencing enables direct sequencing of full-length transcripts, providing an unambiguous view of fusion architecture (Dorney et al., 2023). This includes precise identification of breakpoint locations, exon usage, and transcript structure, allowing improved differentiation between functional and non-functional fusion events. Such resolution is critical for determining the clinical relevance of detected fusions and for guiding targeted therapies.

Beyond fusion detection, long-read sequencing reveals extensive transcript isoform diversity. Alternative splicing events can generate protein variants that influence tumor progression, immune evasion, and therapeutic resistance (van Dijk et al., 2023). By preserving full transcript integrity, long-read sequencing provides a more comprehensive understanding of transcriptional regulation in cancer. For example, BCR–ABL1 fusions in chronic myeloid leukemia represent a classic case where accurate fusion detection directly informs targeted therapy.

➢ *Epigenetic Profiling and Regulatory Landscapes*

While transcriptomic changes reflect functional outputs, epigenetic regulation provides an upstream layer of control that shapes gene expression patterns. Epigenetic alterations, including DNA methylation and chromatin remodeling, play a central role in cancer by regulating gene expression without altering the underlying DNA sequence (Lee et al., 2022). Conventional methods for studying epigenetics often rely on chemical conversion or enrichment strategies, which can introduce bias and limit resolution.

Long-read sequencing technologies, particularly nanopore-based platforms, enable direct detection of DNA methylation and other base modifications from native DNA molecules. This allows simultaneous analysis of genetic and epigenetic information within the same read, preserving long-range context and enabling allele-specific epigenetic profiling.

In cancer diagnostics, methylation patterns are increasingly used for tumor classification and early detection (Tsui et al., 2025). Long-read sequencing enhances these applications by capturing extended methylation patterns across genomic regions, improving the ability to distinguish tumor types and identify tissue of origin. Furthermore, integration of epigenetic and genetic information within single reads provides insight into regulatory mechanisms driving tumor progression. These molecular insights are increasingly being translated into minimally invasive approaches, most notably through liquid biopsy.

➢ *Liquid Biopsy, cfDNA, and Fragmentomics*

Liquid biopsy has emerged as a minimally invasive approach for cancer detection and monitoring, primarily through analysis of circulating tumor DNA (ctDNA), a subset of cell-free DNA (cfDNA) (Wan et al., 2017; Heitzer et al., 2019). However, cfDNA is typically highly fragmented and present at low concentrations, posing challenges for sequencing technologies.

Long-read sequencing introduces new opportunities in this domain by enabling analysis of fragmentomic features, including fragment length distributions, end motifs, and nucleosome positioning patterns (Tsui et al., 2025). These features can help distinguish tumor-derived DNA from normal background and provide additional diagnostic information beyond sequence variation alone.

In addition, long-read platforms enable detection of structural variants and methylation patterns within cfDNA, which are difficult to capture using short-read approaches (Tan et al., 2025). Although current limitations in input requirements and compatibility with fragmented DNA

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may155

remain, ongoing improvements in library preparation and sequencing sensitivity are expected to enhance the applicability of long-read sequencing in liquid biopsy workflows. Taken together, these advances converge in the broader framework of precision oncology, where multi-layered data integration is essential.

➢ *Precision Oncology and Integrated Tumor Profiling*

Precision oncology relies on comprehensive molecular characterization of tumors to guide treatment decisions (Nurk et al., 2022). This requires integration of genomic, transcriptomic, and epigenetic data to identify actionable alterations and predict therapeutic response.

Long-read sequencing is uniquely positioned to support this integration by providing multi-dimensional information within single reads (Logsdon et al., 2020). For example, phased variants can reveal allele-specific expression, while combined genetic and epigenetic data can inform regulatory mechanisms underlying tumor behavior.

Comprehensive sequencing approaches that integrate whole-genome and transcriptome data have demonstrated improved diagnostic yield in complex cases, such as cancers of unknown primary origin (Rebello et al., 2025). Long-read sequencing further enhances these approaches by resolving complex genomic features that may otherwise remain undetected, thereby improving clinical decision-making and enabling more personalized treatment strategies (Table 3).

Table 3 Clinical Applications of Long-Read Sequencing in Cancer Diagnostics

Application	What is Detected	Clinical Impact	Why Long-Read Matters
Structural variants	Large rearrangements, CNVs	Diagnosis, prognosis	Direct breakpoint resolution
Gene fusions	Fusion transcripts	Targeted therapy selection	Full-length transcript detection
Epigenetics	DNA methylation	Tumor classification	Native DNA analysis
Liquid biopsy	ctDNA, fragmentomics	Early detection, monitoring	Multi-feature detection
Phasing	Allele-specific variants	Drug response prediction	Long-range haplotypes
Transcript isoforms	Alternative splicing	Functional insights	Full-length RNA sequencing

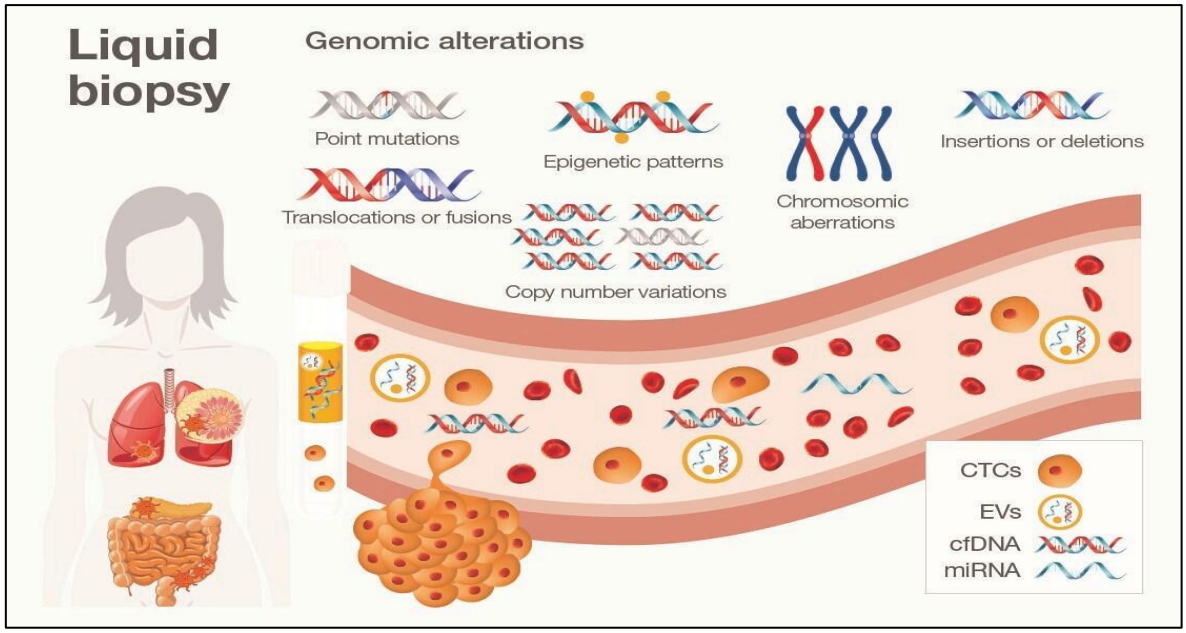

Fig 3 Key Applications of Long-Read Sequencing in Cancer Diagnostics, Including Structural Variant Detection, Gene Fusion Analysis, Epigenetic Profiling, Liquid Biopsy, and Precision Oncology

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may155

V. CHALLENGES AND BARRIERS TO CLINICAL IMPLEMENTATION

Despite its transformative potential, long-read sequencing faces several technical, analytical, and clinical challenges that currently limit its widespread adoption in oncology diagnostics.

A primary barrier remains cost. While sequencing costs have declined substantially, long-read platforms continue to be more expensive than short-read technologies on a per-sample basis (Szakállas et al., 2024). In clinical oncology, where high-throughput testing and cost efficiency are essential, the added value of improved structural resolution must be weighed against economic constraints. This is particularly relevant for routine diagnostic workflows such as targeted panel testing, where short-read approaches remain sufficient for many use cases. Despite the advantages, long-read sequencing does not universally replace short-read approaches, particularly in high-throughput clinical settings where cost, turnaround time, and sequencing depth remain critical considerations.

Sample quality presents a more fundamental limitation. Long-read sequencing relies on high-molecular-weight DNA to achieve optimal performance, yet many clinically relevant samples—particularly formalin-fixed paraffin-embedded (FFPE) tissues—are degraded and chemically modified (Heitzer et al., 2019). This reduces read length and can introduce sequencing artifacts, limiting the ability to resolve structural variants. Similarly, circulating cell-free DNA (cfDNA), which underpins liquid biopsy applications, is inherently fragmented and present at low abundance (Wan et al., 2017). These characteristics pose significant challenges for current long-read library preparation protocols and reduce sensitivity for detecting tumor-derived signals. In practice, successful implementation of sequencing in oncology increasingly depends on the ability to adapt workflows to challenging sample types, rather than relying on ideal input material. This underscores the importance of continued innovation in library preparation and sample processing strategies tailored to clinically relevant specimens.

Computational complexity represents another major barrier. Long-read sequencing generates large and information-rich datasets that require specialized bioinformatics pipelines (Amarasinghe et al., 2020). Error profiles differ from those of short-read sequencing, necessitating tailored approaches for alignment, variant calling, and methylation analysis. In addition, standardized pipelines for clinical use are still evolving, and variability between analytical methods can impact reproducibility and interpretation—an important consideration in regulated diagnostic environments.

Interpretation of results remains a critical bottleneck. Long-read sequencing enables detection of complex genomic and epigenetic features, many of which lack well-established clinical annotations. This increases the likelihood of identifying variants of uncertain significance, particularly in heterogeneous tumors (Szakállas et al., 2024). The clinical utility of such findings depends on the availability of robust reference datasets and functional validation, both of which are still developing.

In parallel with advances in sequencing technologies, innovations in library preparation and workflow automation are beginning to address cost and scalability barriers. For example, bulk RNA barcoding and sequencing approaches such as BRB-seq enable early multiplexing of samples, substantially reducing reagent usage and per-sample cost (Alpern et al., 2019). When combined with automated liquid handling platforms for nucleic acid extraction and library preparation, these strategies can further improve throughput and reproducibility while minimizing hands-on time (Bredenoord et al., 2021; Muir et al., 2020). Although originally developed for transcriptomic applications, such approaches highlight the broader potential of workflow-level optimization to complement advances in sequencing platforms, particularly in high-throughput or resource-limited settings.

Importantly, many of these challenges are not static. Advances in library preparation for low-input and degraded samples, improvements in sequencing accuracy, and the development of standardized bioinformatics frameworks are actively addressing current limitations (van Dijk et al., 2023). As these solutions mature, the barriers to clinical implementation are expected to diminish, paving the way for broader adoption.

VI. FUTURE PERSPECTIVES AND CLINICAL TRANSLATION

The next phase of sequencing in oncology is likely to be defined not by incremental improvements in individual technologies, but by the integration of multiple data modalities into cohesive diagnostic frameworks. Long-read sequencing is positioned to play a central role in this transition due to its ability to capture genomic, transcriptomic, and epigenetic information within single molecules (Nurk et al., 2022).

In the near term, one of the most impactful developments will be the refinement of long-read sequencing for challenging clinical samples. Advances in library preparation protocols tailored for fragmented DNA, including cfDNA and FFPE-derived material, are expected to significantly expand the applicability of long-read platforms in routine diagnostics (Tan et al., 2025). This will be particularly important for liquid biopsy applications, where improved sensitivity and resolution could enable earlier cancer detection and more accurate monitoring of disease progression.

Another key direction is the integration of long-read sequencing with multi-omics approaches. Combining genomic, transcriptomic, and epigenetic data provides a more comprehensive view of tumor biology and has the potential to improve diagnostic accuracy and therapeutic stratification (Logsdon et al., 2020). Long-read sequencing facilitates this

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may155

integration by preserving long-range context and enabling simultaneous analysis of multiple layers of information.

Looking ahead, reductions in sequencing cost will likely depend not only on improvements in sequencing hardware but also on innovations in upstream workflows. Approaches such as early multiplexing strategies (e.g., BRB-seq) and increasing adoption of laboratory automation are expected to play an important role in enabling scalable and cost-effective sequencing in clinical and research settings.

Artificial intelligence and machine learning are expected to play an increasingly important role in interpreting the complex datasets generated by long-read sequencing (Shafin et al., 2020). These approaches can identify patterns across multi-omic data that are not readily apparent through conventional analysis, supporting applications such as tumor classification, prediction of treatment response, and identification of novel biomarkers (Tsui et al., 2025).

In addition, real-time sequencing capabilities—particularly those enabled by nanopore platforms—introduce the possibility of rapid, near-patient diagnostics (van Dijk et al., 2023). This could be especially valuable in clinical scenarios requiring timely decision-making, such as aggressive malignancies or treatment-resistant disease (Amarasinghe et al., 2020).

Over the next several years, the clinical impact of long-read sequencing will depend on the convergence of technological innovation, cost reduction, and standardization of analytical workflows. As these factors align, long-read sequencing is likely to transition from a complementary research tool to a core component of precision oncology, enabling more comprehensive, accurate, and actionable cancer diagnostics.

VII. CONCLUSIONS

Long-read sequencing is steadily reshaping how we approach cancer diagnostics. While short-read technologies have laid the foundation for modern genomic profiling, their limitations in resolving structural complexity have become increasingly apparent. In this context, long-read platforms offer a more complete view of tumor genomes, enabling clearer identification of structural variants, gene fusions, and regulatory features that are often difficult to capture using conventional approaches.

What is becoming equally evident is that the impact of these technologies depends not only on sequencing itself, but on the broader ecosystem in which they are applied. Improvements in library preparation, the growing use of laboratory automation, and advances in bioinformatics are all contributing to more robust and scalable workflows. At the same time, artificial intelligence is beginning to play a meaningful role in making sense of increasingly complex datasets, helping translate sequencing results into clinically actionable insights.

Despite ongoing challenges related to cost, sample quality, and standardization, the overall trajectory is clear. As these barriers continue to be addressed, long-read sequencing is likely to move from a complementary tool to a more central component of precision oncology. In the coming years, its ability to provide a more complete and integrated view of tumor biology may ultimately help enable more accurate, personalized, and clinically meaningful cancer diagnostics.

REFERENCES

[1]. Alpern, D., Gardeux, V., Russeil, J., et al. (2019). BRB-seq: ultra-affordable high-throughput transcriptomics enabled by bulk RNA barcoding and sequencing. *Genome Biology*, 20, 71.
[2]. Amarasinghe, S.L., Su, S., Dong, X., et al. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biology*, 21, 30.
[3]. Bredenoord, A.L., et al. (2021). Automation in genomics and its impact on reproducibility. *Nature Methods*, 18, 104–110.
[4]. Chaisson, M.J.P., Sanders, A.D., Zhao, X., et al. (2019). Multi-platform discovery of haplotype-resolved structural variation. *Nature Communications*, 10, 1784.
[5]. Dorney, K., et al. (2023). Detection of gene fusions using long-read sequencing. *Briefings in Bioinformatics*, 24, bbac519.
[6]. Gansauge, M.T., and Meyer, M. (2020). Single-stranded DNA library preparation for highly degraded DNA. *Nature Protocols*, 15, 2041–2054.
[7]. Goodwin, S., McPherson, J.D., and McCombie, W.R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17, 333–351.
[8]. Heitzer, E., Haque, I.S., Roberts, C.E.S., and Speicher, M.R. (2019). Current and future perspectives of liquid biopsies in genomics-driven oncology. *Nature Reviews Genetics*, 20, 71–88.
[9]. Jansson, L., et al. (2024). Assessment of DNA quality for whole genome library preparation. *Analytical Biochemistry*, 688, 115123.
[10]. Jennings, L.J., Arcila, M.E., Corless, C., et al. (2020). Guidelines for validation of next-generation sequencing–based oncology panels. *The Journal of Molecular Diagnostics*, 22, 341–365.
[11]. Lee, H., et al. (2022). Nanopore sequencing and methylation analysis. *Nature Biotechnology*.
[12]. Logsdon, G.A., Vollger, M.R., and Eichler, E.E. (2020). Long-read human genome sequencing and its applications. *Nature Reviews Genetics*, 21, 597–614.
[13]. Metzker, M.L. (2010). Sequencing technologies — the next generation. *Nature Reviews Genetics*, 11, 31–46.
[14]. Muir, P., et al. (2020). The real cost of sequencing: scaling computation to keep pace with data generation. *Genome Biology*, 21, 53.
[15]. Nurk, S., Koren, S., Rhie, A., et al. (2022). The complete sequence of a human genome. *Science*, 376, 44–53.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may155

[16]. O'Neill, K., et al. (2024). Long-read sequencing of cancer genomes reveals structural complexity. *Cell Genomics*.
[17]. Pignatta, S., et al. (2025). Comparative analysis of library preparation approaches for degraded samples. *Scientific Reports*, 15, 12992.
[18]. Rebello, R., et al. (2025). Whole genome and transcriptome sequencing improves cancer diagnostics. *Nature Communications.*
[19]. Robins, H. (2013). Immunosequencing: applications of immune repertoire deep sequencing. *Current Opinion in Immunology*, 25, 646–652.
[20]. Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *PNAS*, 74, 5463–5467.
[21]. Shafin, K., Pesout, T., Lorig-Roach, R., et al. (2020). Nanopore sequencing and deep learning variant calling. *Nature Biotechnology*.
[22]. Sung, H., Ferlay, J., Siegel, R.L., et al. (2021). Global cancer statistics 2020. *CA: A Cancer Journal for Clinicians*, 71, 209–249.
[23]. Szakállas, N., et al. (2024). Long-read sequencing in clinical genomics. *Pathology & Oncology Research*.
[24]. Tsui, D.W.Y., et al. (2025). Fragmentomics and cancer detection using cfDNA. *Cancer Cell*.
[25]. van Dijk, E.L., Jaszczyszyn, Y., and Thermes, C. (2023). The long-read sequencing revolution. *Trends in Genetics*, 39, 3–16.
[26]. Wan, J.C.M., Massie, C., Garcia-Corbacho, J., et al. (2017). Liquid biopsies in cancer. *Nature Reviews Cancer*, 17, 223–238.
[27]. Wenger, A.M., Peluso, P., Rowell, W.J., et al. (2019). Accurate circular consensus long-read sequencing. *Nature Biotechnology*.