Clinica Chimica Acta 551 (2023) 117568

Contents lists available at ScienceDirect

Clinica Chimica Acta

journal homepage: www.elsevier.com/locate/cca

Review

Next-Generation sequencing transforming clinical practice and precision medicine

Deepali Yadav [a,b], Bhagyashri Patil-Takbhate [a], Anil Khandagale [b], Jitendra Bhawalkar [c], Srikanth Tripathy [a,*], Priyanka Khopkar-Kale [a,*]

[a] Central Research Facility, Dr. D.Y Patil Medical College, Hospital & Research Centre, Dr. D. Y. Patil Vidyapeeth, Pimpri Pune 411018, India
[b] Department of Biotechnology, Dr. D. Y. Patil Arts Science and Commerce College, Pimpri Pune 411018, India
[c] Department of Community Medicine, Dr. D.Y Patil Medical College, Hospital & Research Centre, Dr. D. Y. Patil Vidyapeeth, Pimpri Pune 411018, India

ARTICLE INFO

Keywords:
Next- generation sequencing
Precision medicine
Clinical diagnosis
Genetic disorders

ABSTRACT

Next-generation sequencing (NGS) has revolutionized the field of genomics and is rapidly transforming clinical diagnosis and precision medicine. This advanced sequencing technology enables the rapid and cost-effective analysis of large-scale genomic data, allowing comprehensive exploration of the genetic landscape of diseases. In clinical diagnosis, NGS has proven to be a powerful tool for identifying disease-causing variants, enabling accurate and early detection of genetic disorders. Additionally, NGS facilitates the identification of novel disease-associated genes and variants, aiding in the development of targeted therapies and personalized treatment strategies. NGS greatly benefits precision medicine by enhancing our understanding of disease mechanisms and enabling the identification of specific molecular markers for disease subtypes, thus enabling tailored medical interventions based on individual characteristics. Furthermore, NGS contributes to the development of non-invasive diagnostic approaches, such as liquid biopsies, which can monitor disease progression and treatment response. The potential of NGS in clinical diagnosis and precision medicine is vast, yet challenges persist in data analysis, interpretation, and protocol standardization. This review highlights NGS applications in disease diagnosis, prognosis, and personalized treatment strategies, while also addressing challenges and future prospects in fully harnessing genomic potential within clinical practice.

1. Introduction

DNA sequencing has undergone significant advancements since the discovery of DNA sequencing techniques like Maxam-Gilbert and Sanger sequencing [1] Genome sequencing has achieved widespread adoption, becoming a routine practice on a large scale since the successful completion of the first human genome sequence in 2003 [2]. The advent of NGS has overcome the limitations of traditional Sanger sequencing and facilitated high-throughput, affordable sequencing of genomes [3]. NGS platforms, such as Illumina and Roche pyrosequencing, have spurred research in genomics, transcriptomics, epigenomics, and metagenomics by providing quantitative sequencing capabilities. Ongoing advancements in NGS technologies, such as improved chemistry, longer read lengths, and deeper genome exploration, have paved the way for "third-generation sequencing" platforms like Pacific Biosciences' SMRT sequencing and Oxford Nanopore Technologies' nanopore sequencing. These technologies offer extended sequencing capabilities, real-time analysis of DNA molecules, and the ability to capture structural variations and epigenetic modifications. Third-generation sequencing holds promise for expanding our understanding of the genome. The field of genomics has rapidly evolved, with the introduction of next-generation sequencing (NGS) platforms revolutionizing DNA sequencing [4]. (Fig. 1) NGS has democratized genome sequencing by progressing alongside the development of bioinformatics tools. Research groups, regardless of their size, can now generate draft genome sequences for any organism of interest. NGS encompasses various applications, such as transcriptome sequencing (RNA-seq), whole-exome sequencing (WES), targeted or candidate gene sequencing (CGS), and methylation sequencing (MeS) [5]. RNA-seq, for instance, enables comprehensive identification of transcriptional activities, including coding and non-coding transcripts, with higher precision and sensitivity compared to microarrays. This makes it a powerful tool for analyzing diverse samples. WES, in contrast, provides comprehensive coverage of over 95 % of human exons, focusing on the protein-coding regions of the genome.

* Corresponding authors.
E-mail addresses: director.medicalresearch@dpu.edu.in (S. Tripathy), priyanka.kale@dpu.edu.in, priyankakhopkar@gmail.com (P. Khopkar-Kale).

https://doi.org/10.1016/j.cca.2023.117568
Received 8 July 2023; Received in revised form 27 September 2023; Accepted 27 September 2023
Available online 13 October 2023

This approach is particularly cost-effective and practical for investigating coding variants, including Single Nucleotide Polymorphisms (SNPs), in human genetics and disease studies. While WES has its advantages, Whole-Genome Sequencing (WGS) offers a more extensive dataset with better uniformity of read coverage across disease-associated variants. It can reveal polymorphisms outside coding regions and genomic rearrangements. Therefore, the choice between WGS and WES depends on specific research or clinical objectives.

The review runs through rapid progresses made in the NGS technology, combined with concurrent bioinformatics tool development, has revolutionized genome sequencing and empowered precision medicine. NGS has become invaluable in diagnosing rare diseases, cancers, and infectious diseases, providing essential genetic insights, and improving patient outcomes through personalized medicine approaches.

2. Overview of next-generation sequencing platforms

The origin of NGS is rooted in the historical development of DNA sequencing techniques. While Sanger sequencing, with its pioneering dideoxy chain termination method, served as the cornerstone of DNA sequencing for nearly three decades, its throughput limitations and cost constraints prompted the search for innovative approaches. The genesis of sequencing DNA was first laid down by Sanger sequencing which was, the first DNA sequencing method published in 1977 it was the primary sequencing technology between 1977 and 2005 which relies on the principle of sequencing by synthesis approach which employs a radioactively labeled DNA strand complementary to the template strand using the dideoxy chain termination technique the fragments are further analyzed through gel electrophoresis [1]Sanger sequencing is still been considered as the gold standard for DNA sequencing due to its ability to generate high-quality DNA sequences of relatively long lengths (500–1000 bp) with high accuracy. Despite this Sanger sequencing has limitations in its throughput capacity and cost however it remains valuable for validating DNA sequences and target resequencing. Improvements to Sanger's chain-termination method led to the introduction of the first automatic sequencing machine, AB370, by Applied Biosystems in 1987. This advancement included the adoption of fluorometric-based detection and capillary-based electrophoresis. The AB370 automatic sequencing machine achieved 96-base detection, processed 500 K bases/day, and had read lengths of up to 600 bases. The subsequent AB3730xl model increased capacity to 2.88 M bases/day with read lengths up to 900 bases, revolutionizing DNA sequencing with greater automation and efficiency. The introduction of pyrosequencing technology by 454 Life Sciences in 2005 which is now called (Roche 454) marked the beginning of the next-generation sequencing (NGS) revolution [6]. These second-generation sequencing technologies utilized luminescence-based detection of pyrophosphate synthesis to determine the nucleotide sequence, enabling parallel sequencing reactions and significantly increasing the amount of DNA sequenced in a single run generating both short (50–400 bp) and long (1–100 kb) reads [7]. Roche 454 technologies' latest instrument GS FLX + system, can produce reads up to 1000 base pairs in length and generate around 1 million reads per run. These systems demonstrated improved performance in terms of throughput, accuracy, and cost compared to Sanger sequencing. However, a primary issue with this method is the detection of errors in the form of insertions and deletions, primarily caused by homopolymer regions. These errors occur when the size of homopolymers is incorrectly determined based on the intensity of emitted light during pyrosequencing. If the signal intensity is too high or too low, it can lead to the miscounting of nucleotides, resulting in inaccuracies in nucleotide identification whilst sequencing. Despite variations in biochemistry and arrays across different platforms, the workflows employed generally adhere to a similar set of steps. These steps typically entail DNA extraction: involving the isolation of DNA from the target sample; Library preparation: encompassing the fragmentation of DNA through mechanical or enzymatic means, the addition of adaptors and barcodes/indexes, and subsequent amplification; Template preparation: achieved via either bridge amplification or emulsion PCR techniques; Automated Sequencing: wherein the prepared templates are subjected to sequencing using automated platform (Fig. 2) [8]. Over time, the NGS platforms evolved to offer advantages in read length, accuracy, applications, consumables manpower requirements, and informatics infrastructure. (SOLiD) system [Applied Biosystems (Waltham, MA, USA)], for instance, employed a two-base sequencing method based on ligation sequencing and achieved high accuracy [8]. Its read length and data output were progressively improved, culminating in the release of the SOLiD 5500xl system which generated an 85-bp read length, 99.99 % accuracy, and a data output of 30 G per run. The strength of the ABI/

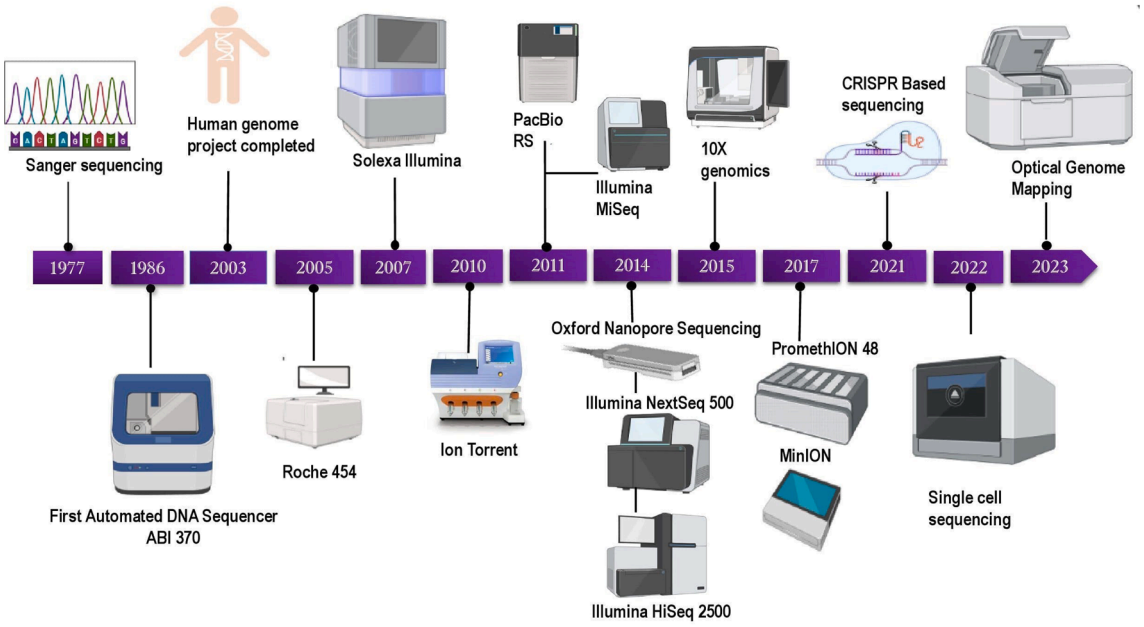

Fig. 1. Developmental Timelines of Next-Generation Sequencing Technologies Figure Legend: Evolution of Next-generation sequencing technologies and platforms, from Sanger sequencing to emerging technologies. *Abbreviations*: Deoxyribonucleic acid (DNA), Clustered regularly interspaced short palindromic repeats (CRISPR).

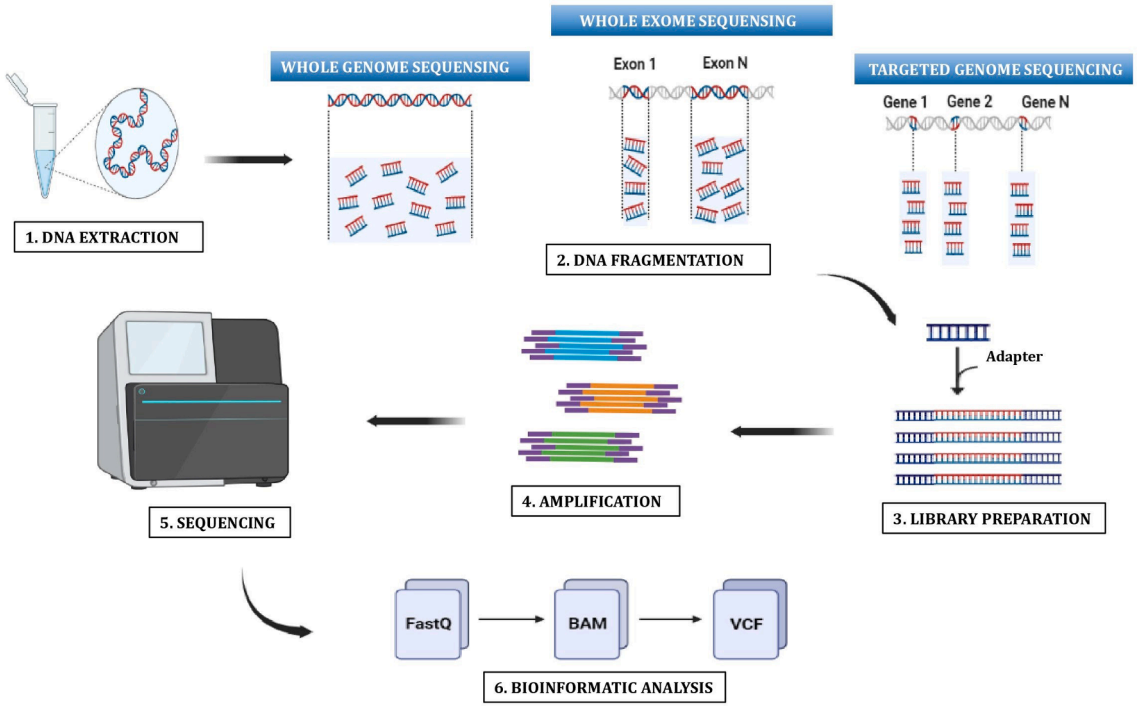

Fig. 2. Next-generation sequencing workflow Fig. 2 Legend: Sequential NGS workflow, including DNA/RNA extraction, library preparation, template amplification, sequencing, Bioinformatic data analysis & interpretation *Abbreviations*: BAM -Binary Alignment Map, VCF -Variant call format.

SOLiD platform is high accuracy because each base is read twice while the drawback is the relatively short reads and long run times, errors of sequencing in this technology are due to signal to noise ratio during the ligation cycle which causes error identification of bases. The main type of error is substitution. Despite its strengths in certain applications, SOLiD's short read length and limited scope for resequencing remained notable limitations. The applications of SOLiD encompassed whole genome resequencing, targeted resequencing, transcriptome research, and epigenomic studies.

Another second-generation short-read sequencing technology includes the Illumina platform which utilizes reversible terminator chemistry and an array-based DNA sequencing-by-synthesis approach [9]. The system includes various sequencing machines such as the NextSeq 500, HiSeq series (2500, 3000, 4000), HiSeq X series (five and ten), and the compact laboratory sequencer called MiSeq. The Illumina system offers mid to high sequencing data output per run (120–1500 Gb) and the MiSeq provides 0.3 to 15 Gb output suitable for targeted sequencing. It boasts advantages such as cost-effectiveness, shorter run times, and low error rates. The system produces read lengths of approximately 100 base pairs and can generate 600 GB of sequencing data per run. However, it has some limitations viz. its sensitivity to sample loading, as sample overloading can lead to issues like overlapping clusters thereby reducing sequencing quality. This technology typically exhibits an overall error rate of approximately 1 %, with nucleotide substitutions being the most prevalent type of error. These errors primarily arise from difficulties in accurately identifying the incorporated nucleotides during the sequencing process. These platforms support Whole-Genome Sequencing (WGS), Target Gene Sequencing (TGS), gene expression profiling, miRNA and sRNA analysis, and metagenomic sequencing. The NextSeq 550 Series and NextSeq 1000 & 2000 have additional applications such as exome sequencing, chip-seq analysis, methylation sequencing, and cell-free sequencing [10]. Illumina's innovative synthetic long-read method using TruSeq technology enhances *de novo* assembly and resolves complex, highly repetitive transposable elements improving NGS quality. This bridge amplification method pioneered by Solexa, later acquired by Illumina, stands out as a remarkable and highly successful second-generation sequencing technique [11]. Introduced in 2011, Ion Torrent is a sequencing method based on Semiconductor sequencing [13]. The mechanism of Ion Torrent sequencing involves detecting changes in pH that occur during DNA synthesis. Each nucleotide incorporation releases a hydrogen ion, which is detected by a semiconductor sensor. This technology eliminates the need for fluorescent labels or optics, simplifying the sequencing process. Over time, Ion Torrent sequencing has achieved improvements in read length, accuracy, and data output. Initially, the read lengths were relatively short, typically around 100–200 base pairs. However, with advancements in technology, read lengths of over 400 base pairs have been achieved. There are four Ion Torrent instruments available. The GeneXus system stands out for its ability to generate data analysis reports within a single day, employing an automated workflow with only two touch points. It is an economical option suitable for low sample inputs and can be easily installed in a laboratory or even in-house use, irrespective of the level of expertise in Next-Generation Sequencing (NGS). The Ion Gene Studio S5 systems, on the other hand, offer efficient, scalable, and cost-effective targeted sequencing capabilities. The major advantage of this sequencing technology is that it offers longer read lengths and faster sequencing times, typically 2 to 8 h. However, it faces challenges in interpreting homopolymer sequences, particularly those exceeding 6 base pairs, leading to insertion and deletion errors (indels) at a rate of around 1 %. Nevertheless, its advantages include longer read lengths, flexible workflow, shorter turnaround times, and cost-effectiveness compared to other platforms which override its limitations.

While second-generation sequencing (2GS) technologies have significantly advanced DNA sequencing, they do come with certain drawbacks. These drawbacks include limitations in read length, difficulties in sequencing repetitive regions, challenges in resolving complex genomic structures, and biases introduced during library preparation. To overcome these limitations, the need for third-generation sequencing (3GS) technologies emerged [12]. Third-generation sequencing (3GS) technologies, such as nanopore sequencing, Single-Molecule Real-Time (SMRT) sequencing, and synthetic nanopore sequencing, directly

sequence individual DNA molecules without amplification or fragmentation. They offer long read lengths and the ability to capture complex genomic regions. Each technology operates on distinct principles, including nanopore-based detection, real-time fluorescence signals, or synthetic nanopores [13]. The PacBio RSII platform, based on single-molecule real-time sequencing technology, enables the synthesis of DNA using fluorescently labeled nucleotides. It utilizes tiny wells with immobilized DNA polymerase to generate long DNA fragments of 20 kb or more in just a few hours. However, the platform has low throughput, high per-sample cost, and a larger instrument footprint, making it suitable for sequencing core facilities focused on high-quality finished genomes. The PacBio method has a higher error rate (11 %-15 %) compared to short-read technologies, necessitating error correction algorithms during assembly. Pacific Biosciences has released a more affordable and higher throughput version of RSII called Sequel. The technology relies on Zero-Mode Wavelength nanostructures (ZMW) to measure fluorescently labeled nucleotide incorporation in real time, producing reads up to 10 kilobases (kb) in length [14]. Oxford Nanopore Technologies introduced their latest single-molecule sequencing systems in 2012. The MinION Mk1 is a portable DNA and RNA sequencing device, similar in size to a USB drive, that can be directly connected to a computer or laptop. On the other hand, the PromethION system is a compact benchtop device. These systems utilize nanopores, which are protein-based channels capable of detecting different DNA bases. The concept behind nanopore sequencing is that each nucleotide passing through the pore alters the ionic current, which is evaluated in real-time. The shape of the translocating molecule affects the measurements. The main advantage of this approach is the minimal sample preparation required, with read lengths in the kilobase (kb) range. However, optimizing the speed of DNA translocation through the nanopores remains a significant challenge, as it affects both the reliability of current measurements and the reduction of high error rates [15]. (Table 1) consists of a brief comparison of the various sequencing technologies.

3. NGS in clinical diagnosis: A deep dive

Omics-integrated NGS has transformed clinical diagnosis and precision medicine, offering a comprehensive view of an individual's genetic and molecular profile. Integrating genomics, transcriptomics, epigenomics, and proteomics data on the NGS platform is crucial for understanding complex human diseases and is advancing personalized healthcare.

3.1. Genomics

Genomics studies utilizing NGS have revolutionized our understanding of the genetic makeup of living organisms. Various approaches, including whole-genome sequencing, whole-exome sequencing, and targeted sequencing, are employed with this advanced technology. It allows researchers to decipher the complete DNA sequence of an organism quickly and cost-effectively.

3.1.1. Whole genome sequencing (WGS)

WGS has been instrumental in genomics since the completion of the Human Genome Project in 2003. It enables comprehensive analysis of an organism's entire genetic code and has found critical applications in clinical diagnostics and precision medicine. WGS identifies genetic variations, guiding personalized healthcare [16]. Two primary WGS approaches are short-read and long-read sequencing, with the latter excelling in capturing complex genomic rearrangements. Challenges include data volume and cost, but WGS continues to drive genomic advancements and tailored therapies. By sequencing all the DNA in an organism's genome, WGS enables the identification of genetic variations, ranging from single-nucleotide polymorphisms (SNPs) to larger structural changes such as insertions, deletions, and rearrangements. WGS encompasses two sequencing approaches, primarily distinguished by genome size. Large whole-genome sequencing focuses on genomes exceeding 5 Mb, typically eukaryotes, while small whole-genome sequencing targets smaller genomes, often prokaryotes under 5 Mb in size [17]. Short-read sequencing is commonly employed for mutation identification, whereas long-read sequencing proves advantageous for assembling genomes accurately, even in the absence of a reference sequence. The fusion of short and long-read sequencing has proven successful in precisely assembling novel genomes.

3.1.2. Whole exome sequencing

WES specifically targets the exome, the protein-coding regions of the genome, providing a comprehensive analysis of genetic variations in these critical regions. There are two primary types of WES: hybrid capture-based and PCR-based. The former uses biotinylated probes to selectively capture and sequence exonic DNA, while the latter employs Polymerase Chain Reaction (PCR) to amplify and sequence exonic DNA fragments. Despite its advantages, WES has limitations, including limited coverage of non-coding regions and potential ethical dilemmas related to incidental findings [18]. However, its clinical applications have been transformative. It aids in identifying rare and inherited genetic disorders, benefiting patients with unexplained symptoms. In oncology, it contributes to cancer genomics by identifying somatic mutations for personalized cancer treatments. Additionally, in pharmacogenomics, WES informs drug selection and dosing based on genetic variants that influence drug response. It is also useful in prenatal diagnosis, enabling early detection of genetic abnormalities in the developing fetus. As genomics continues to expand, whole exome sequencing remains essential, bridging genetic research and clinical practice, and deepening our understanding of the genetic basis of health and disease.

3.1.3. Targeted sequencing

It involves selectively sequencing specific regions of the genome, such as known disease-associated genes or particular exons, rather than the entire genome. This approach offers several advantages, including cost-effectiveness and faster turnaround times, making it suitable for diagnosing specific genetic disorders efficiently [19]. Targeted sequencing can also provide a higher depth of coverage, increasing the likelihood of detecting low-frequency mutations. However, its main limitation is the narrow focus on predefined genomic regions, potentially missing novel or unexpected variants and Single Nucteotide Polymorphisms (SNPs) outside the targeted regions.

3.2. Transcriptomics

NGS has revolutionized transcriptomics by significantly enhancing our ability to study gene expression and RNA molecules. Transcriptomics focuses on gene expression at the RNA level, offering comprehensive genome-wide information about gene structure and function to unveil the molecular mechanisms involved in specific biological processes. NGS technologies provide high-throughput and cost-effective methods for profiling and analyzing RNA molecules, enabling researchers to gain deep insights into gene expression, alternative splicing, non-coding RNA regulation, and various biological processes and diseases. Different types of transcriptomics strategies include [22].

3.2.1. mRNA-Seq

Messenger RNA sequencing (mRNA-Seq) is a high-throughput molecular technique that quantifies and analyzes messenger RNA (mRNA) molecules within a biological sample. It offers a digital readout of gene expression levels, making it a powerful tool for gene expression profiling, differential gene expression analysis, and the discovery of alternative splicing events. mRNA-Seq data provides precise quantification of gene expression levels and differential gene activity, discover new transcript variants, explore alternative splicing, and track gene expression changes over time across the transcriptome [20,21].

Table 1
Comparison of Next Generation Sequencing Platforms.

Producer	Platform	Amplification method	Sequencing method	Read length(bp)	Output	Run time	Pros	Cons	Applications	Refrences
APPLIED BIOSYSTEM	AB370	PCR -Based	Sanger Sequencing	400–900	1 kb	3 h	Established technology, accurate sequencing	Limited read length, lower throughput compared to NGS	Sanger sequencing, targeted sequencing	[1]
	AB370xL	Emulsion PCR	Capillary Electrophoresis	900	1.9–8.4 Kb	3 h	High accuracy, long read length	Low throughput, labor-intensive	Genetic research, diagnostics	
Roche	Roche 454 GS FLX	Emulsion PCR	Pyrosequencing	700–1,000	700 MB	23 h	Long reads, moderate throughput	Expensive reagents, homopolymer errors	*De novo* sequencing, amplicon sequencing	
Life Technologies	ABI SOLiD 5500xL	PCR -Based	Sequencing By Ligation(SBL)	75	10—15 GB	6–7 days	High accuracy, ability to detect DNA modifications	Short reads, limited coverage depth	Whole-exome sequencing, small RNA sequencing	[2]
Illumina/ Solexa	MiniSeq	PCR -Based	Sequencing By Synthesis(SBS)	2 × 150	Up to 7.5 Gb, 25 Million reads	4—24 h	Compact, user-friendly, cost-effective	Lower throughput compared to larger Illumina platforms	Targeted sequencing, small genome sequencing	
	MiSeq	PCR -Based	Sequencing By Synthesis(SBS)	2 × 150	Up to15Gb, 25 million reads	5–55 h	Versatile, flexible read lengths	Moderate throughput, higher reagent costs	Targeted sequencing, amplicon sequencing	[3]
	NextSeq	PCR -Based	Sequencing By Synthesis(SBS)	2 × 301	Up to120 Gb, 400 million reads	12–30 h	Flexibility, scalability, broad range of applications	Lower throughput compared to some other Illumina platforms	Whole-genome sequencing, targeted sequencing, transcriptomics	
	HiSeq	PCR -Based	Sequencing By Synthesis(SBS)	2 × 150	Up to 1,500 Gb, 5 billion reads	1–6 h	High-throughput, customizable read lengths	Moderate throughput, higher reagent costs	Whole-genome sequencing, RNA-Seq, ChIP-Seq	
	HiSeqX	PCR -Based	Sequencing By Synthesis(SBS)	2 × 150	Up to 1,800 Gb, 6 billion reads	< 3 days	Ultra-high throughput, scalable	Longer run times, higher instrument and maintenance costs	Population genomics, large-scale sequencing projects	
Life Technologies	Ion Torrent PGM	Emulsion PCR	Ion Semiconductor Sequencing	Up to 200	Up to 2 Gb	< 1.5 hors for 100 base runs	Fast turnaround time, low instrument cost	Longer run times, higher instrument and maintenance costs	Targeted sequencing, microbial genomics	[4]
(ThermoFisher)	Proton	Emulsion PCR	Ion Semiconductor Sequencing	Up to 400	Up to 10 Gb	~2 h for 100 base runs	Scalable, cost-effective, moderate throughput	Higher error rates, shorter read lengths	Cancer genomics, clinical research	
	Gene Studio S5	PCR -Based	Ion Semiconductor Sequencing	Up to 200	Up to 15 Gb	2.5 h for 200 bp	Fast turnaround time, ease of use	Higher error rates, shorter read lengths	Targeted sequencing, small genome sequencing	
Pacific Biosciences	PacBio RSII	No Amlification required	Single Molecule Real -Time	>1,000	Up to 55,000 reads, 1 Gb	0.5—4 h	Long reads, ability to detect structural variations	Moderate throughput, limited read length	Structural variant detection, isoform sequencing	[5]
	Sequel	No Amlification required	Single Molecule Real -Time	10,000	7.6 Gb	0.5—6 h	Ultra-high throughput, long reads	Higher error rates, lower throughput compared to short-read NGS	*De novo* assembly, full-length transcript sequencing	

(*continued on next page*)

Table 1 (*continued*)

Producer	Platform	Amplification method	Sequencing method	Read length(bp)	Output	Run time	Pros	Cons	Applications	Refrences
Oxford Nanopore Technologies	MinION	No Amlification required	Nanopore Sequencing	100,000	Up to 30 Gb	72 h	Portable, real-time sequencing, long reads	Higher error rates, lower throughput compared to short-read NGS	Field sequencing, rapid pathogen detection, metagenomics	[6]
	GridION	No Amlification required	Nanopore Sequencing	100,000	Up to 150 Gb	72 h	High-throughput, scalable, long reads	Higher error rates, longer run times	Large-scale sequencing projects, population genomics	
	PromethION	No Amlification required	Nanopore Sequencing	2.3 Mb	Up to 8.6 Tb	64 h	Ultra-high throughput, long reads	Higher error rates, longer run times	Large-scale sequencing projects, population genomics	

Abbreviations – **PCR** (Polymerase Chain Reaction), **Gb** (Gigabyte), **Tb** (Terabyte), **KB** (Kilobase), **MB** – (Megabyte),**DNA** (Deoxy-Ribonucleic Acid).

3.2.2. Single-cell transcriptomics or scRNA-seq

NGS-based scRNA-seq allows for high-throughput profiling of gene expression in thousands of single cells simultaneously. This technology overcomes the limitations of traditional bulk RNA-seq by uncovering cellular heterogeneity, identifying rare cell types, and capturing dynamic gene expression changes within complex tissues or biological systems. The resulting scRNA-seq data offers insights into cellular diversity, allowing researchers to identify distinct cell types, trace developmental trajectories, and uncover rare or novel cell populations[22].

3.2.3. Long Non-Coding RNA (lncRNA) and Small-RNA analysis

Long Non-Coding RNAs (lncRNAs) and small RNAs are emerging as pivotal players in the regulation of gene expression and cellular processes. Analyzing these non-coding RNAs has become crucial for understanding the complexity of the genome. LncRNAs, typically longer than 200 nucleotides, have diverse roles in gene regulation. They can act as scaffolds, guides, or decoys, influencing chromatin structure and modulating transcription. Identifying and characterizing lncRNAs involves techniques like RNA-seq and CLIP-seq (Cross-Linking Immunoprecipitation followed by Sequencing), which help uncover their functions in development, disease, and epigenetic regulation. Small RNAs, on the other hand, include microRNAs (miRNAs), small interfering RNAs (siRNAs), and piwi-interacting RNAs (piRNAs) [23,24]. They exert post-transcriptional gene silencing, critical for gene regulation, and are essential in pathways like RNA interference. Studying small RNAs relies on small RNA-seq and related methods, revealing their involvement in processes such as immune response, development, and pathogen defense. Combined, the analysis of lncRNAs and small RNAs expands our knowledge of RNA-mediated gene regulation, offering promising avenues for therapeutic interventions and a deeper understanding of the intricate molecular mechanisms governing cellular biology [24].

3.2.4. Spatial transcriptomics

Spatial transcriptomics, known as 'spatially-resolved transcriptomics,' is a cutting-edge field that revolutionizes the study of gene expression by preserving spatial context within tissues. This technology has wide-ranging applications, from providing insights into the cellular context within tissues to deciphering complex cell signaling pathways by detecting mRNA associated with cell surface-bound proteins. It also delves into the sub-cellular localization of mRNAs, shedding light on how it influences gene function and protein production. Additionally, spatial transcriptomics plays a pivotal role in comprehensive tissue analysis, offering a holistic perspective on gene expression in its native spatial environment, ultimately enhancing our understanding of cellular and tissue biology [25].

3.3. Epigenomics

NGS has revolutionized epigenomic studies by enabling comprehensive profiling of epigenetic marks across the genome. Epigenomics refers to the study of epigenetic modifications and their impact on gene expression and chromatin structure. These modifications are heritable changes that influence how genes are turned on or off without altering the underlying DNA sequence. Epigenomics sequencing encompasses DNA Methylation Sequencing, Histone Modification Profiling, Non-Coding RNA Sequencing, and Chromatin Structure Analysis. These techniques collectively provide insights into the regulation of gene expression and chromatin structure, influencing various aspects of biology and disease. In clinical diagnosis, epigenetic data plays a pivotal role, especially in diseases like cancer. Aberrant DNA methylation patterns, for instance, are associated with the silencing of tumor suppressor genes or the activation of oncogenes. These epigenetic changes can serve as diagnostic biomarkers, aiding in the early detection and classification of cancers. Some key methods include[26]:

3.3.1. Bisulfite sequencing

This technique analyzes DNA methylation patterns with single-nucleotide resolution. It converts unmethylated cytosines to uracil, leaving methylated cytosines unaltered. By sequencing the converted DNA and comparing it to the reference genome, researchers can identify sites of differential DNA methylation, aiding in the classification and early detection of diseases [27].

3.3.2. ChIP-Seq (Chromatin Immunoprecipitation Sequencing)

ChIP-Seq maps histone modifications and transcription factor binding sites across the genome. It immunoprecipitates chromatin fragments associated with specific modifications or transcription factors, offering insights into chromatin structure and gene regulation changes linked to diseases [28].

3.3.3. Atac-seq (Assay for Transposase-Accessible chromatin Sequencing)

ATAC-Seq identifies regions of open chromatin, indicating accessible regulatory elements. By inserting sequencing adaptors into open chromatin regions, it maps active regulatory elements with precision, revealing disease-specific changes in chromatin accessibility [29].

3.3.4. RNA-Seq for Non-Coding RNAs

RNA-Seq of non-coding RNAs, such as microRNAs and long non-coding RNAs, provides insights into their roles in gene regulation and disease. Dysregulated non-coding RNAs often correlate with diseases, making them valuable diagnostic markers for disease status, classification, and prognosis in clinical settings [30].

3.4. Pharmacogenomics

NGS can predict drug responses and adverse effects, aiding in proper drug dosage calculations. This capability is central to pharmacogenomics (PGx), a genomics branch studying how an individual's genetics affect drug responses. PGx analyzes genetic variants, like drug-metabolizing enzymes, to customize drug treatments, optimizing efficacy and safety. NGS underpins PGx, enabling personalized medicine by matching drugs to a patient's unique genetic makeup [31].

4. Bioinformatic analysis

(NGS) is a powerful technique that generates vast amounts of DNA and RNA sequence data, necessitating the application of sophisticated bioinformatics methodologies to extract meaningful insights. The NGS data analysis is a crucial step and its workflow comprises of several stages, commencing with the acquisition of raw sequencing data, which undergoes rigorous preprocessing steps, including quality control and adapter trimming to ensure data integrity. Subsequently, these sequences are aligned to reference genomes or transcriptomes, enabling the identification of genetic variations like single-nucleotide polymorphisms (SNPs) and insertions/deletions (indels), as well as quantification of gene expression levels [32]. Further downstream, differential expression analysis sheds light on genes exhibiting varying expression across different experimental conditions, while a detailed exploration of the transcriptome allows characterization of alternative splicing events and gene isoforms. In cases where reference genomes are unavailable, de novo assembly techniques are employed. Advanced bioinformatics tools play a pivotal role in identifying genetic variations, extending beyond SNPs to encompass copy number variations (CNVs) and structural variants. These tools also support integrative analyses, combining NGS data with other genomic and functional sources to investigate complex regulatory networks. A list of various bioinformatic tools and methods are enlisted in (Table 2). The methodology for identifying genetic variants involves six key steps: data quality control/adapter trimming, read alignment, alignment file preprocessing, variant detection, variant annotation, and variant prioritization. These steps are critical in variant detection workflows, with each step having its own set of software tools for execution. Data quality control, such as adapter trimming using tools like Trimmomatic, ensures high-quality input data. Read alignment with tools like BWA aligns the reads to a reference genome, followed by preprocessing steps to optimize the data. Variant detection algorithms, such as GATK for SNVs and small indels, identify genetic variations, and raw variant lists are generated. Filtering and annotation are used to refine these lists, removing false positives and determining the variants' impact on genes. Finally, variant prioritization involves ranking the variants based on their potential pathogenicity, often relying on custom strategies and domain expertise that are skilled workforce, however these need to be updated with the evolving pace of time [33].

5. Clinical applications of next- next-generation sequencing

5.1. Rare genetic disorders

Rare diseases, encompassing approximately 7,000 distinct conditions, significantly impact an estimated 260 to 440 million individuals, accounting for approximately 3.5 % to 5.9 % of the world's population. According to the Orphanet database, about 71.9 % of these diseases are genetic, with nearly 69.9 % exclusively affecting pediatric cases[96]. Rare diseases pose a substantial challenge in healthcare, with many remaining undiagnosed. Accurate diagnosis is essential for improved disease management and therapy identification, while also preventing unnecessary treatments. Inherited rare diseases require knowledge of the causative variant and inheritance mode for family planning decisions. However, patients with rare diseases often endure a lengthy and emotionally distressing diagnostic journey, averaging 4–5 years or more. NGS has revolutionized genetic testing, enhancing cost-effectiveness and accuracy. Exome sequencing, particularly in family trios, successfully identifies causative variants in undiagnosed diseases. Despite advances, some patients still await definitive molecular diagnoses. Innovative strategies such as WES, transcriptome sequencing, metabolomics, proteomics, and methyl profiling tailored to the patient's phenotype aim to address this diagnostic gap and provide answers to those affected by rare diseases. Various studies have compared the diagnostic performance and cost-effectiveness of NGS with other DNA testing methods. Traditional G-banding identifies chromosomal aberrations with a 3 % diagnostic yield for unexplained constitutional disorders. High-resolution Chromosomal Microarray Analysis (CMA) has a higher yield of 15–20 % for the same disorders. NGS techniques include WES with a 25 % diagnostic yield for Mendelian disorders and WGS with a 27 % yield for genetic diseases [97]. However, WES has limitations in detecting non-coding mutations. In a recent study by Burdick *et al.* [98], it was found that 28 % of diagnoses among Undiagnosed Diseases Network participants couldn't be resolved using WES alone. These cases required WGS or other omics techniques due to WES's inability to detect pathogenic non-coding variants, copy number variations, and repeat expansions. To address this, WGS is being explored for rare disease diagnosis, offering a more comprehensive approach. The UK100K project also highlighted the discovery of disease-associated non-coding variants disrupting gene transcription using WGS. Notably, they reported a case of a boy with autism spectrum disorder and thrombocytopenia, where WGS revealed a hemizygous deletion of a GATA1 enhancer, explaining his persistent low platelet count and other blood parameter abnormalities typically associated with a GATA1 mutation. [99] However, WES, which focuses on the protein-coding regions of the genome, has been particularly valuable in identifying disease-causing variants. In a study by [39], WES successfully identified disease-causing mutations in 48 % of patients with rare diseases, leading to accurate diagnoses and tailored treatment plans. NGS-based approaches have facilitated comprehensive analysis of the DMD gene in Duchenne muscular dystrophy, leading to the identification of disease-causing mutations [100] Neurodevelopmental disorders like autism spectrum disorder and intellectual disability have also benefited from NGS, with studies revealing significant contributions of de novo mutations in genes such as CHD8 and DYRK1A [101]. Using Life Technologies' SOLiD platform. Lupski et al., in their study, made a significant discovery regarding the presence of compound, heterozygous causative alleles of the SH3TC2 gene within the primary author's genome, as well as in their relatives who were affected by Charcot-Marie-Tooth neuropathy. This hereditary demyelinating disorder impacts the peripheral nervous system. The identification of these alleles was accomplished through the utilization of whole genome sequencing, a comprehensive genetic analysis technique [102]. The field of inherited metabolic disorders NGS has greatly improved diagnosis accuracy, as evidenced by the identification of disease-causing mutations in patients suspected of mitochondrial disorders [103]. NGS has also proven instrumental in diagnosing rare pediatric diseases with complex clinical presentations, including skeletal dysplasia and cardiomyopathies [104]Additionally, by utilizing NGS technologies such as whole exome sequencing, undiagnosed genetic disorders have been successfully elucidated [105]. NGS-based panels have played a crucial role in identifying germline mutations associated with cancer predisposition syndromes, enabling early detection and tailored surveillance strategies [46]. In a study by [106] WES was used to identify de novo mutations in the KMT2A gene as the cause of a rare neurodevelopmental disorder called Wiedemann-Steiner syndrome. The researchers performed trio-based whole-exome sequencing and found that the identified mutations were not present in the parents, confirming their de novo nature has proven to be a valuable tool in the diagnosis of mosaic mutations, which can account for a significant portion of the mutational load in rare diseases [107]utilized NGS to identify mosaic mutations in individuals with Sturge-Weber syndrome (SWS), a rare neurocutaneous disorder. SWS is caused by somatic mosaic

Table 2
NGS Data Analysis Tools and Methods.

Analysis Type	Tools	Description	Key Features	Input Data	Output Data	References
Common Analysis	FastQC, FASTX-toolkit, MultiQC	Quality check of sequences	QC reports, adapter trimming, quality assessment	FASTQ files	QC reports	[34–40]
	Trimmomatic, Cutadapt, fastp	Trimming of adaptors and low-quality bases	Adapter removal, quality filtering, flexibility	FASTQ files	Trimmed FASTQ files	
	BWA, Bowtie, dragMAP	Alignment of sequence reads to reference genome	Short-read alignment, indexing, speed	FASTQ/BAM files	BAM files	
	MultiQC	Reports visualization	Consolidates QC reports	MultiQC report files	HTML, PDF reports	
Whole-Genome Sequencing/ Whole-Exome Sequencing/Targeted Panel	Picard, Sambamba	Removal of duplicate reads	Marking duplicates	BAM files	BAM files	[41–61]
	GATK, freeBayes, Platypus, VarScan, DeepVariant, Illumina Dragen	Variant calling (SNPs and indels)	Haplotype-based, joint calling, high accuracy	BAM files	VCF files	
	bcftools	Filter and merge variants	VCF file manipulation	VCF files	Filtered/Merged VCF files	
	ANNOVAR, ensemblVEP, snpEff, NIRVANA	Variant annotation	Functional annotation, pathogenicity prediction	VCF files	Annotated VCF files	
	DELLY, Lumpy, Manta, GRIDDS, Wham, Pindel	Structural variant calling	Detection of SVs	BAM files	VCF files	
	CNVnator, GATK gCNV, cn.MOPS, cnvCapSeq, ExomeDepth	Copy number variation (CNV) calling	Segmentation-based, read depth analysis	BAM files	CNV files	
Transcriptomics	TopHat2, HISAT2, STAR	Alignment of reads to reference	Spliced alignment, gene annotation, transcript quantification	FASTQ files	BAM files	[62–73]
	featureCounts, HTSeq-count, Salmon, Kallisto	Transcript quantification	Quantification methods	BAM files (sorted)	Count tables	
	DESeq2, EdgeR, DAVID, clusterProfiler, Enrichr	Differential gene expression analysis	Enrichment analysis, pathway analysis	Count tables	Differential gene lists, pathway results	
Epigenomics-Methyl Seq	Bwameth, BS-Seeker2, Bismark	Sequence aligners	Bisulfite mapping, methylation calling	FASTQ files	BAM files	[74–77]
	MethylDackel	Methylation level quantification	Efficient cytosine methylation quantification	BAM files	Methylation level files	
	Metilene, BSsmooth, methylKit	Differential methylation	Identification of DMRs	Methylation level files	Differentially methylated regions	
Epigenomics-ChIP seq	Samtools	Removal of PCR duplicates	PCR duplicate removal	BAM files	BAM files	[78–88]
	MACS2, SICER2, SPP	Peak calling	Identifying ChIP-enriched regions	BAM files	Peak-called regions	
	Bedtools	Peak filtering	Customized peak filtering	Peak-called regions	Filtered peak-called regions	
	ChipQC, Phantompeakqualtools	Enrichment quality control	Assessing ChIP-seq data quality	BAM files	Quality control reports	
	diffBind, MAnorm, MMDiff	Enrichment comparison	Identifying differentially bound regions	Peak-called regions	Differential binding regions	
	MemeCHiP, Homer, RSAT	Motif analysis	Discovering DNA motifs enriched in peaks	Peak-called regions	Motif enrichment results	
16 s rRNA seq	QIIME2, mothur, USEARCH	16S rRNAseq analysis pipelines	Microbial community analysis	16S rRNA sequences	Microbial community profiles	[89–93]
	Greengenes, Silva, RDP	Ribosomal RNA databases	Taxonomic classification of 16S rRNA sequences	16S rRNA sequences	Taxonomic classifications	
Shotgun Metagenomics	MetaPhlAn4, Kaiju, Kraken	Taxonomic classification	Species-level and functional classification	Metagenomic sequences	Taxonomic and functional profiles	[94–96]
	metaSPAdes, metaIDBA	Assembly of metagenomic reads	De novo metagenomic assembly	Metagenomic sequences	Metagenomic assemblies	
	NCBI non-redundant protein database	Protein databases for taxonomic classification	Protein database search	Protein sequences	Taxonomic classifications	
	Prokka, MetaGeneMark	Gene annotation	Functional annotation of genes	Metagenomic assemblies	Annotated gene sets	
	COG, KEGG, GO	Databases for functional annotation of genes	Functional pathway analysis	Annotated gene sets	Functional pathway results	

Abbreviations - **ANNOVAR**-ANNOtate VARiation; **BWA**-Burrows Wheeler Aligner; **cn.mops** -Copy Number Estimation by a Mixture Of PoissonS; **COG**-Clusters of Orthologous Groups of Proteins; **DAVID**-A Database for Annotation, Visualization and Integrated Discovery; **Ensembl VEP**-Ensembl Variant Effect Predictor; **Fastp-Fsatq** Preprocessor; **GATK**-Genome Analysis Tool Kit; **GO**-Gene Ontology; **HISAT2**-Hierarchical Indexing for Spliced Alignment of Transcripts; **HOMER**-Hypergeometric Optimization of Motif EnRichment; **Htseq-count**-High-Throughput Sequence Analysis in Python; **KEGG-** Kyoto Encyclopedia of Genes and Genomes; **NCBI**-National Center for Biotechnology Information; **MACS -** Model-Based Analysis for ChIP-Seq; **MEME**-Multiple EM for Motif Elicitation; **Meta-IDBA**-Meta-Iterative De

Bruijn Graph De Novo Short-Read Assembler; **MetaPhlAn**-Metagenomic Phylogenetic Analysis; **metaSPAdes**-meta St Petersburg Genome Assembler; **QIIME**-Quantitative Insights Into Microbial Ecology; **RDP**-Ribosomal Database Project; **RSAT**-Regulatory Sequence Analysis tools; **SICER**-Spatial Clustering Approach for the Identification of ChIP-Enriched regions; **SPP**-The Signaling Pathways Project; **STAR**-Spliced Transcripts Alignment to a Reference.

mutations in the GNAQ gene. The researchers performed deep sequencing of affected tissues from SWS patients and detected low-frequency somatic mutations in the GNAQ gene that were not present in the germline DNA. This provided strong evidence for the presence of mosaic mutations contributing to the disease phenotype. The use of NGS in large-scale collaborative efforts, such as the Deciphering Developmental Disorders (DDD) project, has led to significant advancements in rare disease diagnosis. The DDD project involved sequencing the exomes of over 13,000 individuals with severe undiagnosed developmental disorders, resulting in the identification of novel disease genes and improving diagnostic rates [108]. NGS has played a pivotal role in the diagnosis of Mendelian diseases as these are genetic disorders caused by mutations in a single gene. They follow predictable patterns of inheritance and can have significant impacts on health and quality of life. It is estimated that around 1 in 200 births is affected by a Mendelian disease. This translates to approximately 0.5 % of the population being affected by these disorders. Within the vast human genome, consisting of approximately 20,000–25,000 protein-coding genes, a remarkable 3348 genes have been discovered to be associated with Mendelian diseases. A notable case study using WES is the diagnosis of Miller Syndrome, a rare craniofacial disorder where [109] identified the causative gene DHODH by analyzing the exome of affected individuals. Another study employed WES to diagnose a complex neurodevelopmental disorder, leading to the identification of a de novo mutation in the KCNH1 gene [110]. NGS is vital in diagnosing congenital disorders caused by genetic abnormalities present at birth. It enables comprehensive genetic analysis to identify the underlying causes of these conditions. One notable case where NGS played a crucial role in the diagnosis of a congenital disorder is the identification of the genetic cause of Kabuki syndrome. Kabuki syndrome is a rare developmental disorder characterized by distinctive facial features, intellectual disability, and various physical anomalies by performing WES on the affected individual and their family members, the researchers identified a de novo (spontaneous) mutation in the KMT2D gene. This gene encodes a protein involved in regulating gene expression and is known to be associated with Kabuki syndrome [111]. In another case study NGS was used to identify the role of the CHD7 gene in CHARGE syndrome (a rare congenital disorder characterized by multiple anomalies, including coloboma of the eye, heart defects, atresia of the choanae, retardation of growth and development, genital and/or urinary abnormalities, and ear abnormalities.) Through NGS technologies, pathogenic variants in CHD7 were found in individuals with the syndrome, establishing it as a major causative gene and providing insights into the disorder's molecular basis [112] In a study by [113] WES identified a novel homozygous mutation (c.1227G > A) in the *SUOX* gene in a 5-day-old Chinese female child with Isolated Sulfite Oxidase Deficiency (ISOD). This mutation resulted in a non-functional sulfite oxidase enzyme due to truncation. It was inherited recessively from the child's parents and was absent in 200 healthy controls. In a study [114] involving 68 patients with right-sided colon cancer (RCC), left-sided coloncancer (LCC), and rectal cancer (RC). This large-scale, high-depth multiregional sequencing effort examined 206 tumor regions, encompassing primary tumors, lymph nodes (LN), and extra nodal tumor deposits (ENTD). The results illuminated the significance of WES in uncovering intratumor heterogeneity (ITH) and the dynamics of clonal evolution. Notably, the study revealed a Darwinian pattern of evolution, with LCC and RC displaying more intricate and divergent evolutionary pathways than RCC. In another study [115], WES identified a novel *CEP290* gene variant in a fetus with Meckel syndrome (MKS4). MKS is a severe multisystemic disorder, and the identified variant led to a truncated CEP290 protein and altered cilia length. Whole exome sequencing unveiled a rare homozygous missense mutation (c.2326C > T; p.R776C) in the *DAG1* gene, absent in healthy controls [116]. This unique mutation, impacting a conserved region, resulted in reduced α-dystroglycan expression in muscle tissue. This discovery sheds light on a novel form of muscular dystrophy-dystroglycanopathy, MDDGC9, observed in a Chinese family, emphasizing the power of genetic sequencing in uncovering rare conditions. In a Chinese patient with Acute Intermittent Porphyria (AIP) [117], WES revealed a significant discovery. The 28-year-old male patient presented with severe abdominal symptoms and depression, while his parents were unaffected. WES and Sanger sequencing identified a unique heterozygous mutation (c.809delC) in the *HMBS gene*, leading to a truncated HMBS protein. This mutation was absent in the patient's parents and 100 healthy controls. Following American College of Medical Genetics and Genomics (ACMG) guidelines, the variant was classified as "likely pathogenic." This finding broadens our understanding of *HMBS* gene-related AIP and underscores the critical role of WES in the diagnosis and screening [118]of this rare condition, especially in cases with unusual clinical features like mild anemia. Apart from genome sequencing techniques, RNA-seq stands out as a powerful gene expression analysis technology using NGS. It offers higher precision than microarrays, reaching down to the base-pair level. RNA-seq also excels at detecting alternative splicing and gene fusions, which pose challenges for WES and WGS. In a recent study by Oliver et al.[119], RNA-seq analysis of 47 individuals with undiagnosed rare genetic diseases revealed 11 potentially pathogenic fusion transcripts, including examples like *SAMD12-EXT1* fusion in a patient with multiple exostoses and *ATM-SLC35F2* fusion in a patient with severe combined immunodeficiency.

However, with the wealth of genetic information generated by NGS comes the challenge of identifying the subset of variants with clinical relevance. The process of filtering and annotating genetic variants involves multiple components, and the decision to include or exclude variants relies on data resources and functional annotation algorithms. Additionally, the validation of candidate genetic variants depends on the availability of supporting evidence for their clinical significance and functional impact.

Standardized guidelines, such as those provided by the ACMG, have been instrumental in guiding the interpretation of sequencing variants in clinical laboratories. These guidelines employ a systematic scoring system to evaluate variant evidence and classify variants into categories such as pathogenic, likely pathogenic, uncertain significance, likely benign, and benign. Nevertheless, the terminology and interpretation of variants can still vary among laboratories, highlighting the need for ongoing standardization efforts.

Variant databases play a crucial role in curating and interpreting disease-associated variants, assisting in clinical variant management. These databases can be categorized into population-based resources and disease-specific variant atlases. While population-based databases provide information on variant frequencies in large populations, they may lack representation of certain ethnic groups or genders. Disease-specific databases, although valuable, may have limitations in terms of data quality and statistical power for specific diseases.

Efforts are underway to enhance the quality and representation of genetic variant curation. Initiatives like the MyCode Project aim to collect genetic data from patient populations to advance precision medicine. Collaboration platforms like GeneMatcher help gather rare disease patient samples that share variants in candidate disease genes. These endeavors are crucial for improving the accuracy of genetic variant interpretation and expanding our understanding of rare diseases [120,121].

5.2. Reproductive and newborn screening

The Newborn Screening (NBS) program, introduced by the World Health Organization (WHO) following the Wilson and Jungner guidelines, is a public health initiative aimed at identifying medical conditions in newborns before they become symptomatic. Over the years, the number of conditions covered by NBS has expanded, including phenylketonuria (PKU), hemoglobinopathies, sickle cell disease, glucose-6-phosphate dehydrogenase (G6PD) deficiency, congenital hypothyroidism (CH), congenital adrenal hyperplasia (CAH), tyrosinemia type I (HT1), cystic fibrosis (CF), and more[122].

Next-generation sequencing (NGS) has revolutionized NBS by enabling the rapid and cost-effective analysis of a wide range of genetic conditions. NGS can sequence multiple genes simultaneously, providing a more comprehensive and efficient screening approach. WES and WGS can investigate all protein-coding genes or entire genomes in a single test.

While NGS offers numerous benefits, including early diagnosis and improved carrier identification, its widespread use raises ethical and practical considerations. The inclusion of conditions in NGS-NBS panels should be based on solid scientific evidence, especially when there is no current actionable treatment. Long-term outcomes and psychological impacts also need to be studied before implementing NGS in population screening.

Several initiatives and programs, such as NeoSeq and NESTS, have explored the feasibility and cost-effectiveness of NGS-based NBS, with promising results. NGS can detect causal mutations in inherited metabolic disorders using dried blood spots, making it ideal for identifying monogenic forms of these conditions. Additionally, amplicon-based panels have been validated for diagnostic purposes and as second-tier tests for NBS[122].

NGS has transformed prenatal diagnosis and non-invasive prenatal testing (NIPT). A series of case studies by [123]demonstrated the clinical utility of NGS-based NIPT in identifying fetal chromosomal abnormalities. The study included over 15,000 pregnancies and showed that NGS-based NIPT accurately detected trisomy 21 (Down syndrome), trisomy 18 (Edwards syndrome), and trisomy 13 (Patau syndrome) with a low false-positive rate. It can accurately detect chromosomal abnormalities, microdeletion/microduplication syndromes, single gene disorders, determine fetal sex, and perform paternal haplotyping. NGS-based NIPT analyzes fetal genetic information from maternal blood, providing valuable insights without invasive procedures The use of NGS significantly enhances the precision of pre-implantation genetic testing (PGT). Pre-implantation genetic testing (PGT) is a crucial tool in assisted reproductive technology, improving live birth chances and preventing genetic disorders. Couples with fertility challenges due to factors like advanced maternal age, chromosomal issues, or specific genetic mutations benefit from PGT [124]. It screens embryos before implantation, identifying those likely to result in healthy pregnancies. Advanced methods like NGS enhance accuracy by assessing all 24 chromosomes. Two major platforms, Illumina MiSeq, and Thermo Fisher Ion PGM, are used for NGS in PGS, each with unique features.

Illumina MiSeq employs sequencing by synthesis, accurately detecting single bases as they incorporate into DNA strands. It can identify whole chromosome aneuploidy, mosaicism, mitochondrial copy number, and even single-gene disorders. The error rate is low at 0.8 %, and it can detect segmental imbalances of around 14 Mb or larger [125].

Thermo Fisher Ion PGM uses semiconductor sequencing technology, detecting proton release during nucleotide incorporation. It can identify whole chromosome aneuploidy, deletions, duplications, mosaicism, mitochondrial copy number, and single-gene disorders with an error rate of 1.71 %. It offers advantages like reduced cost and faster sequencing.

NGS has shown excellent concordance with Array Comparative Genomic Hybridization (aCGH) in clinical trials, with a 99.5 % concordance rate. Euploid blastocyst transfers after NGS testing resulted in high clinical pregnancy rates. NGS has the potential benefits of reduced cost, high throughput, automation, and improved detection of partial aneuploidies, and mosaicism [126].

NGS can detect mosaicism, as demonstrated in a study where some blastocysts initially diagnosed as euploid by aCGH (Array Comparative Genomic Hybridization) were found to be mosaic by NGS. However, mosaic embryos are associated with a higher risk of miscarriage.

5.3. Infectious diseases

Infectious diseases continue to pose a significant global health threat, but diagnosing the responsible pathogens accurately and rapidly can be challenging. Current diagnostic methods often require a combination of tests such as culture, serologic assays, and nucleic acid amplification, leading to time-consuming processes. Culturing common pathogens can take days or even weeks, and certain microorganisms, like viruses, are impossible to culture. Obtaining samples can be invasive, and prior antibiotic use may affect test sensitivity.

Although syndromic multiplex PCR panels, 16 s ribosomal DNA detection, and matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry (MALDI-TOF MS) have improved turnaround times, up to 60 % of infectious diseases still have unknown etiologies. Delayed or missed diagnoses often result in the overuse of broad-spectrum antibiotics or antifungal drugs, hindering targeted and effective treatment. Metagenomic next-generation sequencing (mNGS) offers a rapid and comprehensive approach to identifying bacteria, fungi, viruses, and parasites directly from clinical specimens [127]. It excels in detecting mycobacteria, anaerobes, atypical pathogens, and viruses, outperforming culture-based methods. Importantly, mNGS is less affected by prior antibiotic exposure. It has proven its effectiveness in rapid infectious disease diagnostics, as exemplified by a case study by Zeng et al. [128]. They successfully detected *Klebsiella pneumoniae* DNA in cerebrospinal fluid and blood from a patient with liver abscesses and meningitis, even when traditional cultures yielded negative results.

Furthermore, Michel et al. [129]. utilized metagenomic Whole Genome Sequencing (mWGS) to diagnose a lung infection caused by *Ureaplasma spp*. and *Mycoplasma hominis* in a post-lung transplant patient. This approach not only identified the pathogens but also facilitated the effective treatment of the infection.

In another study by Yan *et al*. [130], the diversity of bile bacteria in cholecystitis was explored through a combination of culture, conventional 16 s rRNA sequencing, and Whole Genome Sequencing (WGS). This comprehensive approach provided valuable insights into the role of bile bacteria in cholecystitis.

By transforming diagnostics, surveillance, and research in infectious diseases NGS has revolutionized the tracking, transmission, and evolution of pathogens during infectious disease outbreaks. The West Africa Ebola virus outbreak in 2015 showcased the speed and efficiency of NGS in the field, where researchers utilized Nanopore technology to sequence over 140 Ebola virus samples directly in the field within hours [131]. Similarly, during the early stages of the COVID-19 pandemic, NGS played a crucial role in rapidly determining the sequence of the SARS-CoV-2 virus, enabling disease diagnosis, surveillance, and the development of treatments and vaccines [132]. [138]NGS has also proven valuable in identifying and categorizing clinical isolates during outbreaks, aiding in the identification of emergent resistant or virulent strains [133]. In addition to clinical care, NGS has significantly advanced public health surveillance. It has improved the speed and accuracy of tracking and sourcing foodborne illnesses, allowing for faster detection and response [140]It is a crucial role in monitoring influenza strains, aiding in the development of annual vaccines.The detection of human papillomavirus (HPV) detection techniques has progressed from low-throughput methods to high-throughput NGS screening. It has emerged as a valuable tool for HPV detection due to its high sensitivity in identifying low-copy-number HPV types, novel strains, and variants that standard molecular methods may miss. Whole-

genome NGS, in particular, allows for precise characterization of HPV sequences, including variants and subvariants. It has revealed critical information about HPV-associated diseases, such as identifying high-risk sublineages linked to cancer and pinpointing specific genetic factors influencing viral persistence. Despite its potential, NGS adoption in clinical diagnostics has been limited due to challenges like standardization, cost, and infrastructure requirements [134]. Nanopore sequencing, for instance, has enabled rapid identification of pathogens and antibiotic resistance markers, allowing for faster administration of appropriate antibiotics [135]. NGS can also resolve discrepancies between different diagnostic approaches and detect multiple co-infections that may complicate standard testing results [145146] Integrating sequencing data of both pathogens and the host's microbiome has enhanced the predictive value of testing. By combining host gene expression profiles and microbiome analysis, NGS has accurately identified causative pathogens and assessed the patient's immune response. Additionally, NGS has provided insights into the overall health of patients, monitoring changes in the host microbiome and viral load in immunocompromised individuals [136]. By employing WES [137] successfully identified drug-resistant mutations in *Mycobacterium tuberculosis*, providing valuable insights into the prevalence of drug resistance in tuberculosis. This information is crucial for surveillance purposes and facilitates the development of targeted treatment strategies to combat drug-resistant strains. In the case of fungal infections, WGS has played a crucial role in identifying and tracking the emerging multi-drug resistant yeast, *Candida auris*. This pathogen has caused fatal infections and outbreaks in healthcare facilities worldwide. Standard laboratory methods struggle to differentiate *C. auris* from other yeasts due to their similar appearance and biochemical characteristics. Conventional tests often misidentify it as *Candida haemulonii* or other yeasts, and an enrichment broth method, although available, is time-consuming in terms of patient care. WGS has been instrumental in confirming *C. auris'* identity and tracking its epidemiology on a global scale, enabling better management of this challenging pathogen. NGS has revolutionized the field of infectious disease prevention and management. Its speed, accuracy, and ability to detect pathogens and their genetic characteristics have transformed outbreak tracking, pathogen identification, and diagnosis. NGS also holds promise for integrating host and pathogen sequencing data, leading to more comprehensive and personalized approaches to infectious disease diagnosis and treatment.

5.4. Cancer

Cancer is fundamentally a genomic disease, and the identification of specific genomic aberrations has become a cornerstone of precision medicine in oncology. Pioneering initiatives such as The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC) have significantly contributed to refining cancer classification systems [138]. They have deepened our understanding of the intricate interplay between DNA mutations, RNA expression, and epigenomic patterns in cancer. NGS plays a pivotal role in uncovering various genomic alterations commonly observed in both hematologic and solid malignancies, including single nucleotide variants (SNVs), small insertions and deletions (indels), copy number variations (CNVs), and fusion genes. While the availability of whole-genome, exome, or transcriptome sequencing has increased, targeted gene sequencing remains the preferred method in clinical laboratories for cancer diagnosis owing to optimal sequencing quality, cost-effectiveness, and rapid turnaround time. Inherited cancer testing using NGS panels has become a valuable tool in clinical settings, particularly for identifying individuals at risk of hereditary cancer predisposition syndromes (HCPS). HCPS, such as hereditary breast and ovarian cancer syndrome (HBOC) and Lynch syndrome, are typically associated with high-penetrant genes like *BRCA1* and *BRCA2*. However, recent studies have expanded our understanding of these syndromes by uncovering additional moderate-risk genes and low-penetrance alleles using NGS panels.

For instance, one study employing a 21-gene NGS panel for HBOC in ovarian, peritoneal, or fallopian tube carcinoma patients revealed that 22.8 % carried mutations in various genes, including *BRCA1* and *BRCA2*. Additionally, a 25-gene panel in breast cancer patients identified mutations not only in *BRCA1/2* but also in genes like *PALB2, CHEK2*, and *ATM* [139].

In another study by [140] Targeted NGS was utilized to investigate the presence of germline mutations in DNA Mismatch Repair (MMR) pathway genes, associated with Lynch Syndrome (LS), in individuals with hereditary breast cancer. The study included 711 hereditary breast cancer patients, 60 sporadic breast cancer patients, and 492 healthy donors. Results revealed that 9.7 % of hereditary breast cancer patients harbored mutations in MMR pathway genes, with 4.5 % having pathogenic or likely pathogenic mutations. Importantly, these mutations were significantly more frequent in hereditary breast cancer patients compared to healthy donors (0.4 %). This study highlights the power of targeted sequencing in identifying LS-related mutations, suggesting the need to include LS-mutation testing in genetic counseling for hereditary breast cancer patients.

In the case of rare cancers, cancers of unknown primary (CUP), and pediatric cancers, Whole Genome and Transcriptome Sequencing (WGTS) holds great promise for enhancing and refining the diagnostic process. For instance, a recent clinical study involving 83 sarcoma patients demonstrated that WGS improved diagnosis in 14 % of cases, identified actionable biomarkers in 36 % of tumors, and detected germline cancer predisposition variants in 8 % of patients. This potential impact is particularly significant in sarcoma, which is among the most genetically diverse cancer types, with the World Health Organization recognizing 175 different soft tissue and bone tumors, each distinguishable by distinct genomic features. By employing WGTS [141], healthcare providers can potentially avoid assigning patients to treatments that are unlikely to be effective, thus reducing clinical toxicity and alleviating financial burdens on the healthcare system. particularly RNA sequencing (RNAseq), has transformed cancer genetics by providing extensive genomic and transcriptomic data [142]. These approaches link mutations and polymorphisms to individual tumor responses to therapies, offering invaluable insights for personalized cancer treatment. Combining both NGS and RNAseq approaches can be synergistic in molecular diagnostics for oncology.

In a review by Wang et al.[141], RNAseq methods in cancer research and diagnostics were discussed, showcasing the rapid progress from bulk RNAseq to single-cell RNAseq and digital spatial RNA profiling. Each technology has unique strengths, making it suitable for various clinical oncology applications. For instance, bulk RNAseq is cost-effective and ideal for whole-transcriptome biomarker discovery, while single-cell RNAseq characterizes diverse cell types within tumors. Case studies illustrate the clinical relevance of RNAseq data, such as Zhigalova *et al.'s* [143] investigation of T-cell receptor diversity in response to anti-PD-1 treatment and Borger *et al.'s* analysis of molecular processes during hepatectomy. Additionally, Poddubskaya *et al.* [144] used RNAseq to guide personalized drug combination therapy for an *ALK*-positive lung cancer patient, significantly extending survival.

Furthermore, studies by Tang *et al.* [145] highlight the significance of RNA sequencing in identifying non-coding RNAs as prognostic biomarkers and regulators of cancer cell behavior, offering insights into potential therapeutic targets.

In the area of oncology, NGS has revolutionized cancer molecular diagnosis by enabling the simultaneous sequencing of numerous target genes, providing valuable early diagnostic markers. Large-scale projects like TCGA and the ICGC have contributed to refining cancer classification systems and understanding the interplay between DNA mutations, RNA expression, and epigenomic patterns. Targeted genetic tests are currently used for diagnosis and prognosis, while more comprehensive genomic tests are expected to become routine. WES is commonly employed due to its coverage of disease-causing mutations, while WGS allows for the comparison of tumor progression and treatment response

NGS panels are being used in clinical laboratories to identify biomarkers for cancer diagnosis, prognosis, and treatment. In acute myeloid leukemia (AML), mutated genes such as *NPM1, CEBPA, FLT3, DNMT3A, TET2, IDH1, IDH2, NRAS, ASXL1, KMT2A, WT1*, and *RUNX1* are frequently observed [146]. Each mutation has implications for patient outcomes. For instance, *NPM1* mutations are associated with favorable outcomes, while FLT3-ITD mutations indicate an increased risk of relapse and shorter overall survival [147].

The FDA has approved targeted therapies for specific genetic mutations in AML and lung cancer. In AML, midostaurin is approved for newly diagnosed patients with *FLT3* mutations, while enasidenib and ivosidenib are approved for relapsed/refractory AML with IDH2 and IDH1 mutations, respectively. In lung cancer, *EGFR* inhibitors like afatinib, gefitinib, erlotinib, and Osimertinib are approved for patients with *EGFR* mutations, while *ALK* inhibitors such as crizotinib, alectinib, and ceritinib are used for *ALK* rearrangements [148]. *ROS1* mutations in lung cancer can be targeted by crizotinib and entrectinib [148].

Apart from disease-specific biomarkers, some pan-cancer biomarkers have been identified. Microsatellite instability (MSI) and high tumor mutational burden (TMB) are tissue-agnostic markers that predict response to immunotherapy. Pembrolizumab and nivolumab are FDA-approved immunotherapies for MSI-high or mismatch repair deficient (dMMR) tumors. *NTRK* fusions are also tissue-agnostic biomarkers and larotrectinib and entrectinib are FDA-approved for solid tumors with *NTRK* fusions [149].

By harnessing the power of NGS, clinicians have unlocked new possibilities for improved clinical utility. In the realm of breast cancer diagnosis, the integration of genetic biomarkers such as *BRCA1, BRCA2, HER2, ER*, and *PR* has refined the profiling of tumors [150]. This advancement allows for more accurate diagnostics and personalized treatment strategies. Commercially available microarray-based tests like Oncotype DX and MammaPrint have enabled precise characterization, while the discovery of cancer-associated genes has led to the development of targeted therapies, exemplified by trastuzumab for *HER2* + breast cancers [151]. Liquid biopsy complimented using NGS, identifies circulating tumor DNA (ctDNA), offers a comprehensive cancer molecular profile. It aids cancer diagnosis, and treatment monitoring, and is promising for cancer screening and early detection, although implementation challenges persist. Traditional biopsies can be challenging, especially in metastatic cancers. Liquid biopsy, focusing on ctDNA in the blood, is less invasive and comprehensive for assessing tumor heterogeneity.[152] NGS enhances ctDNA sequencing sensitivity compared to tissue biopsies, making it valuable for various diagnostic purposes. Current ctDNA assays show promise in detecting early-stage cancers, including lung cancer, by identifying mutations and utilizing DNA methylation detection. An FDA-approved test for colorectal cancer uses *SEPT9* gene methylation [153].

These breakthroughs signify a transformative era in breast cancer management, where NGS-driven insights shape diagnostic precision and guide tailored therapeutic interventions. NGS-based liquid biopsies have shown promise in detecting early-stage cancers by analyzing ctDNA or CTCs. These approaches offer the potential for non-invasive cancer screening and early intervention. A study [154]demonstrated the ability of a targeted NGS panel to detect multiple cancer types with high sensitivity and specificity, including colorectal, breast, lung, and ovarian cancers. NGS is also employed in Minimal Residual Disease (MRD) monitoring and liquid biopsy. MRD refers to the small number of cancer cells remaining during or after treatment [155]. NGS-based tests, such as the ClonoSEQ test, can detect MRD at extremely low levels, aiding in treatment effectiveness evaluation and predicting the risk of relapse [156]. Liquid biopsy, which analyzes circulating cancer cells or DNA in bodily fluids, holds promise for MRD detection, real-time monitoring, and early cancer diagnosis. However, challenges such as low cancer-derived DNA fraction and clonal hematopoiesis need to be addressed for wider clinical application of liquid biopsy. In summary, NGS panels are used to identify disease-specific and pan-cancer biomarkers for cancer diagnosis, prognosis, and treatment selection. NGS also plays a role in clinical trial enrollment, MRD monitoring, and liquid biopsy for non-invasive disease monitoring and early detection [157].

6. Precision medicine

Precision medicine holds significant promise in various healthcare areas, including treating rare genetic disorders such as cystic fibrosis, oncology, and pharmacogenomics. Cystic fibrosis, an autosomal recessive disease caused by mutations in the *CFTR* gene, has seen advancements in targeted therapies. For instance, the drug ivacaftor was developed specifically for patients with cystic fibrosis and has shown improvements in pulmonary function, particularly for patients with specific mutations that affect *CFTR* channel activity. Combination approaches involving ivacaftor and lumacaftor have been approved, providing further benefits for patients with common genotypes [158]. Drug resistance is a significant challenge in cancer treatment, often arising from DNA mutations and metabolic changes that enhance drug degradation or inhibition. Genetic analysis can inform personalized treatment strategies to overcome drug resistance. An example of the importance of alternative therapeutic approaches is seen in the overexpression of *HER2*, a receptor tyrosine kinase, in 30 % of breast cancer patients. Prolonged use of kinase inhibitors targeting *HER2* can lead to treatment resistance[159]. NGS analysis helps identify gene mutations causing medication resistance, potentially opening avenues for restoring drug sensitivity through gene editing. NGS has transformed cancer diagnostics and treatment by providing in-depth genetic information. It offers significant advantages over traditional PCR methods, particularly in accuracy and mutation detection sensitivity. NGS can detect mutations with a Minor Allele Frequency (MAF) below 1 %, making it valuable for identifying rare mutations. In thyroid cancer, NGS enhances fine needle aspiration (FNA) cytology by analyzing genomic changes and helps stratify patients based on their cancer risk. It also aids in classifying molecular tumor types and predicting recurrence and metastasis. NGS has revealed additional somatic alterations beyond well-known mutations [160]. In lung cancer, NGS guides personalized treatment decisions and identifies new biomarkers for early diagnosis. WGS identified recurrent somatic variations, expanding our understanding of lung cancer genetics. NGS also plays a crucial role in immunotherapy, identifying target antigens for CAR-T cell therapy and supporting individualized treatment plans [161]. Breast cancer subtyping benefits from NGS, as it helps classify cancer subtypes based on variations in genes like *TP53, PIK3CA,* and *GATA3*. NGS enables the discovery of unknown genes related to treatment response and drug resistance [162]. A case study demonstrated NGS's potential in guiding immunotherapy for a patient with metastatic breast cancer. By identifying somatic cell mutations, NGS informed the use of immunotherapy, highlighting its role in personalized cancer treatment [163] In oncology, the focus has shifted from tissue of origin to molecular classification. Genomic characterization has become essential in guiding personalized therapy. For example, specific mutations such as *EGFR* and *ALK* in lung adenocarcinoma can be targeted with kinase inhibitors, while *BRAF* inhibition has shown promise in treating BRAF-mutant melanoma. Immune checkpoint therapy has also emerged as a successful approach, harnessing the immune system to target tumor cells. Genomic approaches and checkpoint therapy have shown potential in improving treatment outcomes. Pharmacogenomics, an early application of personalized medicine, aims to optimize drug response based on genetic variations. Genetic testing has been utilized to optimize dosing for drugs like warfarin and clopidogrel [164]. Despite some challenges, such as cost-effectiveness and limited availability of genomic information, the potential for pharmacogenomics in individualizing medication regimens remains substantial [164].

7. Limitations and future perspectives

The cost of NGS has undergone remarkable reductions since the first human genome sequencing in 2003, which originally costed $2.7 billion. Despite the remarkable progress in making NGS more cost-effective, it still faces substantial challenges, particularly when analyzing highly polymorphic regions of the genome. Managing the vast data volumes generated, especially in WGS, is a significant hurdle. This task necessitates efficient data transfer, storage, and high-capacity computing resources. Moreover, data analysis is complex due to short read lengths, elevated sequencing errors, and the risk of false positives, demanding rigorous validation. Distinguishing non-pathogenic variants from potentially disease-related ones is further complicated by the sheer volume of data. Nonetheless, researchers have been tackling these challenges effectively by utilizing automated data processing pipelines, precise data filtering techniques, and functional impact assessment tools. These strategies enable scientists to navigate the intricacies of NGS data analysis, even in regions of the genome with a high degree of polymorphism. While NGS has undeniably propelled clinical diagnosis forward, it is not without limitations. The analysis and interpretation of the vast amount of data generated by NGS can be intricate and time-consuming, requiring specialized bioinformatics expertise and robust computational infrastructure. Financial considerations also come into play, as NGS can be relatively costly and may extend the turnaround time for diagnostic results compared to traditional methods. Furthermore, establishing the clinical significance of genetic variants detected by NGS poses a challenge, given the prevalence of uncertain or unknown implications associated with many variants. Technical constraints, including sequencing errors and biases, can impede the precision and sensitivity of NGS. Ethical concerns, such as patient privacy and data security, have arisen due to the substantial volume of genomic information generated. Deploying NGS-based tests within clinical laboratories necessitates rigorous validation and standardization of procedures in the mainstay and background.

Future applications of NGS in clinical diagnosis and precision medicine hold enormous promise. NGS integration with other 'omics' technologies, like proteomics and metabolomics, will give researchers a more thorough understanding of how diseases work and how they respond to various treatments[165]. Artificial intelligence and machine learning algorithms will improve data analysis and interpretation by spotting associations and patterns that are clinically pertinent. Single-cell sequencing methods will make it possible to investigate rare cell populations and complex disease dynamics in greater detail[166]. Liquid biopsies are being improved, and this opens the door to non-invasive, real-time monitoring of disease progression, treatment effectiveness, and the detection of minimal residual disease[167]. However, it is essential to address the NGS's ethical, legal, and social ramifications, guaranteeing data security, privacy, and equal access to NGS-based therapies.

8. Conclusion

In conclusion, NGS has changed the landscape of clinical diagnostics and precision medicine. This ability to quickly and cost-effectively analyze genomic data has transformed our understanding and treatment of diseases. NGS plays an important role in genetic diseases, cancer genomics, infectious diseases, prenatal diagnosis, and pharmacogenomics, enabling the detection of mutations, gene fusions, and various infections by integrating NGS with other 'omics' technologies, leveraging artificial intelligence and machine learning, adopting single-cell sequencing, and refining liquid biopsies, the future of NGS in clinical practice holds immense potential. However, responsible implementation and addressing ethical considerations are paramount to ensure the maximum benefits of NGS are realized, leading to personalized and targeted therapies that greatly improve patient outcomes in the era of precision medicine.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

[1] F. Sanger, S. Nicklen, A.R. Coulson, DNA sequencing with chain-terminating inhibitors, Proc. Natl. Acad. Sci. 74 (12) (Dec. 1977) 5463–5467, https://doi.org/10.1073/pnas.74.12.5463.

[2] M. V. Olson, 'The human genome project', *Proceedings of the National Academy of Sciences of the United States of America*, vol. 90, no. 10. 1993. doi: 10.1073/pnas.90.10.4338.

[3] R.K. Saiki, et al., Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase, Science 239 (4839) (1979) 1988, https://doi.org/10.1126/science.2448875.

[4] B. E. Slatko, A. F. Gardner, and F. M. Ausubel, 'Overview of Next-Generation Sequencing Technologies', *Curr Protoc Mol Biol*, vol. 122, no. 1, Apr. 2018, doi: 10.1002/cpmb.59.

[5] B.-S. Petersen, B. Fredrich, M.P. Hoeppner, D. Ellinghaus, A. Franke, Opportunities and challenges of whole-genome and -exome sequencing, BMC Genet. 18 (1) (Dec. 2017) 14, https://doi.org/10.1186/s12863-017-0479-5.

[6] M. Margulies, et al., Genome sequencing in microfabricated high-density picolitre reactors, Nature 437 (7057) (Sep. 2005) 376–380, https://doi.org/10.1038/nature03959.

[7] M. Ronaghi, S. Karamohamed, B. Pettersson, M. Uhlén, P. Nyrén, Real-Time DNA Sequencing Using Detection of Pyrophosphate Release, Anal. Biochem. 242 (1) (Nov. 1996) 84–89, https://doi.org/10.1006/abio.1996.0432.

[8] K.J. McKernan, et al., Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding, Genome Res. 19 (9) (Sep. 2009) 1527–1541, https://doi.org/10.1101/gr.091868.109.

[9] M. Fedurco, BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies, Nucleic Acids Res. 34 (3) (Feb. 2006) e22–e, https://doi.org/10.1093/nar/gnj023.

[10] M.L. Metzker, Sequencing technologies — the next generation, Nat. Rev. Genet. 11 (1) (Jan. 2010) 31–46, https://doi.org/10.1038/nrg2626.

[11] D.A. Wheeler, et al., The complete genome of an individual by massively parallel DNA sequencing, Nature 452 (7189) (Apr. 2008) 872–876, https://doi.org/10.1038/nature06884.

[12] E.E. Schadt, S. Turner, A. Kasarskis, A window into third-generation sequencing, Hum. Mol. Genet. 19 (R2) (Oct. 2010) R227–R240, https://doi.org/10.1093/hmg/ddq416.

[13] F. Haque, J. Li, H.-C. Wu, X.-J. Liang, P. Guo, Solid-state and biological nanopore for real-time sensing of single chemical and sequencing of DNA, Nano Today 8 (1) (Feb. 2013) 56–74, https://doi.org/10.1016/j.nantod.2012.12.008.

[14] M. J. Levene, J. Korlach, S. W. Turner, M. Foquet, H. G. Craighead, and W. W. Webb, 'Zero-Mode Waveguides for Single-Molecule Analysis at High Concentrations', *Science (1979)*, vol. 299, no. 5607, pp. 682–686, Jan. 2003, doi: 10.1126/science.1079700.

[15] D. Branton, et al., The potential and challenges of nanopore sequencing, Nat. Biotechnol. 26 (10) (Oct. 2008) 1146–1153, https://doi.org/10.1038/nbt.1495.

[16] G. Costain, R.D. Cohn, S.W. Scherer, C.R. Marshall, Genome sequencing as a diagnostic test, Can. Med. Assoc. J. 193 (42) (Oct. 2021) E1626–E1629, https://doi.org/10.1503/cmaj.210549.

[17] G.A. Logsdon, M.R. Vollger, E.E. Eichler, Long-read human genome sequencing and its applications, Nat. Rev. Genet. 21 (10) (Oct. 2020) 597–614, https://doi.org/10.1038/s41576-020-0236-x.

[18] A. Warr, C. Robert, D. Hume, A. Archibald, N. Deeb, and M. Watson, 'Exome Sequencing: Current and Future Perspectives', *G3 Genes|Genomes|Genetics*, vol. 5, no. 8, pp. 1543–1550, Aug. 2015, doi: 10.1534/g3.115.018564.

[19] X.M. Pei, et al., Targeted Sequencing Approach and Its Clinical Applications for the Molecular Diagnosis of Human Diseases, Cells 12 (3) (Feb. 2023) 493, https://doi.org/10.3390/cells12030493.

[20] J.-W. Chen, L. Shrestha, G. Green, A. Leier, and T. T. Marquez-Lago, 'The hitchhikers' guide to RNA sequencing and functional analysis', *Brief Bioinform*, vol. 24, no. 1, Jan. 2023, doi: 10.1093/bib/bbac529.

[21] R. Stark, M. Grzelak, J. Hadfield, RNA sequencing: the teenage years, Nat. Rev. Genet. 20 (11) (Nov. 2019) 631–656, https://doi.org/10.1038/s41576-019-0150-2.

[22] A. Kulkarni, A.G. Anderson, D.P. Merullo, G. Konopka, Beyond bulk: a review of single cell transcriptomics methodologies and applications, Curr. Opin. Biotechnol. 58 (Aug. 2019) 129–136, https://doi.org/10.1016/j.copbio.2019.03.001.

[23] M. Kolanowska, A. Kubiak, K. Jażdżewski, A. Wójcicka, 'MicroRNA Analysis Using next-Generation Sequencing' (2018) 87–101, https://doi.org/10.1007/978-1-4939-8624-8_8.

[24] S.R. Atkinson, S. Marguerat, J. Bähler, Exploring long non-coding RNAs through sequencing, Semin. Cell Dev. Biol. 23 (2) (Apr. 2012) 200–205, https://doi.org/10.1016/j.semcdb.2011.12.003.

[25] C.G. Williams, H.J. Lee, T. Asatsuma, R. Vento-Tormo, A. Haque, An introduction to spatial transcriptomics for biomedical research, Genome Med. 14 (1) (Jun. 2022) 68, https://doi.org/10.1186/s13073-022-01075-1.

[26] K.C. Wang, H.Y. Chang, Epigenomics, Circ. Res. 122 (9) (Apr. 2018) 1191–1199, https://doi.org/10.1161/CIRCRESAHA.118.310998.

[27] D. Barros-Silva, C. Marques, R. Henrique, C. Jerónimo, Profiling DNA Methylation Based on Next-Generation Sequencing Approaches: New Insights and Clinical Applications, Genes (Basel) 9 (9) (Aug. 2018) 429, https://doi.org/10.3390/genes9090429.

[28] R. Nakato, T. Sakata, Methods for ChIP-seq analysis: A practical workflow and advanced applications, Methods 187 (Mar. 2021) 44–53, https://doi.org/10.1016/j.ymeth.2020.03.005.

[29] F.C. Grandi, H. Modi, L. Kampman, M.R. Corces, Chromatin accessibility profiling by ATAC-seq, Nat. Protoc. 17 (6) (Jun. 2022) 1518–1552, https://doi.org/10.1038/s41596-022-00692-9.

[30] A. Arrigoni *et al.*, 'Analysis RNA-seq and Noncoding RNA', 2016, pp. 125–135. doi: 10.1007/978-1-4939-6380-5_11.

[31] A. K. Daly, 'Pharmacogenetics: a general review on progress to date', *Br Med Bull*, pp. 1–15, Oct. 2017, doi: 10.1093/bmb/ldx035.

[32] R. Pereira, J. Oliveira, M. Sousa, Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics, J. Clin. Med. 9 (1) (Jan. 2020) 132, https://doi.org/10.3390/jcm9010132.

[33] D.C. Koboldt, Best practices for variant calling in clinical sequencing, Genome Med. 12 (1) (Dec. 2020) 91, https://doi.org/10.1186/s13073-020-00791-w.

[34] 'Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 1 June 2023).'.

[35] R. Iyer, V.G. Stepanov, B. Iken, Isolation and molecular characterization of a novel pseudomonas putida strain capable of degrading organophosphate and aromatic compounds, Adv Biol Chem 03 (06) (2013) 564–578, https://doi.org/10.4236/abc.2013.36065.

[36] P. Ewels, M. Magnusson, S. Lundin, M. Käller, MultiQC: summarize analysis results for multiple tools and samples in a single report, Bioinformatics 32 (19) (Oct. 2016) 3047–3048, https://doi.org/10.1093/bioinformatics/btw354.

[37] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (15) (Aug. 2014) 2114–2120, https://doi.org/10.1093/bioinformatics/btu170.

[38] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, Embnet J 17 (1) (May 2011) 10, https://doi.org/10.14806/ej.17.1.200.

[39] 'Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997. 2013 Mar 16.'.

[40] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods 9 (4) (Apr. 2012) 357–359, https://doi.org/10.1038/nmeth.1923.

[41] A. Tarasov, A.J. Vilella, E. Cuppen, I.J. Nijman, P. Prins, Sambamba: fast processing of NGS alignment formats, Bioinformatics 31 (12) (Jun. 2015) 2032–2034, https://doi.org/10.1093/bioinformatics/btv098.

[42] A. McKenna, et al., The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data, Genome Res. 20 (9) (Sep. 2010) 1297–1303, https://doi.org/10.1101/gr.107524.110.

[43] 'Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907. 2012 Jul 17.'.

[44] A. Rimmer, et al., Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications, Nat. Genet. 46 (8) (Aug. 2014) 912–918, https://doi.org/10.1038/ng.3036.

[45] D.C. Koboldt, et al., VarScan: variant detection in massively parallel sequencing of individual and pooled samples, Bioinformatics 25 (17) (Sep. 2009) 2283–2285, https://doi.org/10.1093/bioinformatics/btp373.

[46] R. Poplin, et al., A universal SNP and small-indel variant caller using deep neural networks, Nat. Biotechnol. 36 (10) (Nov. 2018) 983–987, https://doi.org/10.1038/nbt.4235.

[47] H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data, Bioinformatics 27 (21) (Nov. 2011) 2987–2993, https://doi.org/10.1093/bioinformatics/btr509.

[48] K. Wang, M. Li, H. Hakonarson, ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data, Nucleic Acids Res. 38 (16) (Sep. 2010) e164–e, https://doi.org/10.1093/nar/gkq603.

[49] W. McLaren, et al., The Ensembl Variant Effect Predictor, Genome Biol. 17 (1) (Dec. 2016) 122, https://doi.org/10.1186/s13059-016-0974-4.

[50] T. Rausch, T. Zichner, A. Schlattl, A.M. Stütz, V. Benes, J.O. Korbel, DELLY: structural variant discovery by integrated paired-end and split-read analysis, Bioinformatics 28 (18) (Sep. 2012) i333–i339, https://doi.org/10.1093/bioinformatics/bts378.

[51] R.M. Layer, C. Chiang, A.R. Quinlan, I.M. Hall, LUMPY: a probabilistic framework for structural variant discovery, Genome Biol. 15 (6) (2014) R84, https://doi.org/10.1186/gb-2014-15-6-r84.

[52] X. Chen, et al., Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications, Bioinformatics 32 (8) (Apr. 2016) 1220–1222, https://doi.org/10.1093/bioinformatics/btv710.

[53] D.L. Cameron, et al., GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly, Genome Res. 27 (12) (Dec. 2017) 2050–2060, https://doi.org/10.1101/gr.222109.117.

[54] Z.N. Kronenberg, et al., Wham: Identifying Structural Variants of Biological Consequence, PLoS Comput. Biol. 11 (12) (Dec. 2015) e1004572.

[55] K. Ye, M.H. Schulz, Q. Long, R. Apweiler, Z. Ning, Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads, Bioinformatics 25 (21) (Nov. 2009) 2865–2871, https://doi.org/10.1093/bioinformatics/btp394.

[56] A. Abyzov, A.E. Urban, M. Snyder, M. Gerstein, CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing, Genome Res. 21 (6) (Jun. 2011) 974–984, https://doi.org/10.1101/gr.114876.110.

[57] M. Babadi, et al., Abstract 2287: Precise common and rare germline CNV calling with GATK, Cancer Res. vol. 78, no. 13_Supplement (Jul. 2018) 2287, https://doi.org/10.1158/1538-7445.AM2018-2287.

[58] G. Klambauer, et al., cn.MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate, Nucleic Acids Res. 40 (9) (May 2012) e69–e, https://doi.org/10.1093/nar/gks003.

[59] E. Bellos, et al., cnvCapSeq: detecting copy number variation in long-range targeted resequencing data, Nucleic Acids Res. 42 (20) (Nov. 2014) e158–e, https://doi.org/10.1093/nar/gku849.

[60] V. Plagnol, et al., A robust model for read count data in exome sequencing experiments and implications for copy number variant calling, Bioinformatics 28 (21) (Nov. 2012) 2747–2754, https://doi.org/10.1093/bioinformatics/bts526.

[61] D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S.L. Salzberg, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions, Genome Biol. 14 (4) (2013) R36, https://doi.org/10.1186/gb-2013-14-4-r36.

[62] D. Kim, J.M. Paggi, C. Park, C. Bennett, S.L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype, Nat. Biotechnol. 37 (8) (Aug. 2019) 907–915, https://doi.org/10.1038/s41587-019-0201-4.

[63] A. Dobin, et al., STAR: ultrafast universal RNA-seq aligner, Bioinformatics 29 (1) (Jan. 2013) 15–21, https://doi.org/10.1093/bioinformatics/bts635.

[64] Y. Liao, G.K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, Bioinformatics 30 (7) (Apr. 2014) 923–930, https://doi.org/10.1093/bioinformatics/btt656.

[65] S. Anders, P.T. Pyl, W. Huber, HTSeq—a Python framework to work with high-throughput sequencing data, Bioinformatics 31 (2) (Jan. 2015) 166–169, https://doi.org/10.1093/bioinformatics/btu638.

[66] R. Patro, G. Duggal, M.I. Love, R.A. Irizarry, C. Kingsford, Salmon provides fast and bias-aware quantification of transcript expression, Nat. Methods 14 (4) (Apr. 2017) 417–419, https://doi.org/10.1038/nmeth.4197.

[67] N.L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification, Nat. Biotechnol. 34 (5) (May 2016) 525–527, https://doi.org/10.1038/nbt.3519.

[68] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (12) (Dec. 2014) 550, https://doi.org/10.1186/s13059-014-0550-8.

[69] M.D. Robinson, D.J. McCarthy, G.K. Smyth, <tt>edgeR</tt> : a Bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics 26 (1) (Jan. 2010) 139–140, https://doi.org/10.1093/bioinformatics/btp616.

[70] G. Dennis, et al., DAVID: Database for Annotation, Visualization, and Integrated Discovery, Genome Biol. 4 (9) (Sep. 2003) R60, https://doi.org/10.1186/gb-2003-4-9-r60.

[71] T. Wu, et al., 'clusterProfiler 4.0: A universal enrichment tool for interpreting omics data', *The*, Innovation 2 (3) (Aug. 2021), 100141, https://doi.org/10.1016/j.xinn.2021.100141.

[72] E.Y. Chen, et al., Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool, BMC Bioinf. 14 (1) (Dec. 2013) 128, https://doi.org/10.1186/1471-2105-14-128.

[73] W. Guo, et al., BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data, BMC Genomics 14 (1) (Dec. 2013) 774, https://doi.org/10.1186/1471-2164-14-774.

[74] F. Krueger, S.R. Andrews, Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications, Bioinformatics 27 (11) (Jun. 2011) 1571–1572, https://doi.org/10.1093/bioinformatics/btr167.

[75] F. Jühling, H. Kretzmer, S.H. Bernhart, C. Otto, P.F. Stadler, S. Hoffmann, metilene: fast and sensitive calling of differentially methylated regions from bisulfite sequencing data, Genome Res. 26 (2) (Feb. 2016) 256–262, https://doi.org/10.1101/gr.196394.115.

[76] A. Akalin, et al., methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles, Genome Biol. 13 (10) (2012) R87, https://doi.org/10.1186/gb-2012-13-10-r87.

[77] Y. Zhang, et al., Model-based Analysis of ChIP-Seq (MACS), Genome Biol. 9 (9) (Nov. 2008) R137, https://doi.org/10.1186/gb-2008-9-9-r137.

[78] 'Spatial Clustering for Identification of ChIP-Enriched Regions (SICER) to Map Regions of Histone Methylation Patterns in Embryonic Stem Cells Stem Cell Transcriptional Networks, 2014, Volume 1150 ISBN : 978-1-4939-0511-9'.

[79] S.A. Ochsner, et al., The Signaling Pathways Project, an integrated 'omics knowledgebase for mammalian cellular signaling pathways, Sci. Data 6 (1) (Oct. 2019) 252, https://doi.org/10.1038/s41597-019-0193-4.

[80] A.R. Quinlan, I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, Bioinformatics 26 (6) (Mar. 2010) 841–842, https://doi.org/10.1093/bioinformatics/btq033.

[81] T.S. Carroll, Z. Liang, R. Salama, R. Stark, I. de Santiago, Impact of artifact removal on ChIP quality metrics in ChIP-seq and ChIP-exo data, Front. Genet. 5 (Apr. 2014), https://doi.org/10.3389/fgene.2014.00075.

[82] S.G. Landt, et al., ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia, Genome Res. 22 (9) (Sep. 2012) 1813–1831, https://doi.org/10.1101/gr.136184.111.

[83] Z. Shao, Y. Zhang, G.-C. Yuan, S.H. Orkin, D.J. Waxman, MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets, Genome Biol. 13 (3) (2012) R16, https://doi.org/10.1186/gb-2012-13-3-r16.

[84] Schweikert, G., Cseke, B., Clouaire, T., Bird, A. and Sanguinetti, G., 2013. MMDiff: quantitative testing for shape changes in ChIP-Seq data sets. BMC genomics, 14(1), pp.1-17.'.

[85] T.L. Bailey, J. Johnson, C.E. Grant, W.S. Noble, The MEME Suite, Nucleic Acids Res. 43 (W1) (Jul. 2015) W39–W49, https://doi.org/10.1093/nar/gkv416.

[86] S. Heinz, et al., Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities, Mol. Cell 38 (4) (May 2010) 576–589, https://doi.org/10.1016/j.molcel.2010.05.004.

[87] A. Medina-Rivera, et al., RSAT 2015: Regulatory Sequence Analysis Tools, Nucleic Acids Res. 43 (W1) (Jul. 2015) W50–W56, https://doi.org/10.1093/nar/gkv362.

[88] P.D. Schloss, et al., Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities, Appl. Environ. Microbiol. 75 (23) (Dec. 2009) 7537–7541, https://doi.org/10.1128/AEM.01541-09.

[89] R.C. Edgar, UPARSE: highly accurate OTU sequences from microbial amplicon reads, Nat. Methods 10 (10) (Oct. 2013) 996–998, https://doi.org/10.1038/nmeth.2604.

[90] P. Menzel, K.L. Ng, A. Krogh, Fast and sensitive taxonomic classification for metagenomics with Kaiju, Nat. Commun. 7 (1) (Apr. 2016) 11257, https://doi.org/10.1038/ncomms11257.

[91] D. McDonald, et al., An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea, ISME J. 6 (3) (Mar. 2012) 610–618, https://doi.org/10.1038/ismej.2011.139.

[92] J.R. Cole, et al., Ribosomal Database Project: data and tools for high throughput rRNA analysis, Nucleic Acids Res. 42 (D1) (Jan. 2014) D633–D642, https://doi.org/10.1093/nar/gkt1244.

[93] D.E. Wood, S.L. Salzberg, Kraken: ultrafast metagenomic sequence classification using exact alignments, Genome Biol. 15 (3) (Mar. 2014) R46, https://doi.org/10.1186/gb-2014-15-3-r46.

[94] R. L. Tatusov, E. V. Koonin, and D. J. Lipman, 'A Genomic Perspective on Protein Families', *Science (1979)*, vol. 278, no. 5338, pp. 631–637, Oct. 1997, doi: 10.1126/science.278.5338.631.

[95] A. Blanco-Míguez, et al., Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4, Nat. Biotechnol. (Feb. 2023), https://doi.org/10.1038/s41587-023-01688-w.

[96] S. Nguengang Wakap, et al., Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database, Eur. J. Hum. Genet. 28 (2) (Feb. 2020) 165–173, https://doi.org/10.1038/s41431-019-0508-0.

[97] E. Turro, et al., Whole-genome sequencing of patients with rare diseases in a national health system, Nature 583 (7814) (Jul. 2020) 96–102, https://doi.org/10.1038/s41586-020-2434-2.

[98] K.J. Burdick, et al., Limitations of exome sequencing in detecting rare and undiagnosed diseases, Am. J. Med. Genet. A 182 (6) (Jun. 2020) 1400–1406, https://doi.org/10.1002/ajmg.a.61558.

[99] K. Freson, et al., Platelet characteristics in patients with X-linked macrothrombocytopenia because of a novel GATA1mutation, Blood 98 (1) (Jul. 2001) 85–92, https://doi.org/10.1182/blood.V98.1.85.

[100] D.R. Adams, C.M. Eng, Next-Generation Sequencing to Diagnose Suspected Genetic Disorders, N. Engl. J. Med. 379 (14) (Oct. 2018) 1353–1362, https://doi.org/10.1056/NEJMra1711801.

[101] B.J. O'Roak, et al., Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations, Nature 485 (7397) (May 2012) 246–250, https://doi.org/10.1038/nature10989.

[102] J.R. Lupski, et al., Whole-Genome Sequencing in a Patient with Charcot–Marie–Tooth Neuropathy, N. Engl. J. Med. 362 (13) (Apr. 2010) 1181–1191, https://doi.org/10.1056/NEJMoa0908094.

[103] 'Large-scale discovery of novel genetic causes of developmental disorders', *Nature*, vol. 519, no. 7542, pp. 223–228, Mar. 2015, doi: 10.1038/nature14135.

[104] K. Retterer, et al., Clinical application of whole-exome sequencing across clinical indications, Genet. Med. 18 (7) (Jul. 2016) 696–704, https://doi.org/10.1038/gim.2015.148.

[105] K. Splinter, et al., Effect of Genetic Diagnosis on Patients with Previously Undiagnosed Disease, N. Engl. J. Med. 379 (22) (Nov. 2018) 2131–2139, https://doi.org/10.1056/NEJMoa1714458.

[106] W.D. Jones, et al., De Novo Mutations in MLL Cause Wiedemann-Steiner Syndrome, Am. J. Hum. Genet. 91 (2) (Aug. 2012) 358–364, https://doi.org/10.1016/j.ajhg.2012.06.008.

[107] A.C. Thomas, et al., Mosaic Activating Mutations in GNA11 and GNAQ Are Associated with Phakomatosis Pigmentovascularis and Extensive Dermal Melanocytosis, J, Invest. Dermatol. 136 (4) (Apr. 2016) 770–778, https://doi.org/10.1016/j.jid.2015.11.027.

[108] C.F. Wright, et al., Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data, Lancet 385 (9975) (Apr. 2015) 1305–1314, https://doi.org/10.1016/S0140-6736(14)61705-0.

[109] S.B. Ng, et al., Exome sequencing identifies the cause of a mendelian disorder, Nat. Genet. 42 (1) (Jan. 2010) 30–35, https://doi.org/10.1038/ng.499.

[110] S. Srivastava, et al., Clinical whole exome sequencing in child neurology practice, Ann. Neurol. 76 (4) (Oct. 2014) 473–483, https://doi.org/10.1002/ana.24251.

[111] S.B. Ng, et al., Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome, Nat. Genet. 42 (9) (Sep. 2010) 790–793, https://doi.org/10.1038/ng.646.

[112] L.E.L.M. Vissers, et al., Mutations in a new member of the chromodomain gene family cause CHARGE syndrome, Nat. Genet. 36 (9) (Sep. 2004) 955–957, https://doi.org/10.1038/ng1407.

[113] R. Zhang, et al., Whole exome sequencing identified a homozygous novel mutation in SUOX gene causes extremely rare autosomal recessive isolated sulfite oxidase deficiency, Clin. Chim. Acta 532 (Jul. 2022) 115–122, https://doi.org/10.1016/j.cca.2022.06.005.

[114] S. Banerjee *et al.*, 'Comparative analysis of clonal evolution among patients with right- and left-sided colon and rectal cancer', *iScience*, vol. 24, no. 7, p. 102718, Jul. 2021, doi: 10.1016/j.isci.2021.102718.

[115] R. Zhang, et al., Whole exome sequencing identified a homozygous novel variant in *CEP290* gene causes Meckel syndrome, J. Cell Mol. Med. 24 (2) (Jan. 2020) 1906–1916, https://doi.org/10.1111/jcmm.14887.

[116] Y. Dai, et al., Whole exome sequencing identified a novel *DAG1* mutation in a patient with rare, mild and late age of onset muscular dystrophy-dystroglycanopathy, J. Cell Mol. Med. 23 (2) (Feb. 2019) 811–818, https://doi.org/10.1111/jcmm.13979.

[117] Z. Zhang, et al., A novel pathogenic single nucleotide germline deletion in APC gene in a four generation Chinese family with familial adenomatous polyposis, Sci. Rep. 7 (1) (Sep. 2017) 12357, https://doi.org/10.1038/s41598-017-10395-x.

[118] Y. Zheng, J. Xu, S. Liang, D. Lin, S. Banerjee, Whole Exome Sequencing Identified a Novel Heterozygous Mutation in HMBS Gene in a Chinese Patient With Acute Intermittent Porphyria With Rare Type of Mild Anemia, Front. Genet. 9 (Apr. 2018), https://doi.org/10.3389/fgene.2018.00129.

[119] G.R. Oliver, et al., RNA-Seq detects a SAMD12-EXT1 fusion transcript and leads to the discovery of an EXT1 deletion in a child with multiple osteochondromas, Mol. Genet. Genomic Med. 7 (3) (Mar. 2019) e00560.

[120] N. Sobreira, F. Schiettecatte, D. Valle, A. Hamosh, GeneMatcher: A Matching Tool for Connecting Investigators with an Interest in the Same Gene, Hum. Mutat. 36 (10) (Oct. 2015) 928–930, https://doi.org/10.1002/humu.22844.

[121] D.J. Carey, et al., The Geisinger MyCode community health initiative: an electronic health record–linked biobank for precision medicine research, Genet. Med. 18 (9) (Sep. 2016) 906–913, https://doi.org/10.1038/GIM.2015.187.

[122] S. Ding, L. Han, Newborn screening for genetic disorders: Current status and prospects for the future, Pediatr Investig 6 (4) (Dec. 2022) 291–298, https://doi.org/10.1002/ped4.12343.

[123] M.E. Norton, et al., Non-Invasive Chromosomal Evaluation (NICE) Study: results of a multicenter prospective cohort study for detection of fetal trisomy 21 and trisomy 18, Am. J. Obstet. Gynecol. 207 (2) (Aug. 2012) 137.e1–137.e8, https://doi.org/10.1016/j.ajog.2012.05.021.

[124] C. Sullivan-Pyke, A. Dokras, Preimplantation Genetic Screening and Preimplantation Genetic Diagnosis, Obstet. Gynecol. Clin. N. Am. 45 (1) (Mar. 2018) 113–125, https://doi.org/10.1016/j.ogc.2017.10.009.

[125] 'A review of pre-implantation genetic testing technologies and applications Ren-Qian Du renqian.du@gmail.com, Ding-Ding Zhao, Kai Kang, Fan Wang, Rui-Xia Xu, Chun-Li Chi, Ling-Yin Kong, and Bo Liang boliang880@alumni.sjtu.edu.cn'.

[126] M. Xiao, et al., Next-Generation Sequencing Is More Efficient at Detecting Mosaic Embryos and Improving Pregnancy Outcomes than Single-Nucleotide Polymorphism Array Analysis, J. Mol. Diagn. 23 (6) (Jun. 2021) 710–718, https://doi.org/10.1016/j.jmoldx.2021.02.011.

[127] S. Miller, C. Chiu, The Role of Metagenomics and Next-Generation Sequencing in Infectious Disease Diagnosis, Clin. Chem. 68 (1) (Dec. 2021) 115–124, https://doi.org/10.1093/clinchem/hvab173.

[128] S. Zeng, W. Yan, X. Wu, H. Zhang, Case Report: Diagnosis of Klebsiella pneumoniae Invasive Liver Abscess Syndrome With Purulent Meningitis in a Patient From Pathogen to Lesions, Front Med (lausanne) 8 (Sep. 2021), https://doi.org/10.3389/fmed.2021.714916.

[129] C. Michel, et al., Case Report: About a Case of Hyperammonemia Syndrome Following Lung Transplantation: Could Metagenomic Next-Generation Sequencing Improve the Clinical Management? Front Med (lausanne) 8 (Jul. 2021) https://doi.org/10.3389/fmed.2021.684040.

[130] Q. Yan, et al., Cultivation and Genomic Characterization of the Bile Bacterial Species From Cholecystitis Patients, Front. Microbiol. 12 (Nov. 2021), https://doi.org/10.3389/fmicb.2021.739621.

[131] J. Quick, et al., Real-time, portable genome sequencing for Ebola surveillance, Nature 530 (7589) (Feb. 2016) 228–232, https://doi.org/10.1038/nature16996.

[132] P. Zhou, et al., A pneumonia outbreak associated with a new coronavirus of probable bat origin, Nature 579 (7798) (Mar. 2020) 270–273, https://doi.org/10.1038/s41586-020-2012-7.

[133] G.K. Paterson, et al., Capturing the cloud of diversity reveals complexity and heterogeneity of MRSA carriage, infection and transmission, Nat. Commun. 6 (1) (Mar. 2015) 6560, https://doi.org/10.1038/ncomms7560.

[134] L.S.A. Mühr, D. Guerendiain, K. Cuschieri, K. Sundström, Human Papillomavirus Detection by Whole-Genome Next-Generation Sequencing: Importance of Validation and Quality Assurance Procedures, Viruses 13 (7) (Jul. 2021) 1323, https://doi.org/10.3390/v13071323.

[135] K. Schmidt, et al., Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic

sequencing, J. Antimicrob. Chemother. 72 (1) (Jan. 2017) 104–114, https://doi.org/10.1093/jac/dkw397.
[136] I. De Vlaminck, et al., Temporal Response of the Human Virome to Immunosuppression and Antiviral Therapy, Cell 155 (5) (Nov. 2013) 1178–1187, https://doi.org/10.1016/j.cell.2013.10.034.
[137] M.E. Török, et al., Rapid Whole-Genome Sequencing for Investigation of a Suspected Tuberculosis Outbreak, J. Clin. Microbiol. 51 (2) (Feb. 2013) 611–614, https://doi.org/10.1128/JCM.02279-12.
[138] D. Repana, et al., The Network of Cancer Genes (NCG): a comprehensive catalogue of known and candidate cancer genes from cancer sequencing screens, Genome Biol. 20 (1) (Dec. 2019) 1, https://doi.org/10.1186/s13059-018-1612-0.
[139] N. Tung, et al., Frequency of mutations in individuals with breast cancer referred for *BRCA 1* and *BRCA 2* testing using next-generation sequencing with a 25-gene panel, Cancer 121 (1) (Jan. 2015) 25–33, https://doi.org/10.1002/cncr.29010.
[140] A.G. Nikitin, et al., Lynch Syndrome Germline Mutations in Breast Cancer: Next Generation Sequencing Case-Control Study of 1,263 Participants, Front. Oncol. 10 (May 2020), https://doi.org/10.3389/fonc.2020.00666.
[141] Y. Wang, et al., Changing Technologies of RNA Sequencing and Their Applications in Clinical Oncology, Front. Oncol. 10 (Apr. 2020), https://doi.org/10.3389/fonc.2020.00447.
[142] E. Cuppen, et al., 'Implementation of Whole-Genome and Transcriptome Sequencing Into Clinical Cancer Care', *JCO Precis*, Oncol no. 6 (2022) Dec, https://doi.org/10.1200/PO.22.00245.
[143] E.A. Zhigalova, et al., RNA-Seq-Based TCR Profiling Reveals Persistently Increased Intratumoral Clonality in Responders to Anti-PD-1 Therapy, Front. Oncol. 10 (Apr. 2020), https://doi.org/10.3389/fonc.2020.00385.
[144] E. Poddubskaya, et al., Transcriptomics-Guided Personalized Prescription of Targeted Therapeutics for Metastatic ALK-Positive Lung Cancer Case Following Recurrence on ALK Inhibitors, Front. Oncol. 9 (Oct. 2019), https://doi.org/10.3389/fonc.2019.01026.
[145] L. Tang, et al., Identification and Validation of Potential Pathogenic Genes and Prognostic Markers in ESCC by Integrated Bioinformatics Analysis, Front. Genet. 11 (Dec. 2020), https://doi.org/10.3389/fgene.2020.521004.
[146] E. Papaemmanuil, et al., Genomic Classification and Prognosis in Acute Myeloid Leukemia, N. Engl. J. Med. 374 (23) (Jun. 2016) 2209–2221, https://doi.org/10.1056/NEJMoa1516192.
[147] B.C. Medeiros, A.T. Fathi, C.D. DiNardo, D.A. Pollyea, S.M. Chan, R. Swords, Isocitrate dehydrogenase mutations in myeloid malignancies, Leukemia 31 (2) (Feb. 2017) 272–281, https://doi.org/10.1038/leu.2016.275.
[148] A. Drilon, et al., Entrectinib in ROS1 fusion-positive non-small-cell lung cancer: integrated analysis of three phase 1–2 trials, Lancet Oncol. 21 (2) (Feb. 2020) 261–270, https://doi.org/10.1016/S1470-2045(19)30690-4.
[149] R.C. Doebele, et al., Entrectinib in patients with advanced or metastatic NTRK fusion-positive solid tumours: integrated analysis of three phase 1–2 trials, Lancet Oncol. 21 (2) (Feb. 2020) 271–282, https://doi.org/10.1016/S1470-2045(19)30691-6.
[150] L. J. van 't Veer *et al.*, 'Gene expression profiling predicts clinical outcome of breast cancer', *Nature*, vol. 415, no. 6871, pp. 530–536, Jan. 2002, doi: 10.1038/415530a.
[151] C. Desmedt, T. Voet, C. Sotiriou, P.J. Campbell, Next-generation sequencing in breast cancer, Curr. Opin. Oncol. 24 (6) (Nov. 2012) 597–604, https://doi.org/10.1097/CCO.0b013e328359554e.
[152] S.N. Lone, et al., Liquid biopsy: a step closer to transform diagnosis, prognosis and future of cancer treatments, Mol. Cancer 21 (1) (Mar. 2022) 79, https://doi.org/10.1186/s12943-022-01543-7.
[153] S. Bach, et al., 'Circulating Tumor DNA Analysis: Clinical Implications for Colorectal Cancer Patients, A Systematic Review', *JNCI Cancer Spectr* 3 (3) (Sep. 2019), https://doi.org/10.1093/jncics/pkz042.
[154] J. D. Cohen *et al.*, 'Detection and localization of surgically resectable cancers with a multi-analyte blood test', *Science (1979)*, vol. 359, no. 6378, pp. 926–930, Feb. 2018, doi: 10.1126/science.aar3247.
[155] R.-I. Chin, et al., Detection of Solid Tumor Molecular Residual Disease (MRD) Using Circulating Tumor DNA (ctDNA), Mol. Diagn. Ther. 23 (3) (Jun. 2019) 311–331, https://doi.org/10.1007/s40291-019-00390-5.
[156] D. Wu, et al., Detection of Minimal Residual Disease in B Lymphoblastic Leukemia by High-Throughput Sequencing of *IGH*, Clin. Cancer Res. 20 (17) (Sep. 2014) 4540–4548, https://doi.org/10.1158/1078-0432.CCR-13-3231.
[157] M.F. Berger, E.R. Mardis, The emerging clinical relevance of genomics in cancer medicine, Nat. Rev. Clin. Oncol. 15 (6) (Jun. 2018) 353–365, https://doi.org/10.1038/s41571-018-0002-6.
[158] M. Brodlie, I.J. Haq, K. Roberts, J.S. Elborn, Targeted therapies to improve CFTR function in cystic fibrosis, Genome Med. 7 (1) (Dec. 2015) 101, https://doi.org/10.1186/s13073-015-0223-6.
[159] I. Schlam, S.M. Swain, HER2-positive breast cancer and tyrosine kinase inhibitors: the time is now, npj Breast Cancer 7 (1) (May 2021) 56, https://doi.org/10.1038/s41523-021-00265-1.
[160] Y.J. Cha, J.S. Koo, Next-generation sequencing in thyroid cancer, J. Transl. Med. 14 (1) (Dec. 2016) 322, https://doi.org/10.1186/s12967-016-1074-7.
[161] C. Cainap, O. Balacescu, S.S. Cainap, L.-A. Pop, Next Generation Sequencing Technology in Lung Cancer Diagnosis, Biology (basel) 10 (9) (Sep. 2021) 864, https://doi.org/10.3390/biology10090864.
[162] U. Testa, G. Castelli, E. Pelosi, Breast Cancer: A Molecularly Heterogenous Disease Needing Subtype-Specific Treatments, Medical Sciences 8 (1) (Mar. 2020) 18, https://doi.org/10.3390/medsci8010018.
[163] S. Mukherjee, Genomics-Guided Immunotherapy for Precision Medicine in Cancer, Cancer Biother. Radiopharm. 34 (8) (Oct. 2019) 487–497, https://doi.org/10.1089/cbr.2018.2758.
[164] M.H. Eckman, Cost-Effectiveness of Using Pharmacogenetic Information in Warfarin Dosing for Patients With Nonvalvular Atrial Fibrillation, Ann. Intern. Med. 150 (2) (Jan. 2009) 73, https://doi.org/10.7326/0003-4819-150-2-200901200-00005.
[165] F.S. Collins, H. Varmus, A New Initiative on Precision Medicine, N. Engl. J. Med. 372 (9) (Feb. 2015) 793–795, https://doi.org/10.1056/NEJMp1500523.
[166] N. Anaparthy, Y.-J. Ho, L. Martelotto, M. Hammell, J. Hicks, Single-Cell Applications of Next-Generation Sequencing, Cold Spring Harb. Perspect. Med. 9 (10) (Oct. 2019), a026898, https://doi.org/10.1101/cshperspect.a026898.
[167] M. Chen, H. Zhao, Next-generation sequencing in liquid biopsy: cancer screening and early detection, Hum. Genomics 13 (1) (Dec. 2019) 34, https://doi.org/10.1186/s40246-019-0220-8.