

From DNA to Data: Transcending Beyond the Double Helix and Demystifying the Genetic Alchemy of Life Through NGS to Empower Precision Medicine

Muhammed Ali Siham H R^{1,2*}, Ashwin Prabahar A³, Sriprata R³, Sandra Nixon³, Sandhiya P³, Gnana Sowndariyan G³ and Palak Bhataria²

Abstract

Next-generation sequencing (NGS) has emerged as a transformative technology within the field of genomics, revolutionizing the ability to unravel the intricacies of genetic information. The advent of NGS platforms has facilitated rapid, high-throughput sequencing, enabling the analysis of entire genomes, transcriptomes, and epigenomes. This powerful technology has paved the way for groundbreaking advancements in diverse areas, including personalized medicine, evolutionary biology, and agricultural genomics. By leveraging massively parallel sequencing, NGS has significantly reduced the cost and time required for sequencing, making it accessible to a wide range of researchers and clinicians. Moreover, it has facilitated the identification of genetic variations, such as single nucleotide polymorphisms (SNPs) and structural variants, enabling comprehensive exploration of genetic diversity and disease mechanisms. Despite these remarkable achievements, challenges remain, including data management, analysis, and interpretation. The vast amounts of data generated necessitate sophisticated bioinformatics tools and computational pipelines for accurate alignment, variant calling, and functional annotation. Furthermore, ethical considerations, such as data privacy and consent, pose additional complexities that need to be addressed. In this paper, we provide a comprehensive overview of NGS technologies, applications, challenges, and future prospects, highlighting its immense potential to advance the understanding of genomics and drive transformative discoveries.

Keywords: Next-generation sequencing; Genomics; High-throughput sequencing; Personalized medicine; Genetic variations; Bioinformatics; Data analysis; Computational pipelines.

¹Chettinad Academy of Research and Education, India

²Garden City University, India

³Bharath Institute of Higher Education and Research, India

*Corresponding Author: Muhammed Ali Siham H R, Garden City University, Bengaluru, Karnataka, India.

Received Date: 07-14-2023

Accepted Date: 07-31-2023

Published Date: 08-25-2023

Copyright© 2023 by H R MAS, et al. All rights reserved. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Introduction

A timeline of sequencing technologies and methods: From sanger to next-generation sequencing and beyond

First-generation sequencing, also known as Sanger sequencing, utilizing the chain termination method, was initially established in 1977. It was followed by sequencing through the Chemical degradation method developed by Maxam and Gilbert. In the same year, Sanger successfully determined the 5368 bp genome of phage Φ x174, marking the inception of DNA genome sequencing. The advent of high-throughput sequencing platforms in 2005 paved the way for various Next-Generation Sequencing (NGS) platforms, each with their own accuracy and reproducibility factors influenced by analysis pipelines and platform-specific features [1]. Around 2016, Ion Torrent sequencing exhibited superior sensitivity over pyrosequencing, despite both operating on the "sequencing by synthesis" principle. Another technique, SOLiD (sequencing by oligonucleotide ligation and detection), showcased high accuracy due to the occurrence of each base being expressed twice, albeit limited by shorter read lengths. Subsequently, DNBS (DNA nanoball sequencing) emerged as a method enabling simultaneous sequencing of a vast collection of DNA nanoballs. Illumina-based sequencing employs the "reversible terminator sequencing" methodology. In comparison to microarray technology, High-Throughput Sequencing, or NGS, offers several advantages, such as reduced sequencing costs, identification of unknown sequences, and enhanced speed and accuracy

[2]. More recently, alongside NGS, which provides high accuracy, cost-efficiency, and expeditious results, third-generation sequencing technologies have surfaced for long-read sequencing of individual RNAs. These techniques circumvent the drawbacks associated with PCR amplification and read mapping while significantly reducing the false positive rate of splice sites, thereby meticulously capturing transcript isoform diversity. Single-molecule RNA sequencing methodologies enable the generation of full-length cDNA transcripts without clonal amplification or transcript assembly requirements. Diverse single-molecule sequencing platforms have been developed, including Pacific Biosciences (PacBio), single-molecule real-time (SMRT) sequencing, Helicos single-molecule fluorescent sequencing, and Oxford Nanopore Technology (ONT). Single-cell RNA sequencing was initially established in 2009 to profile transcriptomes at a single-cell resolution. In 2015, Drop-seq and InDrop methodologies were introduced for the analysis of mouse retinal cell and embryonic stem cell transcriptomes, respectively. To address the loss of spatial information during single-cell isolation, spatial transcriptomics emerged in 2016, leveraging positional barcodes to visualize RNA distributions within tissue sections [3]. In subsequent years, methodologies such as Sci-RNA-seq, single-cell combinatorial indexing RNA sequencing, Geo-seq, and integrated scRNA-seq, which employs laser capture microdissection for individual cell isolation, were developed in 2017. In 2018, split-pool ligation-based transcriptome sequencing (SPLiT-seq) was discovered, wherein both

Sci-RNA-seq and SPLiT-seq adopt a combinatorial indexing strategy involving barcode labeling of RNAs to indicate their cellular origin. The year 2019 saw the emergence of Slide-seq, utilizing DNA barcode beads with positional information. Recent advancements in RNA sequencing have led to the development of specialized technologies catering to specific applications. For example, CaptureSeq, a targeted RNA

sequencing approach, utilizes biotinylated oligonucleotide probes to enrich specific transcripts for gene fusion identification. In situ amplification by rolling-circle amplification (RCA) and fluorescent in situ RNA sequencing (FISSEQ) have been devised to enable targeted sequencing of RNA fragments within morphologically preserved tissues or cells, eliminating the need for RNA extraction.

S.no	Sequencing Technique	Year of successful emergence
1	Genome sequencing	1977
2	Whole Genome sequencing	1995
3	Transcriptome sequencing	Early 1990s
4	Bulk RNA sequencing	Mid 2000s
5	Single cell RNA sequencing	2009
6	Exome sequencing	Mid 2000s
7	Whole Exome sequencing	2009
8	ATAC-seq	2013
9	ChIP-seq	2007
10	MeDIP-seq	2005

Table 1: Decades of Discovery: Evolution of Sequencing Techniques (1977-2013).

Navigating the genomic frontier: Key tools used in genome sequencing

DNA extraction kits

They serve as indispensable tools in the field of genome sequencing, enabling the efficient isolation and purification of DNA samples. These kits utilize the alkaline lysis procedure, a widely employed method in molecular biology, to break down cellular membranes and release genomic DNA.

The process involves subjecting the starting material to alkaline conditions, which disrupt the cells and denature proteins, allowing the DNA to be liberated [4].

One of the key advantages of DNA extraction kits utilizing the alkaline lysis system is the rapid and efficient purification of DNA. This method ensures the removal of cellular debris and contaminants, resulting in high-quality DNA samples suitable for downstream applications such as PCR, sequencing, cloning, and other molecular biology techniques.

Moreover, these kits offer flexibility in handling various quantities of starting materials, allowing researchers and clinicians to process samples of different sizes. Exemplifying the advancements in DNA extraction kits, InstaGene Matrix provides

PCR-ready DNA, streamlining the process of preparing DNA samples for PCR amplification. Another notable example is the utilization of Chelex 100 Resin, which offers efficient DNA purification by chelating metal ions and removing impurities from the DNA sample.

The machinery behind DNA sequencing

Sequencing machines and analyzers play a pivotal role in automating the DNA sequencing process, enabling the precise determination of the sequence order of the four DNA bases: adenine (A), thymine (T), guanine (G), and cytosine (C). Certain sequencers function as optical instruments that scrutinize the light signals emitted by fluorochromes attached to nucleotides. Notably, Lloyd M Smith pioneered the development of the first automated DNA sequences, which revolutionized the Sanger Sequencing system. These sequencers find applications in genotyping studies, particularly for analyzing heritable markers and determining DNA fragment lengths [5]. Third-generation DNA sequencers, including PacBio, SMRT, and Oxford Nanopore bring with them some innovative techniques to measure the sequential addition of nucleotides to single DNA molecules. The selection of a DNA sequencer and system relies on the specific experimental requirements and available budget, considering various DNA sequencing methodologies. Automated DNA sequencers or analyzers are well-suited for sequencing DNA and examining DNA fragments for diverse purposes. Capillary electrophoresis-based systems enable DNA fragments to migrate through a polymer matrix, with the

resulting fluorescence signals measured. The use of multiple capillaries allows for efficient loading of samples in a 96-well microplate format. Pyrosequencing technology uses alternative analyzers for rapid sequencing, offering comparable capabilities to Sanger sequencing. These instruments are highly suitable for applications such as genotyping, mutation analysis, and single nucleotide polymorphism (SNP) profiling. Selecting an optimal sequencer involves careful consideration of factors such as desired functionality and throughput requirements in the study of heritable traits.

Leveraging bioinformatics software solutions for NGS

Bioinformatics software assumes a critical role in the intricate landscape of genome sequencing and analysis, empowering scientists to unravel the mysteries encoded within genetic blueprints. Noteworthy exemplars include Splice Center, Phred, DDIALIGN, BioGPS, and Genome Browser, among a multitude of others. SpliceCenter facilitates the estimation and characterization of gene-splicing variations, shedding light on their potential functional consequences [6]. Phred diligently scrutinizes DNA sequence chromatogram lines, dissecting peaks to unveil precise quality scores for each individual base call, ensuring data accuracy and reliability. DDIALIGN emerges as a versatile tool, enabling comprehensive analysis grounded on the intricacies of genome sequences, unraveling intricate patterns and associations. Serving as a reliable gateway to explore the multifaceted world of genome, gene, and protein functions, BioGPS offers a customizable gene

report layout, empowering researchers to extract tailored insights from vast datasets. In the area of genome assembly, sophisticated software solutions take center stage. Genome assembly software programs expertly map additional sequences and unveil genomic variations, providing a comprehensive summary of statistical data and facilitating comparisons against gold standard reference assemblies. PBJelly takes advantage of the scaffolding approach, utilizing SMRT long reads to bridge sequence gaps in genome assembly, while HAGAP emerges as a powerful tool for de novo assembly. Drawing upon computationally optimized components, FALCON emerges as a leading-edge haplotype genome assembly tool, illuminating the complexities of genetic variation [7].

Genome annotation software, with its diverse repertoire, holds immense value in deciphering the intricacies of genomic elements. Notable members of this software cadre include tRNAScanSE, which meticulously predicts tRNA genes across entire genomes, and RNAmmer, a specialized tool for annotating rRNA species across diverse kingdoms of organisms. Prodigal assumes the mantle of annotating microbial genomes, while GenMark stands as a versatile gene prediction tool catering to prokaryotes, eukaryotes, viruses, phages, plasmids, and transcripts. By capitalizing on anonymous genomic sequences of varying lengths, Metagene Annotator emerges as a key resource for precise gene prediction in prokaryotes, showcasing its extensive utility in microbial genome studies and genome annotation endeavors.

Whole genome sequencing (WGS)

Whole-genome sequencing (WGS) has emerged as a powerful tool for an in-depth analysis of an organism's genetic information, spanning both coding and non-coding regions. This technique entails the determination of the precise order of nucleotide bases, namely adenine, guanine, cytosine, and thymine that constitute an individual's genetic code. By making use of advanced technologies like next-generation sequencing, it enables the identification of genetic variations and mutations that may be implicated in various diseases comprising cancer, genetic disorders, and inherited conditions. It offers invaluable insights into the genetic underpinnings of diseases, thereby facilitating the development of personalized treatment strategies [15].

However, ethical and privacy concerns regarding the utilization and storage of individuals' genetic information remain pertinent. The analysis involves a multifaceted procedure aimed at determining the complete DNA sequence of an organism's genome. It relies on high-throughput DNA sequencing technologies to sequence the entire genome, thereby obtaining a systematic understanding of its genetic makeup and identifying relevant genetic variations associated with specific traits or diseases. The steps involved in WGS constitute DNA extraction, library preparation, sequencing, and subsequent data analysis. DNA extraction calls for the isolation of DNA from a sample, typically derived from blood, saliva, or tissue. Library preparation involves the fragmentation of

DNA, attachment of adapters, and amplification to generate a library of fragments amenable to sequencing. Subsequently, high-throughput sequencing technology is utilized to determine the sequence of each fragment within the library. The resulting sequence reads are then aligned to a reference genome during data analysis, allowing for the identification of genetic variations and the interpretation of findings [16-18]. However, it is important to note that WGS generates prodigious amounts of data, thereby requiring the utilization of advanced bioinformatics tools and computational resources to ensure efficient analysis and interpretation. Analysis of such large genomic data holds immense potential in the field of medicine, offering diverse applications ranging from the precise diagnosis of genetic disorders to the implementation of personalized medicine approaches and the evaluation of disease susceptibility. Moreover, it also serves as an indispensable part of scientific research, enabling in-depth exploration of disease genetics and the elucidation of novel therapeutic targets.

On the flip side, WGS also has several limitations that need to be considered. Firstly, its cost is relatively high, limiting its accessibility in resource-limited settings. Additionally, it generates massive amounts of data, requiring substantial computational resources for effective storage and management [19]. Interpreting the vast genetic information obtained can be challenging, particularly in distinguishing clinically relevant genetic variations from benign variants. Furthermore, there is very limited understanding of non-coding regions,

which play crucial roles in gene regulation and disease development. In addition to all this, false positives and false negatives can also occur in its results, leading to potential misdiagnoses or missed diagnoses. Technical limitations, such as incomplete genomic coverage and challenges in identifying certain genetic variations, exist as well. It is equally important to note the ethical and social implications that arise from data privacy, information sharing, and genetic discrimination. Therefore, it is important to be aware of these limitations to appropriately interpret its results and make informed decisions in clinical and research applications. Ongoing advancements are focused on addressing these limitations and further enhancing its utility.

Practical applications of genome sequencing

Genome sequencing, a transformative technology with profound implications in genetics, finds diverse applications in environmental science, agriculture, and other scientific domains. This groundbreaking process involves the meticulous determination of nucleotide sequences (A, T, C, and G) within an organism's DNA.

Commencing with the acquisition of a DNA sample from the target organism, the purification step is put into action to isolate and refine the genetic material, which is subsequently fragmented to generate a comprehensive library of sequences. The actual sequencing of the DNA is achieved through advanced techniques like single-molecule real-time sequencing or reversible terminator sequencing. The ensuing raw

sequence data undergoes rigorous quality control measures and filtering procedures, culminating in the assembly of these sequences to reconstruct the entire genome. Bioinformatics tools play an instrumental role in the subsequent analysis, enabling the identification of genes and functional elements within the genome. These sophisticated computational tools empower researchers to unravel the intricate genetic architecture and gain a deeper understanding of the organism's biological functions. Through the utilization of these tools, scientists can decipher disease-causing mutations, unravel evolutionary relationships by comparing genomes, and delve into the intricate workings of specific genes. The seamless facilitation of genomics, bioinformatics, and other innovative research methodologies facilitates an extensive exploration of an organism's genetic makeup, propelling breakthroughs across diverse scientific disciplines.

Transcriptome sequencing

RNA sequencing, also known as transcriptome sequencing, is a powerful analytical method exercised to unveil the presence and quality of RNA molecules within a biological sample, providing valuable insights into their developmental stage or physical condition. This technique allows for the comprehensive analysis of the continuous and intricate alterations in the transcriptome, thereby deciphering the functional elements of the genome [8,9]. Through this sequencing, the molecular constituents of cells and tissues can be elucidated, facilitating an enhanced understanding of developmental processes and disease

pathogenesis. In various cell types, different combinations of genes are switched on or off, resulting in a variety of structures and functions observed.

Among the numerous next-generation sequencing (NGS) techniques available, RNA sequencing stands out as the most widely referenced approach for quantifying and characterizing RNA transcripts. Two fundamental variations of this technique involve the use of either random primers or oligo(dt) primers during the sequencing reaction. Oligo(dt) primers, with their pronounced 3'-end bias, are particularly suited for analyzing mRNA abundance, while the introduction of bias can be mitigated by fragmenting the input RNA when using random primers.

RNA sequencing also serves as a highly sensitive and accurate tool for measuring gene expression across the entire transcriptome, facilitating the detection of previously undetected changes that occur in various disease states, under diverse environmental conditions, in response to therapeutics, and across a broad range of experimental designs. Each RNA sequence corresponds to the DNA sequence from which it was transcribed [10].

Thus, by examining the transcriptome, we can ascertain the precise timing and location of gene activation or inactivation within the cells and organs of an organism. One of its major advantages lies in its ability to identify a wide range of molecular features without constraints imposed by prior information, encompassing well-known properties as well as novel transcript isoforms, gene fusions,

single nucleotide variants, and other distinctive characteristics.

Notably, over 95% of the published RNA-seq data available in the Short Read Archive (SRA) has been generated using the Illumina short-read sequencing method, which has become the standard technology for RNA sequencing. However, as researchers seek approaches that can provide superior isoform-level data, long-read cDNA sequencing and emerging dRNA-seq technologies may soon challenge its dominance. The field of transcriptomics has experienced a profound revolution owing to the advent of high-throughput next-generation sequencing (NGS), which enables the high throughput analysis of RNA through complementary DNA (cDNA) sequencing. This groundbreaking technique, known as RNA sequencing (RNA-Seq), has completely transformed the knowledge of the intricate and dynamic nature of the transcriptome [11]. With RNA-Seq, gene expression, alternative splicing events, and allele-specific expression can be better understood and quantified. Recent advancements in RNA-Seq methodology, spanning sample preparation, sequencing platforms, and bioinformatics data interpretation, have enabled deep profiling of the transcriptome, giving more insights into numerous physiological and pathological conditions.

Compared to earlier techniques such as Sanger sequencing and microarray-based approaches, RNA-Seq offers significantly broader coverage and improved resolution of the dynamic nature of the transcriptome. The data generated by RNA-Seq not only facilitates the identification of alternatively

spliced genes but also enables the discovery of novel transcripts and the detection of allele-specific expressions. Moreover, it allows for accurate quantification of gene expression, opening avenues for extensive transcriptomic investigations.

Bulk-RNA sequencing

Bulk RNA sequencing, also known as RNA-Seq, is a robust and widely utilized methodology for assessing the expression levels of genes within a biological sample. It provides a comprehensive and holistic view of the entire transcriptome at a given time point, enabling the identification of novel transcripts, alternative splicing events, and non-coding RNA molecules [12]. This powerful technique has found extensive applications in diverse scientific disciplines, including genomics, molecular biology, and biomedical research. To ensure the generation of high-quality and physiologically relevant data in bulk differential gene expression (DGE) RNA-seq studies, meticulous experimental design is crucial. Factors such as the level of replication, sequencing read depth, and the utilization of single- or paired-end sequencing reads must be carefully considered. The term "bulk" in this context refers to the entirety of RNA derived from a population of cells, facilitating comprehensive examination and analysis. Consequently, bulk sequencing methodologies enable the assessment of all molecular constituents comprising the transcriptome. Interestingly, the total RNA pool encompassing ribosomal RNA (rRNA), pre-mRNA, and various classes of non-coding RNA (ncRNA) can be subjected to

sequencing, or specific RNA types can be selectively depleted or enriched prior to or during library construction [13,14]. Numerous techniques have been developed to achieve the targeted removal or enrichment of specific RNA molecules, tailored to different RNA types and starting materials. Bulk RNA-Seq represents a broad term encompassing sequencing methodologies that leverage averaged gene expression from a population of cells to determine the presence and abundance of RNA in a given sample. Consequently, bulk-based approaches enable the discrimination of diverse sample conditions. While careful consideration of various factors is essential to ensure the generation of high-quality data in bulk sequencing experiments, this method is not excessively constrained by technical applications in practice.

Single cell RNA sequencing: exploring cellular diversity

Recent therapeutic trends have prioritized specific targeting and personalization to cater to the diverse needs of patients with different diseases. However, achieving improved targeting and personalized medicine requires an in-depth study of cellular and molecular characteristics at both the overall and individual cell levels.

This necessitates the accurate sequencing of individual cell genomes and understanding their genomic sequences, post-translational modifications, RNA profiles, and chromatin structure. Studying gene expression is essential for comprehending cellular and metabolic responses within a cell [20]. Gene expression calls for the production of proteins

encoded by genes, which undergo post-translational modifications to perform specific functions. Each gene encodes a unique protein, and in a single cell, thousands of genes within DNA sequences are transcribed into mRNA, ultimately giving rise to a vast array of proteins known as the proteome. Researchers worldwide are actively investigating the proteome of single cells, investing in the latest advancements in single-cell RNA sequencing and proteomics/transcriptomics analysis. Single-cell RNA sequencing enables the analysis and study of the entire mRNA complement produced within a single cell's genome, leading to a comprehensive understanding of a cell's chemical and physical nature, along with its proteomic profile. The advent of its sequencing in 2009 sparked significant interest among researchers, recognizing its potential to unravel the heterogeneity among cells of the same type, tissue, or organ. Conducting high-resolution genomic studies on a single cell is significantly more challenging than bulk genomic analysis of an organism [21]. However, technological barriers have been overcome, paving the way for the implementation of single-cell analysis on a whole-genomic scale, albeit with ongoing challenges. The transcriptome encompasses the complete set of mRNAs generated from a genome.

Single-cell RNA sequencing (scRNA-seq) analysis provides a highly detailed examination of the transcriptome within individual cells, facilitating the characterization of cellular individuality and enabling comparisons between different cells. This approach proves particularly valuable in

distinguishing normal cells from cancerous cells originating from similar sources. Analyzing the disparities in transcriptomes and protein profiles between cells enhances the understanding of heterogeneity among normal and cancerous cells. Furthermore, these differences aid in identifying cell lines with similar functions and rare cell lines exhibiting enhanced cellular activities, such as heightened immune responses and receptor activity. Additionally, it enables the study of cellular heterogeneity based on their origin, transcriptional activities, splicing patterns, post-translational modifications, and gene expression and regulation.

Exploring cellular transcripts: steps in single cell RNA sequencing

Single-cell RNA sequencing analysis (scRNA-seq) represents a pioneering approach that has required significant technological advancements in bioinformatics and computational analysis tools. Although scRNA-seq begins as a wet-lab process, its initial step involves the isolation of viable target cells from a tissue or organ of interest. Various techniques, including micromanipulation, fluorescence-activated cell sorting, and droplet-based methods, are employed for the isolation of single cells intended for scRNA-seq analysis. Following the isolation of single cells, the subsequent step involves cell lysis to liberate their RNA complement for downstream analysis [23,24]. It is imperative to meticulously isolate RNA from lysed cell components to ensure precise analysis, as the utmost precision in these steps enhances the overall accuracy of RNA analysis manifold. The transcriptome of lysed cells can be isolated using oligo-dT beads or

random hexamers. Alternatively, isolation of mRNA alone can be achieved through poly-A enrichment. Since a cell encompasses chromosomal DNA and extra-chromosomal DNA, both of which generate RNA, it is essential to segregate mRNA molecules into distinct pools. To avert interference from extrachromosomal RNAs like ribosomal RNAs, the separation and analysis of polyadenylated mRNA molecules involve the use of poly[T]-primers. Conversely, non-polyadenylated RNAs and extrachromosomal RBNA follow distinct and intricate protocols compared to the analysis of polyadenylated RNA. For the isolated poly[T]-primed mRNA, complementary DNA (cDNA) is synthesized using reverse transcriptase, followed by amplification through polymerase chain reaction (PCR). The resulting amplified cDNA sequences are then fragmented and subjected to high-throughput sequencing technologies, employing various library preparation methods such as Smart-seq2, 10x Genomics Chromium, and Drop-seq [25]. This step encompasses multiple variations depending on the specific needs of the sequencing protocol. The resulting cDNA sequences are subsequently tagged for identification, enabling their segregation into pools for analysis and sequencing using dedicated sequencing tools. The sequenced data is then subjected to analysis to identify the expressed genes in the cells of interest, thereby determining the level of gene expression and discerning significant differences in gene expression levels between individual cells. This is accomplished by aligning the sequencing reads to a reference genome and quantifying the number of reads that map to each gene. To account for discrepancies in

sequencing depth among cells of the same or different types, data normalization techniques are applied to facilitate accurate analysis of gene expression.

Tools used in single cell RNA sequencing and analysis

Just like other sequencing methods, single cell RNA sequencing and analysis also makes use of two sets of tools that are sequencing tools and computational analysis tools. These tools help to sequence the RNA and analyze it further to elaborate the study of gene expression.

Sequencing tools

Drop seq and indrops

Drop seq technique of single cell RNA sequencing is a widely used strategy that involves profiling individual cells in large numbers using microfluidics. It works by encapsulating RNAs in tiny droplets for analysis. These droplets, on a nanoliter scale, serve as aqueous compartments to carry various substances like RNAs and nanoparticles. They can also act as reaction chambers for PCR and reverse transcriptase reactions. This method utilizes two types of capsules: simple microparticles and hydrogel microparticles. It offers several benefits, including barcoding the single cell transcriptome, reducing sample consumption, and providing high throughput results [26]. In the InDrops method, after isolating single cells from a tissue, mRNAs from the cells are captured using hydrogel probes. The hydrogel is then flowed into a microfluidic device, creating droplets, each containing one cell. The droplets are further

processed to isolate and prepare cDNAs from the mRNAs through reverse transcriptase and PCR amplification before sequencing. By using barcodes, the droplets can be identified, linking the transcriptome to its corresponding cell, facilitating the construction of gene expression profiles for each cell. These profiles are compared and analyzed to understand the similarities and heterogeneity among the cells. The InDrops method provides advantages such as high throughput, increased sensitivity towards small transcripts, and reduced risk of cross-contamination.

Microwell-seq and CEL-seq

Microwell sequencing, also known as microdroplet-based sequencing, utilizes microfluidics to divide multiple samples into individual droplets, each labelled with unique barcodes. This technique enables high-throughput analysis of gene expression, processing data from thousands to millions of cells in a single run. By using microwells as tiny compartments to isolate individual cells and reagents, cross-contamination is minimized. Despite its sensitivity, precautions are necessary to reduce errors. It finds applications in single-cell genomics, transcriptomics, and epigenomics. The 10x Genomics Chromium system is a prominent tool in this approach, using droplets with barcoded gel beads to capture and barcode individual cells [27]. CEL-seq is another powerful technology for single-cell RNA sequencing. Developed in 2013, it allows high-throughput gene expression profiling of single cells. It involves capturing individual cells using microfluidic devices, performing reverse transcription on their RNA molecules

to generate cDNA, and then linearly amplifying the cDNA with sequencing adapters for high-throughput sequencing. It offers advantages like high-quality data from a small number of cells, low technical errors, and accurate measurement of low-abundance transcripts. However, it may be susceptible to amplification bias, where certain transcripts are over-represented in the sequencing data due to linear amplification. It is widely used in various biological studies, including developmental biology, cancer research, and neuroscience.

Computational analysis tools

Seurat, Cell Ranger, and SCENIC are powerful tools used for analyzing single-cell RNA sequencing (scRNA-seq) data. They provide comprehensive frameworks for quality control, normalization, clustering, differential gene expression analysis, and visualization of scRNA-seq data.

Seurat is a widely used R package for scRNA-seq data analysis, offering comprehensive tools for quality control, normalization, clustering, differential gene expression analysis, and data visualization. Its graphical approach to clustering allows identification of rare cell types missed by other methods. The package also integrates data from multiple experiments, identifies marker genes for each cluster, performs dimensionality reduction using t-SNE or UMAP, and visualizes gene expression patterns with heatmaps and dot plots. Additionally, it supports spatial transcriptomics analysis, identifying spatially co-expressed genes. Despite its high sensitivity, Seurat remains a powerful and flexible tool due to regular updates and

continuous development. Cell Ranger, developed by 10x Genomics, is a widely used software for analyzing scRNA-seq data [28]. It processes raw sequencing data into a format ready for downstream analysis, including cell type identification, clustering, and differential gene expression analysis. Cell Ranger's command-line tool takes input as fastq files, producing outputs such as a gene-barcode matrix, cell barcodes, and gene annotations. Its parallelized algorithm efficiently handles large datasets with thousands to millions of cells, making processing time reasonable. It comes with built-in tools like Loupe Cell Browser for scRNA-seq data visualization and Cell Ranger ATAC for analyzing single-cell ATAC sequencing data. SCENIC is a computational method used to study and identify transcriptional regulatory networks in scRNA-seq data. It combines gene expression data with information about transcription factor binding motifs to infer transcription factor activity in each cell. This information is then used to cluster cells with similar regulatory networks. SCENIC is widely adopted for single-cell genomics studies due to its applications in identifying novel cell types, characterizing cell-state transitions, and understanding regulatory mechanisms underlying disease states. It is available as an open-source software package.

Samples that get analysed in scRNA-seq

scRNA-seq analysis has been successfully used to study mammalian cells, including human and mouse primary cells from embryos, tumors, nervous system, and hematopoietic cells, including stem cells and differentiated lymphocytes. The technique

has theoretically shown that any eukaryotic cell can be sequenced using scRNA-seq. The Human Cell Atlas, created with scRNA-seq, contains transcriptomic data from all human cell types and has the potential to revolutionize therapeutics for genetic disorders and advance translational research. However, there are challenges to overcome, such as isolating viable cells and improving analysis accuracy by avoiding extracellular DNA and other cellular organelles. Cells intertwined in collagen or connective tissue present isolation difficulties, even with cell isolation kits. Preserving the genome and transcriptome is crucial as scRNA-seq relies on the transcriptome alone. While isolating individual transcriptomes remains tedious, immune cells, blood cells, and excised tumor cells are more easily isolated due to their lack of neighboring connections [29]. ScRNA-seq is a powerful technique for analyzing gene expression profiles at the single-cell level. It enables the study of cell-to-cell variation, rare cell types, and novel cell states. In this method, individual cells are isolated, and their RNA is sequenced. The obtained data can identify expressed genes and compare their expression levels between cells, providing insights into functional properties and cellular interactions. Examples of cells analyzed using scRNA-seq include various blood cells (red blood cells, white blood cells, and platelets), individual tumor cells to identify cancer-driving mutations and potential therapeutics, different brain cell types (neurons, astrocytes, and oligodendrocytes), and immune cells (T cells, B cells, and natural killer cells). It is also used to analyze stem cells, identifying molecular mechanisms regulating differentiation and

self-renewal, and embryonic cells to understand gene expression during development and cell fate decisions.

scRNA-seq database

Single-cell RNA sequencing database is defined as the collection of gene expression data obtained from the analysis of individual cells. These databases contain information on thousands or millions of individual cells and can provide valuable insights into cellular heterogeneity, development, and cellular metabolism. There are many scRNA-seq databases available and every one of them has their own strengths and limitations. One of the most commonly used scRNA-seq databases include “Single Cell Portal” which is a web-based platform developed by the Broad Institute of MIT and Harvard, hosted on the Cancer Genomics Cloud infrastructure. It facilitates the analysis, visualization, and sharing of single-cell genomics data obtained from sequencing analysis. Researchers can easily access and analyze publicly available data sets or upload and analyze their own data with a user-friendly interface. The portal offers various analyses, including quality control, clustering, differential gene expression, and pathway analysis [30]. It also provides visualization options like heatmaps, t-SNE plots, and UMAP plots. Collaborative analysis is supported, allowing secure data and analysis sharing with others. The cell atlas project is an ongoing effort to create a comprehensive collection of molecular profiles for all cell types in an organism. Its primary goal is to map the genomic and transcriptomic data of all human cells, including gene expression patterns, epigenetic modifications, protein expression,

and cellular interactions. This project aims to provide insights into various diseases, such as cancer, neurological disorders, and autoimmune diseases. High-throughput technologies like single-cell RNA sequencing, mass cytometry, and spatial transcriptomics are used for analyzing cells from different tissues and organs, making the project costly and labor-intensive. Multiple organizations, including the Human Cell Atlas, are collaborating worldwide to create a complete cell atlas of the human body.

Applications of scRNA-seq

Single-cell RNA sequencing (scRNA-seq) is a powerful tool that enables researchers to study gene expression at the single-cell level, offering diverse applications in various research fields. It can identify different cell types within a tissue or organism, including rare cell populations. Moreover, scRNA-seq is used to study cellular and molecular mechanisms driving cell differentiation and development. It aids in understanding molecular changes during disease progression, facilitating the development of new therapies and precision treatments. Additionally, scRNA-seq helps screen for potential drug targets, study immune cell diversity and function, and enhances neurobiological research by studying neuronal development and function. In cancer research, it identifies distinct cell types within tumors and molecular targets for therapy, addressing tumor heterogeneity.

Limitations of scRNA-seq

Despite its benefits, scRNA-seq also has limitations. High cost is a major drawback,

particularly when analyzing large numbers of single cells. Technical variability arises due to differences in cell capture, library preparation, and sequencing. Limited sensitivity can affect poorly expressed or unexpressed genes, reducing accuracy and efficiency. Its throughput is limited, making analysis of large cell samples challenging. Analyzing rare cell populations in large samples is also difficult, as scRNA-seq may struggle to detect them. Handling large data volumes generated by scRNA-seq poses data analysis challenges. Furthermore, scRNA-seq lacks spatial information about cell-cell interactions and tissue architecture since cells are dissociated from their tissue context. It is essential to consider these limitations when designing experiments and interpreting results, despite the technology's powerful capabilities.

Exome sequencing: from bench to bedside, pioneering genetic discoveries

The protein coding segments known as exons are sequenced for identification of rare mutations and the detection of genetic variation prone to cause diseases. The exome capture technology operates exclusively on two categories, one being Solution-based wherein the targeted regions are hybridized selectively by fragmentation of DNA samples followed by biotinylating using streptavidin beads and the other being array-based, the mechanism being more or less similar to solution-based method with variation in the usage of high-density microarray to which the probes are attached [31]. The array-based method identifies common variants along with rare variants, this capture method is also used to determine diseases caused by

monogenic alterations. Array-based capture being the pioneer in exome capture technologies has several limitations such as their inability to scale up due restriction in the probe number, excessive time consumption and the requirement of supplementary equipment to process the microarray.

Illumina, NimbleGen and Agilent are the most widely used programmes for exome capture. NimbleGen, an eminently accurate, sensitive platform that provides cross-examination of expression of genomes from any annotated or sequenced fragments in the form of high-density arrays with a particular pattern that are unique as well as with the help of extended oligo-probes. Illumina brings about millions of sequences reads from a variety of genomes and performs comparisons among the databases that are eminently accurate, relatively cheap and are rapid, this helps in widening the knowledge and understanding to analyse several diseases [32,33].

Agilent is another platform that is known for its high sensitivity and a better performance in comparison to the previous two programmes in terms of variant detection. Exome sequencing has proved to be useful to humans in terms of diagnosis and to provide with prognoses in difficult to diagnose diseases even with minimal symptoms, in case of agriculture identification of gene responsible for inducing noddled dwarfism in barley is done via exome capture. Discovery of new genetic markers and comparison between human genome with other mammals can also be done with the assistance of exome capture.

Clinical exome sequencing

Clinical exome sequencing deals with exome sequences or the splice junctions for the identification and determination of rare variants that cause genetic disorders. Clinical exome sequence is known for its accuracy and precision in comparison with whole exome sequencing as it lays out reliable and extensive analytic outcomes on specific variants and disease related genes. Patients prone to mendelian disorders are advised for clinical exome sequencing as an initial step in the diagnosis as it proves to be accurate and cost effective. In cases of neurological disorder, it plays a very constricted role due to lack of recognition for coverage, to analyse and interpret the results [34]. Although there is very little venture in neurological disorder, it has proved to be highly useful and efficient in detecting myopathy, intellectual disability, neuropathy, epilepsy and ataxia.

Clinical exome sequencing is carried out through single-gene sequencing, multigene sequencing and chromosomal microarray analysis. In single gene sequencing a particular gene is analysed and sequenced for the diagnosis of a particular disease, the most commonly used technique is sanger sequencing. Multiple sequencing aims at detecting heterogeneous diseases like neurological disorders including polyneuropathy, epilepsy and patients suspected with type II diabetes mellitus which has the involvement of multiple genes or of polygene origin [35]. Multiple sequencing is done with sanger sequencing and panel testing. In the case of chromosomal microarray analysis detects the variants specifically based on structural changes like

duplications and deletions in array-oriented form for hybridization using comparative genomes. Congenital anomalies and autism spectrum disorders are analyzed via this technique. Clinical exome sequencing covers the disorders caused by polymorphisms, pathogenic mutations, alterations involving deletion, insertion, duplication, missense and nonsense mutations. In cases of diseases that require more than just the basic diagnostic procedure such as imaging, CSF analysis and a simple phenotypic analysis which would not be able to provide a clear picture, clinical exome sequencing offers an effective and accurate outcome.

Comprehending whole exome sequencing to look for hidden genetic clues

Whole exome sequencing (WES) is a modern sequencing method used to analyze only the protein-coding regions of the genome, known as the "exome." Focusing on these specific regions, which represent less than 2% of the genome but are responsible for approximately 85% of disease-related genetic changes, allows for an economical alternative to whole-genome sequencing [37]. It has a wide range of applications, including population genetics, genetic disease studies, and cancer research. It not only detects variants in coding exons but also includes untranslated regions (UTRs) and microRNA, enabling a more comprehensive view of gene regulation. With the ability to prepare libraries in a day and requiring only a small amount of sequencing per exome, it has been incorporated into pediatric and adult healthcare to detect suspected single-gene disorders. Furthermore, it is now utilized for prenatal

genetic diagnosis in cases where standard tests like chromosomal microarray analysis and karyotyping have not provided a clear diagnosis for pregnancies with fetal anomalies. While it has a higher molecular diagnostic rate compared to chromosomal microarray analysis, there is a possibility of identifying variants of uncertain significance and incidental findings unrelated to the fetal condition [38].

Whole exome sequencing (WES) involves several key steps. The process begins with DNA extraction, where high-quality genomic DNA (gDNA) is collected from biological samples, often obtained from leukocytes in peripheral blood, using traditional "salting out" or spin column-based techniques. Alternatively, formalin-fixed paraffin-embedded (FFPE) samples can be used, although they may yield lower quality DNA. After extraction, the DNA undergoes enzymatic or mechanical shearing to obtain small fragments [39].

Next comes exome library preparation, which includes DNA fragmentation, adaptor ligation, and target enrichment. Mechanically shearing the DNA creates fragments, and adaptors are ligated to the fragment ends. Various platforms employ distinct techniques, such as Sure Select QXT and Nextera, for shearing-free transposase-based library preparation. Specific target regions are then enhanced using platform-specific capture techniques. The process of DNA shearing is essential to produce high-quality sequencing libraries with unbiased shearing and minimal sample contamination [40]. Subsequently, the DNA fragments are converted into a DNA library, achieved

through PCR amplification or DNA fragmentation techniques, such as nebulization or sonication. The fragments are ligated to specific 5' and 3' adapters, enabling attachment to a solid surface for sequencing. The DNA library sequencing follows, where the genomic DNA is used to create a sequencing library by fragmenting and sizing the fragments to a specific length and attaching adaptors to the fragment ends. The quantity of the library is determined using PCR-based techniques like digital PCR or quantitative PCR.

Additionally, the fragmentation process can be performed enzymatically, chemically, or physically, with common techniques like sonication producing fragments with sizes between 100 and 5,000 bp. It is common to use a maximum insert size of 200–250 kb for exome sequencing, considering human exons are approximately 200 bp in length. Library quantification is essential to determine the quantity of nucleic acid molecules in the NGS library, often achieved through PCR-based methods like digital PCR or quantitative PCR. Finally, millions of DNA reads are generated by the sequencer, and specialized computer programs arrange them in the correct order, resulting in a finished genome sequence ready for further examination.

Empowering discoveries with computational tools in whole exome sequencing

The improvement and development of WES technology over the years brought in new bioinformatics methods and computational tools for efficient and detailed analysis as well as interpretation of the given sample [41]. The

majority of WES computational tools are focussed on creating variant calling format (VCF). VCF is further divided into the creation of pre-VCF tools and the post-VCF tools. With the help of pre-VCF tools the raw sequencing reads are aligned into the reference genome followed by variant detection and annotation. On the other hand, post-VCF tools are involved in somatic mutation detection, copy number alteration, driver prediction, pathway analysis and INDEL identification.

Computational tools involved in pre-VCF analysis

Alignment tools

The foremost step in any analysis of sequencing is the alignment of sequencing reads to reference genome, using two of the highly common referencing human genomes hg18 and hg19. There are several tools used for alignment purposes each having different methods and unique features, few among them are BWA, GEM, ELAND, MAQ, SOAP (1&2), mrFAST, Noalign, Bowtie (1&2), stampy, SSAHA and YOABS. There are three most commonly used algorithms one being BWA, which integrates all the given reference sequences into a long sequence using burrow wheeler aligner, the Second-Bowtie (1&2) wherein, the short DNA Sequences are aligned into large genomes, it is an extended version of burrow wheeler aligner which integrates novel quality aware backtracking algorithm that allows for mismatches, lastly SOAP (1&2) which aligns gapped and ungapped short oligonucleotide sequences onto a reference sequence using the same Burrow wheeler aligner.

Auxiliary tools

After the alignment of sequencing reads to reference genome, these auxiliary tools help in filtering the aligned reads to make sure there is high quality data for further downstream analyses. PCR amplification brings out the duplicate reads that have a great influence and affects mapped reads as well as has an impact in downstream analyses [42].

To detect PCR originated duplicate reads several tools are developed among which Picard, FastUniq and SAMtools are widely used. SAMtools finds sequences that start and end at the same position as well as helps in identifying reads with highest quality scores and marks the rest as duplicates for elimination. Picard works by identifying the similar 5' positions and marks them as duplicates for elimination. FastUniq uses a different method that uses a de novo approach to identify the PCR duplicates in a short period of time.

Single nucleotide variant (SNV) calling tools

The detection of variants in WES data is the next step and this is through the calling for single nucleotide variants (SNV). The variant calling is divided into four categories: somatic variants, germline variants, copy number variants and finally structural variants [43]. There are multiple tools that perform one or more SNV variant calling techniques, the commonly used tools among them are GATK, SAMtools and VCMM. GATK and SAMtools have the same operating mechanism but differentiate in terms of errors, GATK works

on the assumption that every error is independent of each other while SAMtools considers that the secondary error carries more weightage. VCMM suppresses the false positive and false negative variant calls in comparison with the other two tools.

Contemporary methods for identifying structural variants (SV)

Structural variants (SV) include insertions and deletions (INDELs) which are more challenging to isolate than single nucleotide variants as they include undefined number of nucleotides. SVs can be identified up to a certain level using SAMtools and GATK tools. Unique and accurate tools for INDELs identification will increase the sensitivity as well as decrease the false discovery rate. Platypus, FreeBayes, Pindel and Splitread are few among the unique and efficient INDEL identification tools. Platypus uses denovo assembly to identify both SNVs and INDELs but is more efficient and sensitive to INDEL identification as well as has lower fosmid false discovery rate [44,45].

FreeBayes uses haplotype-based variant detection to detect INDELs. Pindel was one of the first programmes to be developed for the identification of unidentified INDELs. This tool works by simply identifying the reads where only one end of the sequence is mapped and the other is not.

Splitread was specifically designed to identify INDELs and SVs in WES data, by anchoring one end of the sequence while clustering the other ends to find content, location size of SVs. A compilation of tools like indelMINER, sprites are advised to identify SVs in a read.

Methods used in variant calling format (VCF) annotation

The next step involves the annotation of the variants that were aligned, detected, and called. The annotation tools include ANNOVAR, MuTect, SnpEff, SnpSift and VAT among which ANNOVAR and MuTect being the most popular ones [46]. ANNOVAR accesses over 20+ databases by using gene, region and filter-based annotations for variants annotation. MuTect is widely used in cancer genomics research which uses Bayesian classifiers for detection as well as annotation of variants.

Functional prediction tools

For determining the effect of the variants, several functional prediction tools were designed with slight variation in their algorithms. SIFT, LRT, VEST, CADD, MetaLR, FATHMM, MetaSVM and PolyPhen-2 are some of the function prediction tools that gives the users scores through ANNOVAR. SIFT uses PSI-BLAST algorithm to if a variant is deleterious as well as helps to determine the conservation of amino acids using closely linked sequence alignments. PolyPhen-2 is used in determining if the mutation is benign or deleterious using a pipeline of eight sequence-based methods and three structural based methods.

LRT (likelihood ratio test) is used to determine a particular mutation's functional properties and its impact by exploiting the knowledge of conservation between two closely related species. Multiple mutation predictors are used to detect a wide range of deleterious SNVs [47,48]. FATHMM is used to

predict functional effects of protein mutation with the use of sequence conservation from the hidden markov models. This tool considers mutation on the basis of pathogenicity. MetaSVM and MetaLR are two ensemble mechanisms that integrate ten different predictor scores to predict the deleterious variants. VEST (variant effect scoring tool) works with the help of training set and machine learning for the prediction of a mutation's function, it is specifically designed for mendelian studies [49]. CADD (combined annotation dependent depletion) integrates multiple variants with mutation using an entirely different approach that involves both stimulated mutations as well as naturally survived mutations.

Computational methods involved in post-VCF analysis

Tools involved in determining significant somatic variations

VarSim, Somaticsniper, MuTect, and SomVarIUS are tools that are used to identify the somatic variants among thousands of identified SNVs. Somaticsniper works by comparing the normal and tumor samples to find out the uniqueness in a particular mutation, using genotype likelihood model of MAQ.

MuTect similar to somaticsniper detects somatic mutation with the help of normal and cancer sample inputs using variant detection statistics, MuTect removes common polymorphisms with the help of dbSNP database as well as known sequencing artifacts to identify somatic mutations as well as to minimise the set of candidate genes [50].

VarSim uses simulation as well as experimental data for variant calling and assessing alignment.

SomVarIUS is widely used in tumor samples for the identification of somatic variants, by first identifying the possible variant sites then estimating the occurrence of sequence error and finally by identifying the origin of the variant either germline or somatic.

Tools that estimate CNVs (Copy Number Variations)

Copy number alterations can be estimated using various tools like CONTRA, EXCAVATOR, SegSeq, ADTEX, ExomeCNV, CNV-seq, control-FREEC and VarScan2. The most commonly used VarScan2 estimates the somatic mutations as well as CNAs with the help of normal sample to find somatic CNAs.

Tools that predict drivers in cancer exomes

Driver mutation in cancer is involved in the progression of the cancer, there are several tools to identify these driver mutations which includes Dendrix, CHASM and MutSigCV. Cancer specific high throughput annotation of somatic mutation CHASM is a highly sensitive tool that was developed to differentiate the known driver missense and missense mutation that was randomly generated with the help of COSMIC database and substitution frequencies. MutSigCV is another common tool used to identify drivers in cancer exomes, which reduces the problem of false positive findings and identify the true drivers [51]. Dendrix identifies the de novo driver pathways with the help of two algorithms, one being greedy algorithm and

the other being markov chain monte carlo algorithm.

After the pre-VCF and post-VCF alterations the samples are analysed with databases to identify the pathways, interactions and ultimately to link these variants to the drug targets by providing a bridge that connects genomic data and clinical therapeutic treatments.

Applications and ethical considerations of WES

Applications of Whole Exome Sequencing are seen in diverse fields, including agriculture, medicine, and cancer research. In agriculture, exome sequencing plays a vital role in analyzing natural evolution in plants, studying host-pathogen interactions, and improving crop production by providing insights into the genetic composition of crop varieties, enhancing disease resistance, and optimizing nutrient utilization [52].

In medicine, Whole Exome Sequencing (WES) is extensively utilized for diagnosing rare genetic disorders, identifying pathogenic variants responsible for patient symptoms, and predicting drug metabolism and response, enabling personalized pharmacological therapy. In cancer research, it is pivotal for detecting somatic mutations unique to tumor cells, shedding light on cancer onset and spread, understanding tumor heterogeneity, and identifying genetic biomarkers for diagnosis, prognosis, and therapeutic response.

Despite its potential, WES raises ethical concerns regarding informed consent, data sharing, and the return of incidental findings,

and various professional associations have explored these issues [53]. As WES becomes more prevalent in medical diagnostics, addressing these ethical challenges will be essential to harness its full potential for improving healthcare.

A closer look at targeted sequencing and a glimpse of amplicon sequencing

Targeted sequencing focuses on specific genomic regions of interest, allowing for increased depth of coverage by allocating more sequencing time. This approach reduces the number of sequencing runs and libraries, resulting in a rapid and cost-effective workflow. It is instrumental in the development of targeted therapy applications and personalized medicine. Its greater depth of coverage makes it suitable for various applications, including cancer research and identification of rare diseases [54].

Compared to whole-genome sequencing (WGS), it has advantages such as reduced computational resource utilization due to smaller datasets, increased scalability, and the ability to handle more samples and sequencing runs. The upfront selection and isolation of genes or regions of interest are achieved through PCR amplification or hybridization-based capture methods. For sequencing a small number of targeted regions, PCR amplification in conjunction with Sanger sequencing is employed, while Ion AmpliSeq™ gene panels are utilized for larger numbers of genes or regions, providing the ability to conveniently target and sequence hundreds of genes on the Ion PGM™ System. Importantly, the Ion TargetSeq™ Enrichment System is used to

sequence higher-density target regions of up to ~60 MB due to its cost-saving and customizable solution capture method for the Ion PGM™ System [55].

The working of targeted sequencing involves a step called "target enrichment," which utilizes two methods: hybridization capture and amplicon sequencing. Hybridization capture employs biotinylated oligonucleotide probes to capture regions of interest, followed by separation through magnetic streptavidin-biotin binding complexes. Tiling probes enable the coverage of large contiguous regions from terminal to terminal, overcoming obstacles like repetitive sequences [56]. Hybridization capture is a cost-efficient and effort-saving approach with multiplexing capacity to increase efficiency. The usage of unique molecular identifiers (UMIs) enhances sequencing precision, enabling the identification of specific molecules and rectification of PCR and sequencing errors, especially for low-frequency variants.

Amplicon sequencing, on the other hand, uses PCR to construct DNA sequences called amplicons, allowing analysis of genetic variation in specific genomic regions [57]. Samples are multiplexed and barcoded for sequencing in pools, offering a rapid and straightforward workflow. The adapters facilitate indexed amplicons to adhere to the sequence flow cell. The choice between hybridization capture and amplicon sequencing depends on factors like desired accuracy, budgetary constraints, and downstream sequencing application. Tools like CleanPlex and xGen Pre-Designed Hyp Panels offer highly scalable, sensitive

amplicon-based solutions with uniform multiplex PCR amplification chemistry. CleanPlex's workflow involves amplification of targets, removal of unwanted products, and PCR indexing, offering exceptional sensitivity and the detection of various variants. xGen Pre-Designed Hyp Panels and xGen Custom Hyb Panels provide reliable capture results with uniform coverage and automation-friendly protocols.

Targeted sequencing confers several advantages, including identifying rare variants and causative mutations in a single assay, providing accurate and easily interpretable results, and cost-efficient disease-related gene findings. However, it can have limitations in detecting circulating tumor DNA (ctDNA) [58,59]. Its application spans multiple fields, including cancer research and diagnostics, reproductive health (carrier screening, prenatal testing, preimplantation genetic diagnosis, etc.), industry (agri genomics, food safety, forensics, etc.), and environmental research.

The future scope of targeted sequencing is promising, especially in personalized medicine and gene editing, due to its high efficiency, cost-saving, and accuracy [60]. Ethical concerns arise with targeted sequencing, concerning data release and identifiability, adequacy of consent, and reporting research results, affecting both individuals and populations.

Overall, targeted sequencing offers immense potential for transformative advancements in medical research and applications, underlining the importance of addressing ethical considerations.

Traversing the epigenetic landscape for discoveries through ATAC-seq

ATAC-Seq, a powerful technique, harnesses the hyperactive transposase Tn5 to simultaneously fragment chromatin and incorporate NGS adapters into the fragmented regions. This process is crucial for constructing a next-generation sequencing library, which is subsequently subjected to sequencing to analyze open chromatin regions in the sample genome [61].

This method stands out for its efficient and straightforward two-step procedure, making it easy to handle and requiring only a small sample volume. Recent advancements in this approach have opened new avenues for investigating various areas of interest. For instance, it has proven valuable in studying cancer epigenetics, such as breast cancer, as well as exploring anti-aging phenomena and diseases like age-related macular degeneration.

Furthermore, ATAC-Seq is instrumental in examining immunological interactions, including B-cell maturation, response, and progenitors, thus enhancing the understanding of cellular development and differentiation [62].

Recent versions of this technique have introduced several variations to suit diverse research needs. Examples include Omni ATAC-Seq, single-cell ATAC-Seq, and ENCODE ATAC-Seq. These adaptations broaden the applicability of this approach and enable researchers to delve deeper into the complexities of chromatin accessibility and regulatory mechanisms [63].

Probing the DNA-protein interface with the help of ChIP-seq

Chromatin immunoprecipitation sequencing (ChIP-Seq) is an antibody-based technique that enables the precise determination and localization of protein binding sites within the genome. It sheds light on interactions between DNA and specific proteins by selectively enriching DNA-binding proteins at their respective targets. This method detects the presence of certain proteins or mutations caused by them, including histone modifications, and analyzes interactions between proteins and specific genomic loci [64,65].

ChIP-Seq involves five essential steps: first, forming crosslinks to bind proteins to DNA; second, fragmenting chromatin; third, using antibodies for protein precipitation, resulting in aggregates; fourth, downstream analysis to extract DNA from the precipitated mixture; and finally, performing PCR, microarray, or next-generation sequencing to analyze DNA and its interactions with proteins. By following this systematic approach, ChIP-Seq provides critical insights into the intricate regulatory mechanisms governing gene expression and chromatin structure, thereby advancing the understanding of cellular processes and disease mechanisms.

A window into DNA methylation dynamics by exploring MeDIP-sequencing

The acronym MeDIP -Seq stands for methylated DNA immunoprecipitation sequencing/DNA immunoprecipitation sequencing. It is a large-scale purification

method that isolates methylated DNA fragments using an antibody produced against 5-methylcytosine in order to enrich methylation DNA sequence. It is therefore frequently used to examine 5MC or 5HMC alteration. Methylated DNA is separated from genomic DNA via immunoprecipitation if a 5mc-specific antibody is used [66].

Genomic DNA fragments and anti-5mc antibodies are incubated together, precipitated, and then the DNA is purified and sequenced. While locus-specific techniques like methylation-specific PCR (MS-PCR) rely on prior knowledge of potentially methylated regions for the design of primers and subsequent detection, the anti-5mC antibody's indiscriminate activity for methylated cytosines also enables an informative, unbiased analysis of the genome at specific loci. Combining various pretreatment methods with various later molecular biology techniques, such DNA microarrays and next-generation sequencing (NGS), allows for the mapping of DNA methylation over the entire genome. It is the most effective method for methylation cytosines coverage over the entire genome.

Unlike other procedures, MeDIP does not require enzymatic digestion or bisulfite conversion in order to obtain methylated fragments, making it a quick and generally inexpensive procedure [67]. Combining it with techniques that are better at detecting unmethylated regions is a strategy that is increasingly popular for high-resolution, low-cost methylome research. The 5-hydroxymethylcytosine (hmC) can also be identified throughout the complete genome using the hydroxy methylated DNA

immunoprecipitation (hMeDIP) technique using a 5hmC-specific antibody [68].

MeDIP-seq does not result in the introduction of mutations or uracil-tolerant DNA polymerase. It has a substantial advantage over methods based on enzymatic digestion in that it does not favor nucleotide sequences other than CpGs.

The relationship between enrichment and absolute methylation levels is complicated by elements like CpG density, though [69,70].

MeDIP-Seq can collect roughly the same amount of the methylome and can also find differentially methylated regions (DMRs). Although antibodies in this sequencing capture DNA fragments including any methylated cytosines, non-CpG methylation may be significant for some disorders.

Conclusion

The paper on Next-Generation Sequencing (NGS) has shed light on the remarkable advancements and transformative impact this technology has had on genomics and various fields of research. From its inception to the present, it has revolutionized the understanding of genetics, disease mechanisms, and biodiversity, paving the way for groundbreaking discoveries. With each passing day, it becomes an increasingly indispensable tool in the pursuit of knowledge and innovation, fueling curiosity and empowering humanity to unravel the mysteries of life at the molecular level. As the authors look to the future, it is poised to continue its momentous trajectory, driving advancements in personalized medicine, agriculture, and environmental conservation.

The amalgamation of it with artificial intelligence and machine learning promises to amplify its impact even further, unlocking novel insights from the vast sea of genomic data. However, amidst its triumphs, NGS also faces certain limitations, such as data storage and management challenges, the potential for errors, and ethical considerations surrounding the use of genomic information. Addressing these hurdles will be crucial to maximizing the potential benefits of NGS while ensuring responsible and ethical practices in its application.

Statements and declarations

Funding

The authors affirm that this manuscript was created without receiving any funding, grants, or additional support.

Competing interests

It is hereby stated that the authors have no conflicting relationships that could be perceived as having influenced the content or conclusions of this manuscript.

Consent statement

The authors affirm that all data, tables, and information presented in this review paper have been sourced from publicly available literature, research articles, and reputable sources. No personal or confidential information has been used without proper authorization and consent.

Author contributions

Muhammed Ali Siham H R led the team as the first and corresponding author, overseeing

the paper's conceptualization, execution, and writing. Ashwin Prabahar A conducted thorough literature review, data curation, and contributed to the paper's structure. Sriprata R and Sandra Nixon provided essential insights into data interpretation and ensured methodological rigor and contributed to

scientific review. Sandhiya P enhanced conceptual clarity and overall coherence. Gnana Sowndariyan G and Palak Bhataria managed references and formatting. Together, the collaborative efforts of all authors enriched the paper's quality and depth.

References

1. Cheng J, Smyth GK, Chen Y. Unraveling the Timeline of Gene Expression: A Pseudo-Temporal Trajectory Analysis of Single-Cell RNA Sequencing Data. *bioRxiv*. 2023;2023-05. [CrossRef](#)
2. Roberts L. Timeline: A History of The Human Genome Project. *Science*. 2001;291(5507):1195-200. [PubMed](#) | [CrossRef](#)
3. Harrison A, Parle-McDermott A. DNA Methylation: A Timeline of Methods and Applications. *Front Genet*. 2011;2:74. [PubMed](#) | [CrossRef](#)
4. Ari Ş, Arikan M. Next-Generation Sequencing: Advantages, Disadvantages, and Future. *Plant Omics: Trends Applications*. 2016;109-35. [CrossRef](#)
5. Kamalakaran S, Varadan V, Janevski A, Banerjee N, Tuck D, McCombie WR, et al. Translating Next Generation Sequencing to Practice: Opportunities and Necessary Steps. *Mol Oncol*. 2013;7(4):743-55. [PubMed](#) | [CrossRef](#)
6. Ozsolak F, Platt AR, Jones DR, Reifengerger JG, Sass LE, McInerney P, et al. Direct RNA Sequencing. *Nature*. 2009;461(7265):814-8. [PubMed](#) | [CrossRef](#)
7. Vincent AT, Derome N, Boyle B, Culley AI, Charette SJ. Next-Generation Sequencing (NGS) in The Microbiological World: How to Make the Most of Your Money. *J Microbiol Methods*. 2017;138:60-71. [PubMed](#) | [CrossRef](#)
8. Mutz KO, Heilkenbrinker A, Lönne M, Walter JG, Stahl F. Transcriptome Analysis Using Next-Generation Sequencing. *Curr Opin Biotechnol*. 2013;24(1):22-30. [PubMed](#) | [CrossRef](#)
9. Du H, Bao Z, Hou R, Wang S, Su H, Yan J, et al. Transcriptome Sequencing and Characterization for The Sea Cucumber *Apostichopus Japonicus* (Selenka, 1867). *PloS One*. 2012;7(3):e33311. [PubMed](#) | [CrossRef](#)
10. Lappalainen T, Sammeth M, Friedländer MR, 't Hoen PA, Monlong J, Rivas MA, et al. Transcriptome and Genome Sequencing Uncovers Functional Variation in Humans. *Nature*. 2013;501(7468):506-11. [PubMed](#) | [CrossRef](#)
11. Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T. Transcriptomics Technologies. *Plos Computational Biol*. 2017;13(5):e1005457. [PubMed](#) | [CrossRef](#)
12. Zhu L, Lei J, Devlin B, Roeder K. A Unified Statistical Framework for Single Cell and Bulk RNA Sequencing Data. *Ann Appl Stat*. 2018;12(1):609. [PubMed](#) | [CrossRef](#)
13. Li X, Wang CY. From Bulk, Single-Cell to Spatial RNA Sequencing. *Int J Oral Sci*. 2021;13(1):36. [PubMed](#) | [CrossRef](#)
14. Thind AS, Monga I, Thakur PK, Kumari P, Dindhoria K, Krzak M, et al. Demystifying Emerging Bulk RNA-Seq Applications: The Application and Utility of Bioinformatic Methodology. *Brief Bioinform*. 2021;22(6):bbab259. [PubMed](#) | [CrossRef](#)
15. Ng PC, Kirkness EF. Whole Genome Sequencing. *Methods Mol Biol*. 2010:215-26. [PubMed](#) | [CrossRef](#)
16. Park ST, Kim J. Trends in Next-Generation Sequencing and A New Era for Whole Genome Sequencing. *Int Neurol J*. 2016(Suppl 2):S76. [PubMed](#) | [CrossRef](#)
17. Kwong JC, McCallum N, Sintchenko V, Howden BP. Whole Genome Sequencing in Clinical and Public Health Microbiology. *Pathology*. 2015;47(3):199-210. [PubMed](#) | [CrossRef](#)

18. Van El CG, Cornel MC, Borry P, Hastings RJ, Fellmann F, Hodgson SV, et al. Whole-Genome Sequencing in Health Care. *Eur J Hum Genet.* 2013;21(6):580-4. [PubMed](#) | [CrossRef](#)
19. Dewey FE, Grove ME, Pan C, Goldstein BA, Bernstein JA, Chaib H, et al. Clinical Interpretation and Implications of Whole-Genome Sequencing. *JAMA.* 2014;311(10):1035-45. [PubMed](#) | [CrossRef](#)
20. Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA. The Technology and Biology of Single-Cell RNA Sequencing. *Mol Cell.* 2015;58(4):610-20. [PubMed](#) | [CrossRef](#)
21. Luecken MD, Theis FJ. Current Best Practices in Single-Cell RNA-Seq Analysis: A Tutorial. *Mol Syst Biol.* 2019;15(6):e8746. [PubMed](#) | [CrossRef](#)
22. Saliba AE, Westermann AJ, Gorski SA, Vogel J. Single-Cell RNA-Seq: Advances and Future Challenges. *Nucleic Acids Res.* 2014;42(14):8845-60. [PubMed](#) | [CrossRef](#)
23. Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, et al. Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol cell.* 2017;65(4):631-43. [PubMed](#) | [CrossRef](#)
24. Wu AR, Neff NF, Kalisky T, Dalerba P, Treutlein B, Rothenberg ME, et al. Quantitative Assessment of Single-Cell RNA-Sequencing Methods. *Nat Methods.* 2014;11(1):41-6. [PubMed](#) | [CrossRef](#)
25. Papalexi E, Satija R. Single-Cell RNA Sequencing to Explore Immune Cell Heterogeneity. *Nat Rev Immunol.* 2018;18(1):35-45. [PubMed](#) | [CrossRef](#)
26. Hwang B, Lee JH, Bang D. Single-Cell RNA Sequencing Technologies and Bioinformatics Pipelines. *Exp Mol Med.* 2018;50(8):1-4. [PubMed](#) | [CrossRef](#)
27. Suvà ML, Tirosch I. Single-Cell RNA Sequencing in Cancer: Lessons Learned and Emerging Challenges. *Mol Cell.* 2019;75(1):7-12. [PubMed](#) | [CrossRef](#)
28. Chen G, Ning B, Shi T. Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. *Front Genet.* 2019;10:317. [PubMed](#) | [CrossRef](#)
29. Bacher R, Kendziorski C. Design and Computational Analysis of Single-Cell RNA-Sequencing Experiments. *Genome Biol.* 2016;17(1):1-4. [PubMed](#) | [CrossRef](#)
30. Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, et al. Quantitative Single-Cell RNA-Seq with Unique Molecular Identifiers. *Nat Methods.* 2014;11(2):163-6. [PubMed](#) | [CrossRef](#)
31. Majewski J, Schwartzentruber J, Lalonde E, Montpetit A, Jabado N. What Can Exome Sequencing Do for You? *J Med Genet.* 2011;48(9):580-9. [PubMed](#) | [CrossRef](#)
32. Warr A, Robert C, Hume D, Archibald A, Deeb N, Watson M. Exome Sequencing: Current and Future Perspectives. *G3 (Bethesda).* 2015;5(8):1543-50. [PubMed](#) | [CrossRef](#)
33. Biesecker LG, Green RC. Diagnostic Clinical Genome and Exome Sequencing. *N Engl J Med.* 2014;370(25):2418-25. [PubMed](#) | [CrossRef](#)
34. Tarailo-Graovac M, Shyr C, Ross CJ, Horvath GA, Salvarinova R, Ye XC, et al. Exome Sequencing and The Management of Neurometabolic Disorders. *New N Engl J Med.* 2016;374(23):2246-55. [PubMed](#) | [CrossRef](#)
35. Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A, et al. Whole-Genome Sequencing is More Powerful Than Whole-Exome Sequencing for Detecting Exome Variants. *Proc Natl Acad Sci USA.* 2015;112(17):5473-8. [PubMed](#) | [CrossRef](#)
36. Samuels DC, Han L, Li J, Quanguh S, Clark TA, Shyr Y, et al. Finding the Lost Treasures in Exome Sequencing Data. *Trends Genet.* 2013;29(10):593-9. [PubMed](#) | [CrossRef](#)
37. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, et al. Exome Sequencing Identifies the Cause of a Mendelian Disorder. *Nat Genet.* 2010;42(1):30-5. [PubMed](#) | [CrossRef](#)
38. Biesecker LG, Shianna KV, Mullikin JC. Exome Sequencing: The Expert View. *Genome Biol.* 2011;12(9):1-3. [PubMed](#) | [CrossRef](#)
39. Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, et al. Molecular Findings Among Patients Referred for Clinical Whole-Exome Sequencing. *JAMA.* 2014;312(18):1870-9. [PubMed](#) | [CrossRef](#)
40. Singleton AB. Exome Sequencing: A Transformative Technology. *Lancet Neuro.* 2011;10(10):942-6. [PubMed](#) | [CrossRef](#)
41. Rabbani B, Tekin M, Mahdieh N. The Promise of Whole-Exome Sequencing in Medical Genetics. *J Hum Genet.* 2014;59(1):5-15. [PubMed](#) | [CrossRef](#)

42. Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A, et al. Whole-Genome Sequencing is More Powerful than Whole-Exome Sequencing for Detecting Exome Variants. *Proc Nat Acad Sci*. 2015;112(17):5473-8. [PubMed](#) | [CrossRef](#)
43. Retterer K, Juusola J, Cho MT, Vitazka P, Millan F, Gibellini F et al. Clinical Application of Whole-Exome Sequencing Across Clinical Indications. *Genet Med*. 2016;18(7):696-704. [PubMed](#) | [CrossRef](#)
44. Iglesias A, Anyane-Yeboah K, Wynn J, Wilson A, Truitt Cho M, Guzman E, et al. The Usefulness of Whole-Exome Sequencing in Routine Clinical Practice. *Genet Med*. 2014;16(12):922-31. [PubMed](#) | [CrossRef](#)
45. Best S, Wou K, Vora N, Van der Veyver IB, Wapner R, Chitty LS. Promises, Pitfalls and Practicalities of Prenatal Whole Exome Sequencing. *Prenatal Diag*. 2018;38(1):10-9. [PubMed](#) | [CrossRef](#)
46. Atwal PS, Brennan ML, Cox R, Niaki M, Platt J, Homeyer M, et al. Clinical Whole-Exome Sequencing: Are We there Yet? *Genet Med*. 2014;16(9):717-9. [PubMed](#) | [CrossRef](#)
47. Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, et al. Clinical Whole-Exome Sequencing for the Diagnosis of Mendelian Disorders. *N Engl J Med*. 2013;369(16):1502-11. [PubMed](#) | [CrossRef](#)
48. Srivastava S, Cohen JS, Vernon H, Barañano K, McClellan R, Jamal L, et al. Clinical Whole Exome Sequencing in Child Neurology Practice. *Ann Neurol*. 2014;76(4):473-83. [PubMed](#) | [CrossRef](#)
49. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ et al. De Novo Mutations Revealed by Whole-Exome Sequencing are Strongly Associated with Autism. *Nature*. 2012;485(7397):237-41. [PubMed](#) | [CrossRef](#)
50. Foo JN, Liu JJ, Tan EK. Whole-Genome and Whole-Exome Sequencing in Neurological Diseases. *Nat Rev Neurol*. 2012;8(9):508-17. [PubMed](#) | [CrossRef](#)
51. Tetreault M, Bareke E, Nadaf J, Alirezaie N, Majewski J. Whole-Exome Sequencing as A Diagnostic Tool: Current Challenges and Future Opportunities. *Expert Rev Mol Diagn*. 2015;15(6):749-60. [PubMed](#) | [CrossRef](#)
52. Posey JE, Rosenfeld JA, James RA, Bainbridge M, Niu Z, et al. Molecular Diagnostic Experience of Whole-Exome Sequencing in Adult Patients. *Genet Med*. 2016;18(7):678-85. [PubMed](#) | [CrossRef](#)
53. Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, et al. Comparison and Integration of Deleteriousness Prediction Methods for Nonsynonymous SNVs in Whole Exome Sequencing Studies. *Hum Mol Genet*. 2015;24(8):2125-37. [PubMed](#) | [CrossRef](#)
54. Rehm HL. Disease-Targeted Sequencing: A Cornerstone in the Clinic. *Nat Rev Genet*. 2013;14(4):295-300. [PubMed](#) | [CrossRef](#)
55. Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, et al. Evaluation of Next Generation Sequencing Platforms for Population Targeted Sequencing Studies. *Genome Biol*. 2009;10(3):1-3. [PubMed](#) | [CrossRef](#)
56. Han SW, Kim HP, Shin JY, Jeong EG, Lee WC, Lee KH, et al. Targeted Sequencing of Cancer-Related Genes in Colorectal Cancer Using Next-Generation Sequencing. *PloS One*. 2013;8(5):e64271. [PubMed](#) | [CrossRef](#)
57. Tewhey R, Warner JB, Nakano M, Libby B, Medkova M, et al. Microdroplet-Based PCR Enrichment for Large-Scale Targeted Sequencing. *Nat Biotechnol*. 2009;27(11):1025-31. [PubMed](#) | [CrossRef](#)
58. Schultzhau Z, Wang Z, Stenger D. CRISPR-Based Enrichment Strategies for Targeted Sequencing. *Biotechnol Adv*. 2021; 46:107672. [PubMed](#) | [CrossRef](#)
59. Larridon I, Villaverde T, Zuntini AR, Pokorny L, Brewer GE, Epitawalage N, et al. Tackling Rapid Radiations with Targeted Sequencing. *Front Plant Sci*. 2020;10:1655. [PubMed](#) | [CrossRef](#)
60. Mercer TR, Clark MB, Crawford J, Brunck ME, Gerhardt DJ, Taft RJ, et al. Targeted Sequencing for Gene Discovery and Quantification Using RNA CaptureSeq. *Nat Protoc*. 2014;9(5):989-1009. [PubMed](#) | [CrossRef](#)
61. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Current protocols in molecular biology*. 2015 Jan;109(1):21-9. [PubMed](#) | [CrossRef](#)
62. Sun Y, Miao N, Sun T. Detect Accessible Chromatin Using ATAC-Sequencing, from Principle to Applications. *Hereditas*. 2019;156(1):1-9. [PubMed](#) | [CrossRef](#)
63. Park PJ. ChIP-Seq: Advantages and Challenges of A Maturing Technology. *Nat Rev Genet*. 2009;10(10):669-80. [PubMed](#) | [CrossRef](#)

64. Zhang ZD, Rozowsky J, Snyder M, Chang J, Gerstein M. Modeling ChIP Sequencing in Silico with Applications. PLoS comp bio. 2008;4(8):e1000158. [PubMed](#) | [CrossRef](#)
65. Pepke S, Wold B, Mortazavi A. Computation for ChIP-Seq and RNA-Seq Studies. Nat Methods. 2009;6(Suppl 1):S22-32. [PubMed](#) | [CrossRef](#)
66. Taiwo O, Wilson GA, Morris T, Seisenberger S, Reik W, Pearce D, et al. Methylome Analysis Using MeDIP-seq with Low DNA Concentrations. Nat Protoc. 2012;7(4):617-36. [PubMed](#) | [CrossRef](#)
67. Li N, Ye M, Li Y, Yan Z, Butcher LM, Sun J, et al. Whole Genome DNA Methylation Analysis Based on High Throughput Sequencing Technology. Met. 2010;52(3):203-12. [PubMed](#) | [CrossRef](#)
68. Li D, Zhang B, Xing X, Wang T. Combining MeDIP-Seq and MRE-Seq to Investigate Genome-wide CpG Methylation. Met. 2015;72:29-40. [PubMed](#) | [CrossRef](#)
69. Zhao MT, Whyte JJ, Hopkins GM, Kirk MD, Prather RS. Methylated DNA Immunoprecipitation and High-Throughput Sequencing (MeDIP-seq) Using Low Amounts of Genomic DNA. Cell Reprog. 2014;16(3):175-84. [PubMed](#) | [CrossRef](#)
70. Clark C, Palta P, Joyce CJ, Scott C, Grundberg E, Deloukas P, et al. A Comparison of the Whole Genome Approach of MeDIP-Seq to the Targeted Approach of the Infinium HumanMethylation450 BeadChip® for Methylome Profiling. PloS One. 2012;7(11):e50233. [PubMed](#) | [CrossRef](#)