biology

MDPI

Review

Next-Generation Sequencing Technology: Current Trends and Advancements

Heena Satam [1], Kandarp Joshi [1], Upasana Mangrolia [1], Sanober Waghoo [1], Gulnaz Zaidi [1], Shravani Rawool [1], Ritesh P. Thakare [2], Shahid Banday [2], Alok K. Mishra [2], Gautam Das [1,*] and Sunil K. Malonia [2,*]

[1] miBiome Therapeutics, Mumbai 400102, India
[2] Department of Molecular Cell and Cancer Biology, UMass Chan Medical School, Worcester, MA 01605, USA; alok.mishra@umassmed.edu (A.K.M.)
* Correspondence: gautam@mibiome.com (G.D.); sunil.malonia@umassmed.edu (S.K.M.)

Simple Summary: Next-generation sequencing (NGS) is a powerful tool used in genomics research. NGS can sequence millions of DNA fragments at once, providing detailed information about the structure of genomes, genetic variations, gene activity, and changes in gene behavior. Recent advancements have focused on faster and more accurate sequencing, reduced costs, and improved data analysis. These advancements hold great promise for unlocking new insights into genomics and improving our understanding of diseases and personalized healthcare. This review article provides an overview of NGS technology and its impact on various areas of research, such as clinical genomics, cancer, infectious diseases, and the study of the microbiome.

Abstract: The advent of next-generation sequencing (NGS) has brought about a paradigm shift in genomics research, offering unparalleled capabilities for analyzing DNA and RNA molecules in a high-throughput and cost-effective manner. This transformative technology has swiftly propelled genomics advancements across diverse domains. NGS allows for the rapid sequencing of millions of DNA fragments simultaneously, providing comprehensive insights into genome structure, genetic variations, gene expression profiles, and epigenetic modifications. The versatility of NGS platforms has expanded the scope of genomics research, facilitating studies on rare genetic diseases, cancer genomics, microbiome analysis, infectious diseases, and population genetics. Moreover, NGS has enabled the development of targeted therapies, precision medicine approaches, and improved diagnostic methods. This review provides an insightful overview of the current trends and recent advancements in NGS technology, highlighting its potential impact on diverse areas of genomic research. Moreover, the review delves into the challenges encountered and future directions of NGS technology, including endeavors to enhance the accuracy and sensitivity of sequencing data, the development of novel algorithms for data analysis, and the pursuit of more efficient, scalable, and cost-effective solutions that lie ahead.

Keywords: next-generation sequencing; genomics; microbiome; molecular diagnostics; bioinformatics; Nanopore; PacBio; Illumina; pyrosequencing

Citation: Satam, H.; Joshi, K.; Mangrolia, U.; Waghoo, S.; Zaidi, G.; Rawool, S.; Thakare, R.P.; Banday, S.; Mishra, A.K.; Das, G.; et al. Next-Generation Sequencing Technology: Current Trends and Advancements. *Biology* **2023**, *12*, 997. https://doi.org/10.3390/biology12070997

Received: 22 June 2023
Revised: 9 July 2023
Accepted: 11 July 2023
Published: 13 July 2023
Corrected: 24 April 2024

1. Introduction

Next-generation sequencing (NGS) has revolutionized genomics, expanding our knowledge of genome structure, function, and dynamics. This groundbreaking technology has enabled extensive research and allowed scientists to explore the complexities of genetic information in unprecedented ways. With its high-throughput capacity and cost-effectiveness, NGS has become a fundamental tool for researchers across diverse disciplines, from basic biology to clinical diagnostics [1]. NGS has not only enabled comprehensive genome sequencing but also facilitated transcriptomics, epigenomics, metagenomics, and other omics studies [2]. The advent of advanced NGS platforms, such as Illumina, Pacific

Biosciences, and Oxford Nanopore, has transformed the field of genomics by allowing for the parallel sequencing of millions to billions of DNA fragments [3,4]. This capability has unlocked new opportunities for understanding genetic variation, gene expression, epigenetic modifications, and microbial diversity. NGS has been instrumental in identifying disease-causing variants, uncovering novel drug targets, and shedding light on complex biological phenomena, including the heterogeneity of tumors and developmental processes [3–5]. This review provides a comprehensive overview of NGS technology, highlighting its transformative impact in various fields, including clinical genomics, cancer research, infectious disease, surveillance, and microbiome analysis. We also discuss the future prospects of NGS, including emerging technologies, its potential for advancing genomics research, and its applications in the biomedical sciences.

2. Generations of Sequencing Technologies

Technologies for "reading" DNA sequences have evolved rapidly over the past two decades [6–10]. This rapid progress has paved the way for significant breakthroughs in the field of DNA sequencing, leading to the emergence of three generations of sequencing technologies (Figure 1).

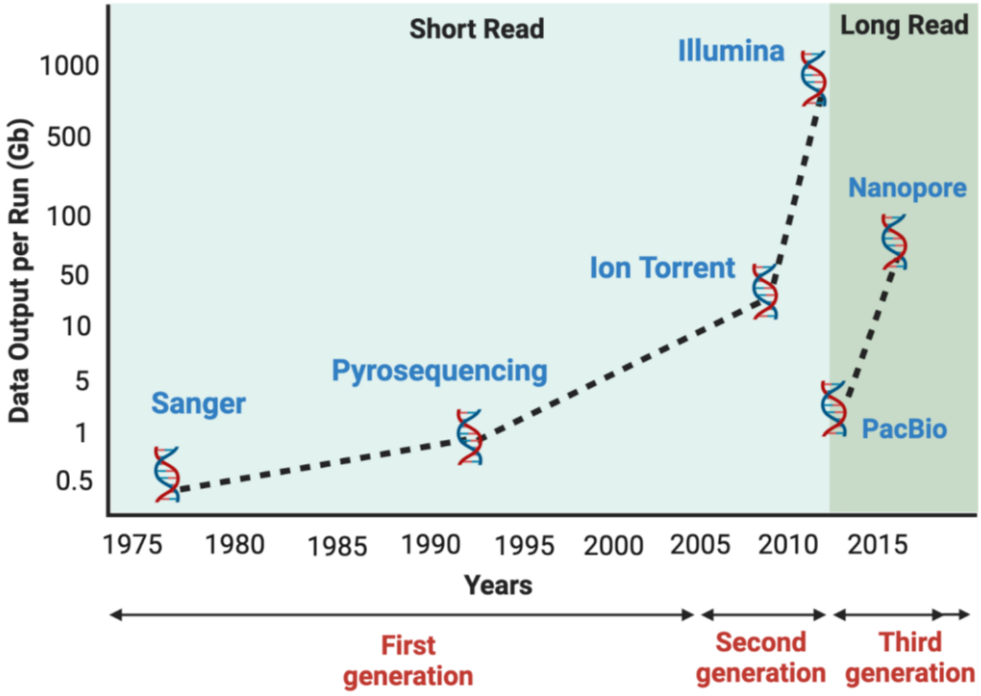

Figure 1. Evolution of sequencing technologies. The development of sequencing technologies over the past four decades can be categorized into three generations. The first generation was represented by Sanger sequencing, providing the foundation for DNA sequencing. The second generation introduced massively parallel sequencing with platforms such as Illumina and Ion Torrent, enabling high-throughput sequencing. The current third generation includes PacBio and Nanopore, offering long-read and single-molecule sequencing capabilities.

2.1. First-Generation Sequencing Technology

The first attempts at sequencing DNA and RNA involved chemical degradation or enzymatic cleavage of the molecules to generate fragments that could be analyzed individually. Robert Holley was the first to sequence a nucleic acid molecule, Alanine tRNA, in 1964 using ribonuclease from *S. cerevisiae* [11]. Similarly, Walter Gilbert and Allan Maxam developed a chemical degradation technique that allowed the sequencing of complete bacteriophage PhiX174 [12]. However, the real breakthrough came with the introduction of the chain termination-based sequencing method by Fredrick Sanger [13]. This technique

used dideoxynucleotides, which terminate the chain elongation of DNA strands during replication, and allowed for the production of sequence reads of up to a few hundred nucleotides in length. Sanger's method was widely adopted and revolutionized the field of molecular biology by allowing for the rapid sequencing of DNA and RNA [12]. In 1987, the first commercial automated sequencing machine, the Applied Biosystems ABI 370, was launched in the United States. This machine used fluorescently labeled dideoxynucleotides and capillary electrophoresis to automate the Sanger sequencing method, significantly increasing the speed and accuracy of DNA sequencing [14,15]. The ABI 370 quickly became the industry standard, and subsequent improvements in the technology led to the development of higher-throughput sequencers capable of producing longer reads [15,16]. While the first-generation technology has been largely superseded by newer, higher-throughput sequencing technologies, it remains an important historical milestone in the development of sequencing techniques. The ability to sequence DNA and RNA has revolutionized many areas of biology and medicine and has led to numerous discoveries and advancements in the understanding of genetics and molecular biology.

2.2. Second-Generation Sequencing Technologies

Second-generation sequencing methods have revolutionized DNA sequencing by enabling the simultaneous sequencing of thousands to millions of DNA fragments. These methods differ from traditional Sanger sequencing in their ability to perform parallel sequencing. Several widely used second-generation sequencing platforms have emerged, one of which is Roche's 454 sequencing method, which relies on pyrosequencing, where the sequence is determined by detecting the release of pyrophosphate when nucleotides are added to the DNA template. Another platform is Ion Torrent sequencing, which detects the release of hydrogen ions during DNA synthesis to determine the sequence. The widely used Illumina sequencing platform utilizes a sequencing-by-synthesis method based on reversible dye terminators. Another upcoming technology, SOLiD sequencing (Sequencing by Oligonucleotide Ligation and Detection), employs a ligation-based approach using reversible terminators to determine the DNA sequence. These second-generation sequencing technologies have significantly increased the throughput and speed of DNA sequencing, enabling a wide range of applications in genomics research and clinical diagnostics [17]. These platforms have enabled whole-genome sequencing, transcriptome analysis, and targeted sequencing, leading to breakthroughs in genetic variation, disease research, and personalized medicine. Many developments in the second generation of sequencing methods have been achieved over the years and are represented in Figure 2 and briefly described in Table 1.

2.3. Third-Generation Sequencing

Third-generation sequencing technologies represent the latest advancements in DNA sequencing, offering new approaches that overcome the limitations of previous generations. These technologies provide long-read sequencing capabilities, enabling the sequencing of much larger DNA fragments compared to earlier methods. Examples include PacBio Sequencing, which uses a single-molecule, real-time (SMRT) approach with fluorescently labeled nucleotides, enabling long-read sequencing of DNA fragments up to tens of kilobases in length. Another technology is Oxford Nanopore sequencing, based on nanopore technology, where a single-stranded DNA molecule passes through a nanopore, and changes in electrical current are measured to determine the DNA sequence. Oxford Nanopore sequencing provides long-read lengths, portability, and real-time analysis. Third-generation sequencing methods have been summarized in Table 1. Figure 3 describes technologies available on NGS and the type of data generated in each type of NGS assay and their brief application.

Table 1. Different generations of NGS platforms.

Sr No.	Platform	Use	Sequencing Technology	Amplification Type	Principle	Read Length (bp)	Limitations	Ref.
1	454 pyrose-quencing	Short read sequencing	Seq by synthesis	Emulsion PCR	Detection of pyrophosphate released during nucleotide incorporation.	400–1000	May contain deletion and insertion sequencing errors due to inefficient determination of homopolymer length.	[18–20]
2	Ion Torrent	Short read sequencing	Seq by synthesis	Emulsion PCR	Ion semiconductor sequencing principle detecting H^+ ion generated during nucleotide incorporation.	200–400	When homopolymer sequences are sequenced, it may lead to loss in signal strength.	[19–21]
3	Illumina	Short read sequencing	Seq by synthesis	Bridge PCR	Solid-phase sequencing on immobilized surface leveraging clonal array formation using proprietary reversible terminator technology for rapid and accurate large-scale sequencing using single labeled dNTPs, which is added to the nucleic acid chain.	36–300	In case of sample overloading, the sequencing may result in overcrowding or overlapping signals, thus spiking the error rate up to 1%.	[19,20,22]
4	SOLiD	Short read sequencing	Seq by ligation	Emulsion PCR	An enzymatic method of sequencing using DNA ligase. 8-Mer probes with a hydroxyl group at 3′ end and a fluorescent tag (unique to each base A, T, G, C) at 5′ end are used in ligation reaction.	75	This platform displays substitution errors and may also under-represent GC-rich regions. Their short reads also limit their wider applications.	[20,23]
5	DNA nanoball sequencing	Short read sequencing	Seq by ligation	Amplification by Nanoball PCR	Splint oligo hybridization with post-PCR amplicon from libraries helps in the formation of circles. This circular ssDNA acts as the DNA template to generate a long string of DNA that self-assembles into a tight DNA nanoball. These are added to the aminosilane (positively charged)-coated flow cell to allow patterned binding of the DNA nanoballs. The fluorescently tagged bases are incorporated into the DNA strand, and the release of the fluorescent tag is captured using imaging techniques.	50–150	Multiple PCR cycles are needed with a more exhaustive workflow. This, combined with the output of short-read sequencing, can be a possible limitation.	[24,25]
6	Helicos single-molecule sequencing	Short-read sequencing	Seq by synthesis	Without Amplification	Poly-A-tailed short 100–200 bp fragmented genomic DNA is sequenced on poly-T oligo-coated flow cells using fluorescently labeled 4 dNTPS. The signal released upon adding each nucleotide is captured.	35	Highly sensitive instrumentation required. As the sequence length increases, the percentage of strands that can be utilized decreases.	[26,27]
7	PacBio Onso system	Short-read sequencing	Seq by Binding	Optional PCR	Sequencing by binding (SBB) chemistry uses native nucleotides, scarless incorporation under optimized conditions for binding and extension (https://www.pacb.com/technology/sequencing-by-binding/, accessed on 1 July 2023).	100–200	The higher cost compared to other sequencing platforms	

Table 1. *Cont.*

Sr No.	Platform	Use	Sequencing Technology	Amplification Type	Principle	Read Length (bp)	Limitations	Ref.
8	PacBio Single-molecule real-time sequencing (SMRT) technology	Long-read sequencing	Seq by synthesis	Without PCR	The SMRT sequencing employs SMRT Cell, housing numerous small wells known as zero-mode waveguides (ZMWs). Individual DNA molecules are immobilized within these wells, emitting light as the polymerase incorporates each nucleotide, allowing real-time measurement of nucleotide incorporation	average 10,000–25,000	The higher cost compared to other sequencing platforms.	[28,29]
9	Nanopore DNA sequencing	Long-read sequencing	Sequence detection through electrical impedance	Without PCR	The method relies on the linearization of DNA or RNA molecules and their capability to move through a biological pore called "nanopores", which are eight nanometers wide. Electrophoretic mobility allows the passage of linear nucleic acid strand, which in turn is capable of generating a current signal.	average 10,000–30,000	The error rate can spike up to 15%, especially with low-complexity sequences. Compared to short-read sequencers, it has a lower read accuracy.	[5,19,30]

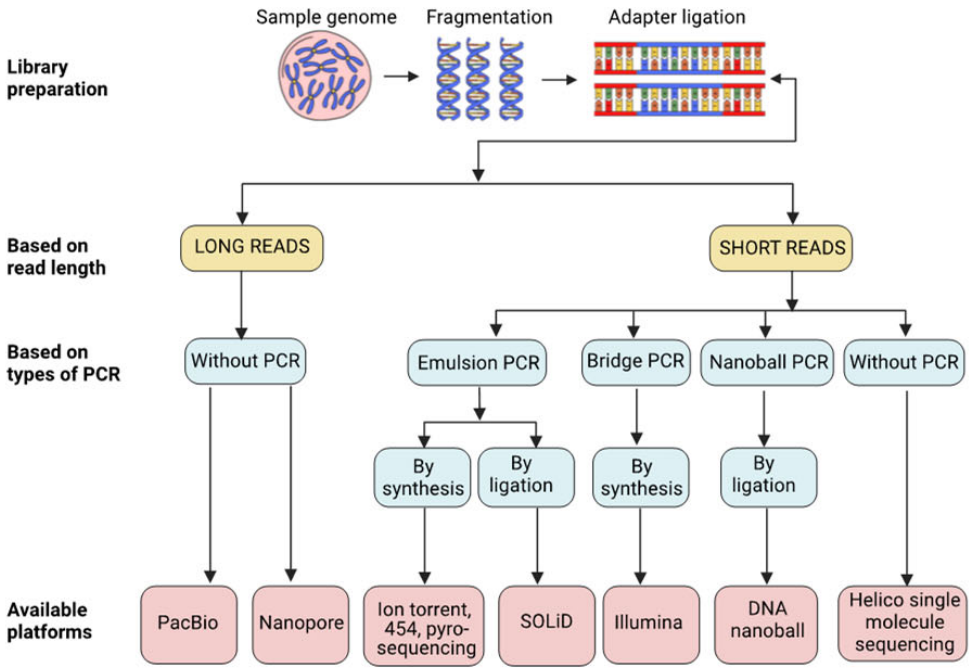

Figure 2. Overview of various NGS technologies with different platforms and principles.

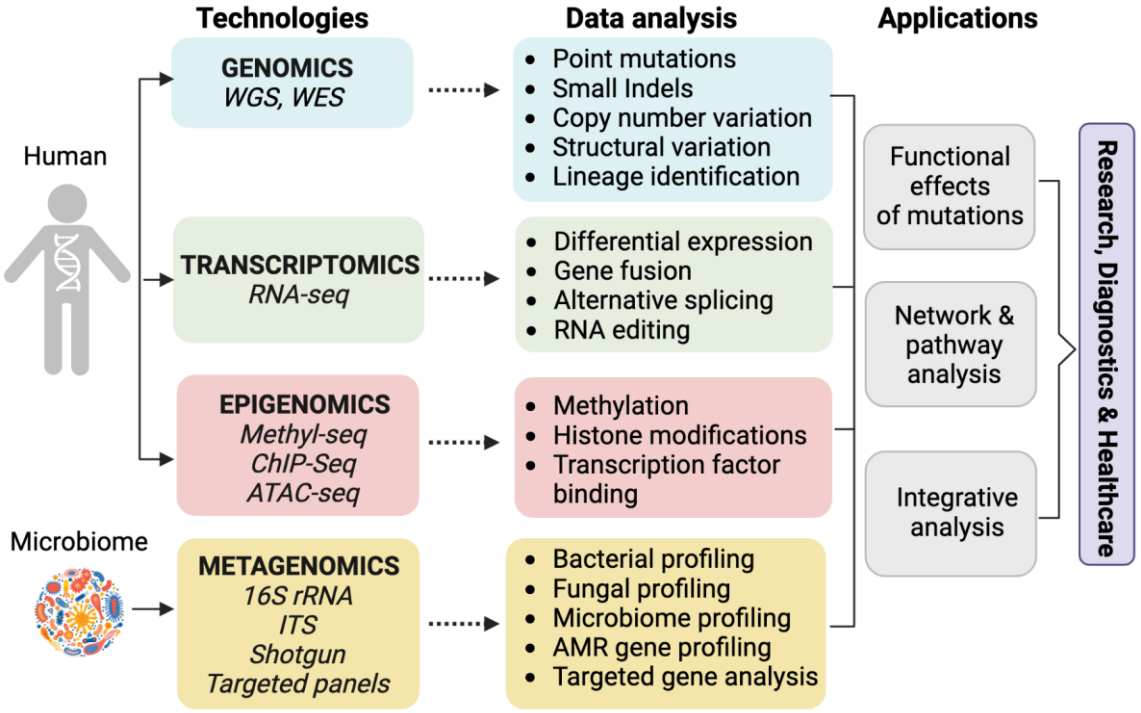

Figure 3. Various approaches used for genome analysis and applications of NGS, including technological platforms, data analysis, and applications. WGS, whole-genome sequencing; WES, whole-exome sequencing; Seq, sequencing; ITS, internal transcribed spacer; ChIP, chromatin immunoprecipitation; ATAC, assay for transposase-accessible chromatin; AMR, anti-microbial resistance.

Long-Read and Short-Read Sequencing

The basic principle for short-read sequencing involves sequencing by synthesis based on enrichment through hybridization, amplification, or fragmentation. Whereas long-read sequencing works on sequence detection either by synthesis or by electrical voltage change/impedance, generating the current as a single base is passed through the biological membrane pore. Long-read sequencing can generate reads up to 25–30 kb, whereas short-read sequencing can generate reads around 600–700 bp. Furthermore, the amplification bias is eliminated in long-read sequencing as opposed to short-read sequencing. As the library preparation is PCR-free, the base modification such as DNA methylation can be easily detected by long-read sequencing. The introduction of high-throughput sequencing platforms has significantly reduced error rates and notably improved the accuracy of long-read sequencing technologies [29,31]. Short-read sequencing is useful for determining the abundance of specific sequences, profiling transcript expression, and identifying variants. However, long-read sequencing technologies excel in providing comprehensive genome coverage, enabling researchers to identify complex structural variants such as large insertions, deletions, inversions, duplications, and more [8,29,31].

3. Next-Generation Sequencing-Based Omics

Understanding complex human diseases requires data integration from multiple omics techniques such as genomics, transcriptomics, epigenomics, and proteomics. Here, we briefly describe various omics technologies that are implemented on the NGS platform:

3.1. Genomics

Genomics studies using NGS profoundly analyze DNA using various approaches such as whole-genome sequencing, whole-exome sequencing, and targeted sequencing.

3.1.1. Whole-Genome Sequencing

Whole-genome sequencing (WGS) is a powerful and comprehensive genomic analysis technique that involves determining the complete DNA sequence of an individual's genome. It provides a detailed blueprint of an individual's genetic makeup, encompassing all the genes, regulatory regions, and non-coding elements present in their genome. It finds its application mainly in discovery science, such as plant and animal research, cancer research, rare genetic diseases, patients with complex disease symptoms, population genetics, and novel genome assembly of eukaryotes and prokaryotes [32]. 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. This wealth of information obtained through WGS offers a multitude of applications in various fields [33]. WGS has two types of sequencing approaches on the basis of genome size viz. (1) large whole-genome sequencing deciphering larger genomes of >5 Mb such as eukaryotes, and (2) small whole-genome sequencing deciphering smaller genomes of <5 Mb mainly of prokaryotes. Short-read sequencing is preferred for mutation calling, while long-read sequencing is preferred for genome assemblies. Combining short and long-read sequencing for sequencing novel genomes has been successfully applied for accurate genome assembly without a reference sequence.

3.1.2. Whole-Exome Sequencing

Whole-exome sequencing (WES) is a sequencing approach that focuses on capturing and sequencing the protein-coding regions of the genome, known as the exome. The exome represents approximately 1–2% of the entire genome but contains the majority of known disease-causing variants. By sequencing the exome, WES enables the identification of genetic variations, including single-nucleotide variants (SNVs), insertions, deletions, and copy number variations (CNVs), within protein-coding genes [34,35]. WES is a cost-effective alternative to WGS for rare clinical diseases with clusters of symptoms, as well as in identifying variants for population and cancer genetics [36]. WES involves the enrichment of exonic regions using hybrid capture or target-specific amplification techniques, followed by high-throughput sequencing. Various exome capture assays from NimbleGen, Agilent, Illumina, Twist, and IDT are available that are compatible with the Illumina NGS platform [37]. The bioinformatic approach used for WES data analysis is the same as that of WGS since WES is a part of WGS.

3.1.3. Targeted Sequencing

Targeted sequencing, as the name suggests, has less exploratory power than WGS or WES as it targets specific regions of the gene and is able to pick up various types of genetic variations from SNVs to small gene deletions, duplications, insertions, or gene rearrangements associated with disease phenotypes. However, advantages include cost-effectiveness and manageable data for clinicians, making clinical decisions easier with more specific disease-relevant information. It can give much deeper coverage up to 5000× for rare alleles in genetic diseases, as well as for low-abundant evolving mutant clones arising as a result of tumor heterogeneity or disease evolution in cancer [38]. The candidate gene approach or commercially available targeted panels is the result of WGS/WES projects carried out at the population scale. The germline, as well as somatic variants, can be tested using targeted NGS panels, few examples of which are listed in Table 2. Targeted panels work on a simple approach of enrichment by amplification using pools of region-specific oligonucleotide primers. Specific size libraries that are produced are then sequenced and analyzed bioinformatically [39].

Table 2. Examples of targeted panels available in research and diagnostic settings.

Disease Condition	Available Panel	Type of Inheritance	Specimen Type
Inherited cardiovascular defects	Cardiovascular research panel	Germline	Blood
Arrhythmias and cardiomyopathies	Arrhythmias and cardiomyopathy research panel	Germline	Blood
Sensitivity to pharmacological drugs	Pharmacogenomics research panel (PGex Seq panel)	Germline	Blood
Antimicrobial treatment efficacy testing	Antimicrobial resistance research panel	Microbial gene testing	Bacterial culture
Infertility conditions	Infertility research panel	Germline	Blood
Homologous recombination defect analysis	HRR gene panel	Somatic	Tumor tissue
myeloid cancers	Myeloid cancer panel	Somatic	Blood
HIV speciation and drug resistance	HIV-Xgene panel	Pathogen detection	HIV-positive plasma
Antimicrobial resistance in MTB	TB research panel	Pathogen detection	MTB-positive specimen
Inborn errors of metabolism	Error of metabolism research panel	Germline	DBS/blood
Hereditary cancers	BRACA and extended breast and ovarian cancer research panel, inherited cancer research panel	Germline	Blood

3.2. Transcriptomics

Next-generation sequencing (NGS) has had a transformative impact on transcriptomics, revolutionizing our ability to study the transcriptome—the complete set of RNA molecules in an organism or specific cell population. NGS technologies offer high-throughput and cost-effective methods for profiling and analyzing RNA molecules, allowing researchers to gain deep insights into gene expression, alternative splicing, non-coding RNA regulation, and various biological processes and diseases [40–43]. Here are some key roles of NGS in transcriptomics:

(a) mRNA Sequencing (RNA-Seq): RNA-seq is a widely used NGS application in transcriptomics. It involves the sequencing and quantification of mRNA molecules, providing a comprehensive snapshot of the expressed genes in a biological sample. By generating millions of short sequencing reads, NGS allows researchers to identify and quantify gene expression levels accurately. RNA-seq data can be analyzed to detect differential gene expression between different conditions, discover novel transcripts, assess alternative splicing events, and study gene expression dynamics over time or across different tissues or cell types [44,45].

(b) Alternative Splicing Analysis: Alternative splicing, a process in which a single gene can generate multiple mRNA isoforms, significantly contributes to transcriptome complexity. NGS provides the ability to study alternative splicing patterns comprehensively. By aligning RNA-seq reads to the reference genome, researchers can identify splice junctions and detect alternative splicing events. This information allows for the quantification and characterization of transcript isoforms, providing insights into isoform diversity, tissue-specific expression, and the functional implications of alternative splicing [46].

(c) Long Non-Coding RNA (lncRNA) and Small-RNA Analysis: NGS facilitates the study of non-coding RNAs, which play critical roles in gene regulation. Techniques such as small-RNA sequencing and long non-coding RNA sequencing enable the identification and characterization of various classes of non-coding RNAs. Small-RNA sequencing allows the profiling of small regulatory RNAs, including microRNAs, piRNAs, and snoRNAs, providing insights into their roles in post-transcriptional gene regulation. Long non-coding RNA sequencing enables the identification and analysis of long non-coding RNA transcripts, which have been implicated in diverse biological processes and diseases [47–49]. Long RNA-seq reads can inform about the connectivity between multiple exons and reveal sequence variations (SNPs) in the transcribed region [50]. Small-RNA sequencing is a non-targeted approach that allows the detection of novel miRNA and other small RNAs [51]. The transcriptome with ChIP-seq studies in cancer biology has helped to understand the emerging role

of ncRNAs such as sncRNAs and lncRNA in gene regulation mechanisms during carcinogenesis/cancer progression [52–54].

(d) Transcriptome Assembly and Annotation: NGS data can be utilized to reconstruct and annotate the transcriptome of an organism. By aligning RNA-seq reads to a reference genome or using de novo assembly approaches, researchers can identify novel transcripts, splice variants, untranslated regions, and other transcript features. This information enhances our understanding of the transcriptome's complexity and improves the annotation of reference genomes, enabling the discovery of previously unknown genes and regulatory elements [55].

(e) Single-Cell Transcriptomics: NGS has facilitated the emergence of single-cell transcriptomics, enabling the study of gene expression profiles at the individual cell level. Single-cell RNA-seq (scRNA-Seq) technologies allow the profiling of transcriptomes from individual cells, providing insights into cellular heterogeneity, cell type identification, cell lineage analysis, and gene expression dynamics in complex tissues or developmental processes [56,57].

(f) Integrative Transcriptomics: NGS data from transcriptomics can be integrated with other omics data, such as genomics, epigenomics, and proteomics, to gain a comprehensive understanding of gene regulation and biological processes. Integrative approaches provide a system-level view of molecular interactions and enable the identification of key regulatory mechanisms underlying cellular processes and diseases [56].

3.3. Epigenomics

Epigenomics refers to the study of epigenetic modifications, which are heritable changes in gene expression patterns that do not involve alterations in the DNA sequence [58,59]. The most common types of epigenetic modifications studied are DNA methylation [60], histone modification, and RNA methylation (epi-transcriptome). These chemical tags in turn alter DNA accessibility, chromatin remodeling, and nucleosome positioning [61]. These modifications are influenced by environmental factors such as nutrients, pollutants, toxicants, and inflammation [62,63]. The knowledge and data generated through whole-genome-wide sequencing in humans, plants, and animals [64] have helped scientists to gain better insights into these epigenetic alterations, especially DNA methylation and hydroxymethylation. Epigenetic alterations have attracted researchers' and clinicians' interest in complex disorders such as behavioral disorders, memory, cancer, autoimmune disease, addiction, neurodegenerative, and psychological disorders [65]. There are various platforms and assays developed to study epigenetic modifications, which have been very well described elsewhere [66]. NGS has been utilized for investigating epigenomics, as discussed below:

(a) DNA Methylation Profiling: DNA methylation is a crucial epigenetic modification that plays a critical role in gene regulation and cellular processes. NGS enables genome-wide profiling of DNA methylation patterns at single-nucleotide resolution [67]. Several strategies, such as whole-genome bisulfite sequencing (WGBS) and reduced representation bisulfite sequencing (RRBS), leverage NGS to identify methylated cytosines [68]. However, RRBS is based on enriching methylated genomic regions using restriction enzymatic digestion [66,69]. These methods allow researchers to study DNA methylation dynamics, uncover differentially methylated regions (DMRs) associated with diseases, and understand the impact of methylation on gene expression.

(b) Chromatin Accessibility Mapping: NGS-based techniques, such as assay for transposase-accessible chromatin using sequencing (ATAC-seq) and DNase-seq, enable the genome-wide profiling of chromatin accessibility. These methods identify regions of the genome that are accessible to DNA-binding proteins and transcription factors, providing insights into gene regulatory elements, enhancers, and promoters. By combining chromatin accessibility data with other epigenetic modifications, gene expression data,

and transcription factor binding data, researchers can unravel the functional elements within the genome [70,71].

(c) Histone Modification Analysis: Histone modifications, including acetylation, methylation, phosphorylation, and more, are critical epigenetic marks that regulate chromatin structure and gene expression. Chromatin immunoprecipitation sequencing (ChIP-seq) enables genome-wide profiling of histone modifications by antibody-based pull down of the protein followed by enrichment of DNA bound to the protein and sequencing. This technique finds application in many different areas of research, such as transcription factor (TF) binding site identification, histone modification analysis of the DNA, and DNA methylation. For studying histone modifications, antibodies targeted to histone modifications are used to pull down the DNA and sequenced using the NGS technique. The resulting reads are aligned to the reference genome, enabling the identification of histone modification patterns at specific genomic regions. ChIP-Seq can provide insights into the epigenetic regulation of gene expression, chromatin states, and the identification of enhancers and other regulatory elements [72–75].

(d) Chromatin Conformation Analysis: NGS-based techniques, such as Hi-C and 4C-seq, allow the investigation of 3D chromatin organization and interactions. These methods capture long-range chromatin interactions and enable the construction of chromatin interaction maps [76,77]. By integrating 3D chromatin conformation data with epigenetic modifications, gene expression data, and functional annotations, researchers can gain insights into the spatial organization of the genome and understand how it influences gene regulation.

(e) In addition to these standalone approaches, NGS data from epigenomics can be integrated with transcriptomics data to unravel the relationship between epigenetic modifications and gene expression. Integration of DNA methylation profiles with RNA-seq data can identify differentially methylated regions (DMRs) associated with gene expression changes. Integration of histone modification and chromatin accessibility data with RNA-seq allows the identification of regulatory elements associated with specific gene expression patterns and the exploration of epigenetic regulatory mechanisms.

3.4. Metagenomics

Metagenomics deals with direct genetic analysis of the prokaryotic genome including bacteria, fungi, and viruses contained in a sample [78] either by targeted approach or adaptor ligation PCR approach for shotgun sequencing in a culture-independent manner. The hypervariable region in 16S or 18S ribosomal RNA genes of bacteria and fungi is used in the targeted approach. A blend of conserved and hypervariable regions helps in the identification of each bacterial species from the sample. Similarly, for fungal species identification, ITS1 and ITS2 regions spanning the 5.8S rRNA gene of the fungal genome are selected for amplification [79]. For viral genome sequencing, reads generated from NGS (shotgun) are again the culture-independent method for studying viral diversity, abundance, and functional potential of viruses in the environment. All filtered reads are mapped with the human reference sequence, and remaining, unmapped reads are mapped against the NCBI RefSeq viral genomic database (Table 3) [80]. The targeted viral and bacterial genome panels are also available, e.g., ChapterDx for HR HPV and microbial infection detection, the HIV drug resistance panel, the AMR panel, the gastrointestinal disorder panel, etc.

Based on the nucleotide sequence similarities, pre-processed sequences are clustered at 97% similarity into operational taxonomic units (OTUs). OTUs are compared with the database to identify the microorganisms [81]. Several analysis pipelines are used for the analysis of 16S amplicon reads (Table 3) [82]. For shotgun metagenomics samples, taxonomic and functional profiles can be obtained by different approaches, as elaborated in Table 3 [83–89]. Microbiome sequencing can identify the full spectrum of microbial species present in the sample. The results are highly quantitative, and one can study the bacterial

communities over a specific interval of conditions. The NGS platform can also generate reads for low-abundance species in a sample.

4. Bioinformatic Approaches for NGS Data Analysis

NGS generates vast amounts of DNA or RNA sequences, necessitating computational methods to handle, analyze, and interpret these data. Raw sequencing data produced by NGS instruments need to be processed, analyzed, and interpreted to derive biological insights. This is where bioinformatic approaches come into play. These approaches encompass a wide range of computational methods, algorithms, and tools that handle preprocessing, alignment, variant calling, gene expression quantification, differential expression analysis, and other specialized analyses. Once processed, various computational techniques, such as de novo assembly, reference-based mapping, and transcriptome analysis, are employed to extract meaningful biological information. Furthermore, advanced bioinformatic tools facilitate the identification of genetic variations, including single-nucleotide polymorphisms (SNPs), copy number variations (CNVs), and structural variants. Integrative analyses, combining NGS data with other genomic and functional data sources, enable the exploration of gene expression and regulatory networks. The various bioinformatics tools used in NGS analysis are listed in Table 3.

Table 3. Bioinformatic steps and tools used for NGS data analysis.

Analysis	Commonly Used Tools
Common Analysis	
Quality check of sequences	FastQC [90], FASTX-toolkit [91], MultiQC [92]
Trimming of adaptors and low-quality bases	Trimmomatic [93], Cutadapt [94], fastp [95]
Alignment of sequence reads to reference genome	BWA [96], Bowtie [97], dragMAP [98]
Reports visualization	MultiQC [92]
Whole-Genome Sequencing/Whole-Exome Sequencing/Targeted Panel	
Removal of duplicate reads	Picard [99], Sambamba [100]
Variant calling (single-nucleotide polymorphisms and indels)	GATK [101], freeBayes [102], Platypus [103], VarScan [104], DeepVariant [105], Illumina Dragen [106]
Filter and merge variants	bcftools [107]
Variant annotation	ANNOVAR [108], ensemblVEP [109], snpEff [110], NIRVANA [111]
Structural variant calling	DELLY [112], Lumpy [113], Manta [114], GRIDDS [115], Wham [116], Pindel [117]
Copy number variation (CNV) calling	CNVnator [118], GATK gCNV [119], cn.MOPS [120], cnvCapSeq(targeted sequencing) [121], ExomeDepth (CNVs from Exome) [122]
Transcriptomics	
Alignment of reads to reference	Splice-aware aligner such as TopHat2 [123], HISAT2 [124], and STAR [125]
Transcript quantification	featureCounts [126], HTSeq-count [127], Salmon [128], Kallisto [129]
Differential gene expression analysis	DESeq2 [130], EdgeR [131], DAVID [132], clusterProfiler [133],
enrichment of gene categories	Enrichr [134]
Epigenomics-Methyl Seq	
Sequence aligners	Bwameth [135], BS-Seeker2 [136], Bismark [137]
Methylation level quantification	MethylDackel *
Differential methylation	Metilene [138], BSsmooth [139], methylKit [140]
Epigenomics-ChIP seq	
Removal of PCR duplicates	Samtools [107]
Peak calling	MACS2 [141], SICER2 [142], SPP [143]
Peak filtering	Bedtools [144]
Enrichment quality control	ChipQC [145], Phantompeakqualtools [146]
Enrichment comparison	diffBind [147], MAnorm [148], MMDiff [149]
Motif analysis	MemeCHiP [150], Homer [151], RSAT [152]

Table 3. *Cont.*

Analysis	Commonly Used Tools
16s rRNA seq	
16S rRNAseq analysis pipelines	QIIME2 [82], mothur [153], USEARCH [154]
Ribosomal RNA databases	Greengenes [155], Silva [156], RDP [157]
Shotgun Metagenomics	
Taxonomic classification	MetaPhlAn4 [158], Kaiju [159], Kraken [160]
Assembly of metagenomic reads	metaSPAdes [86], metaIDBA [87]
Protein databases for taxonomic classification	NCBI non-redundant protein database [83]
Gene annotation	Prokka [88], MetaGeneMark [89]
Databases for functional annotation of genes	COG [161], KEGG [84], GO [85]

Footnote: 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. * Available at: https://github.com/dpryan79/MethylDackel/ (accessed on 1 June 2023). Bold represents the categories of analysis and commonly used bioinformatics tools used for NGS data analysis.

5. NGS Applications in Research and Diagnostics

NGS has revolutionized the field of scientific research and clinical genomics due to high-throughput multiplexing. This power of NGS in translation medicine lies not only in its advanced multiplexing efficiency but also in the equally smart bioinformatic tools used for data curation followed by various reference databases that help researchers, medical practitioners, and drug designers to understand the genetic basis of the disease. Different population genome sequencing projects such as 1000 G, ExAC, ESP6500, UK 100 K, Indigenome, and gnomAD generated vast amounts of data on NGS [162]. Among the reference population databases, gnomAD is the largest and most widely used database generated from harmonized sequencing data incorporating exome and genome sequencing data from 140,000 humans. This has been widely used as a resource for estimating allele frequency in rare diseases, disease gene discovery, and the biological effect of variation [163]. This has led to the creation of knowledge bases and in turn large and small sequencing panels for major applications in clinical research and diagnostics [164]. The large gene panels find their major application in clinical research mainly in cancer genetics.

5.1. Role of NGS in Research

5.1.1. Microbiome Research

Given the ubiquitous nature of microbes, their symbiotic, pathogenic, and commensal characteristics are of importance to humans by forming a highly functioning ecosystem. The microbiome community became an obligatory factor in our survival through evolution [165]. However, a close monitoring and comprehensive understanding of the host–microbiome and microbiome–intercommunity interactions are vital to healthy survival. The approaches include pathogen surveillance, functional dysbiosis, and therapeutic potential. Metagenomic studies have linked the gut microbiome to disorders affecting mental health [166], autoimmune diseases (rheumatoid arthritis) [167], and metabolic disorders (diabetes and obesity) [168], thus instrumental in evaluating the functional potential of the microbiome. This opens doors for more therapeutic approaches and options. Designing targeted panels to pick up mutations (aiding in antibiotic resistance tracking) or identifying the pathogenic genes followed by sequencing can help in detecting pathogens with known antimicrobial

resistance. Research is also underway for the pharmacomicrobiomics of individuals requiring drug treatment. This would aid in identifying the effect of drugs on an individual's microbiome and drug disposition by the microbiome.

5.1.2. Human Disease Research

The focus of NGS-based research is now extended from genomic research to the study of transcriptome, epi-transcriptome, and epigenome. Human genome-based research through WGS and WES has provided novel insights into the biological processes and has found application in wellness research; agriculture and food research; genome-wide association research studies uncovering the wide range of population genetic variants; their genetic linkage and molecular basis to various diseases, including cancer; and the study of new pathogenic/emerging variants such as SARS-CoV-2 variants in human diseases. The redefinition of the mutational landscapes in tumors has resulted in translating this information into clinical research through the ever-growing list of targeted large gene panels such as the 261 gene panel, the 400 gene panel, the TSO 500 panel from Illumina, IDT, Agilent, and Thermo Fisher. These panels assess not only SNVs but also clinically relevant CNVs and RNA fusion transcripts, TMB, and microsatellite instability (MSI) for lung cancer, breast cancer, colorectal cancer, and even for difficult cancers such as ovarian, pancreatic, renal, urothelial cancers, etc.

RNA-seq finds its application mainly in research for analyzing pathogen transcriptomic signatures [169], metastatic biomarkers, therapeutic resistance, immune microenvironment, immunotherapy, and neoantigen research in cancer [170,171]. With NGS, it is now possible to study single-cell behavior with respect to its differentiation, de-differentiation, proliferation, and tumorigenesis in cancer using single-cell RNA-sequencing strategies such as Smart-seq2, MATQ-seq, SUPeR-seq, Drop-seq, Seq-Well, Chromium, DroNC-seq, STRT-seq, etc. [172]. The recent new development of the RiboSeq technique can plot potential ongoing events of translation in the cytosol, which is useful in identifying potentially functional micro-peptides. This is how thousands of sORFs (small open-reading frames) were discovered in lncRNA. Thus, with transcriptomics, Ribo-seq, and MS proteomics, the bifunctional potential of RNA molecules is identified [173,174].

The role of epigenomics in gene regulation, the maintenance of tissue-specific expression, and developmental processes is evident from X chromosome inactivation, embryonic development, genomic imprinting, epigenetic reprogramming, cell identity establishment, and lineage specification studies. Epigenetic signatures are important biomarkers that have promise not only in cancer, malignant transformation, and metastasis but also for their clinical applicability in other disease conditions such as diabetes, neurological conditions, infectious diseases, and immune disorders [175,176]. The reversible nature of epigenetic changes makes them promising candidates for precision medicine in cancer and other conditions [164,176]. Pharmacoepigenomics is an emerging research area, where the relationship between variable drug response and epigenetic status is being studied [59]. Epi-drugs have been developed over the last 40 years, and few are in clinical practice, whereas some are in clinical trials [177]. Non-coding RNAs (ncRNAs) are gene expression regulators apart from epigenetic modifications that are being explored as drug targets. Numerous lncRNAs are subsequently identified and found to be aberrantly expressed in various tumors [58]. Increasing studies have shown miRNAs as biomarkers of multiple cancers as their abnormal quantity has been correlated with the stage of pathology and prognosis [178]. The applications of miRNA analog or anti-miRNAs have shown promising outcomes in vitro and in vivo cancer studies, suggesting that miRNA-based drugs are emerging as a novel strategy for cancer therapy [179]. Apart from cancer, multiple FDA-approved drugs exist for DMD, SMA, familial hypercholesterolemia, CMV retinitis, etc. [178].

5.2. NGS in Diagnostics

A decisive approach is important when selecting an NGS assay. Type of variant, disease symptoms, and probable genetic associations are important aspects when selecting

NGS-based tests in clinical decision making, as per recommendations by the National Comprehensive Cancer Network (NCCN), the College of American Pathologists (CAP), the American Society of Clinical Oncology (ASCO), the Association of Molecular Pathology (AMP), the American College of Medical Genetics (ACMG), and the European Society of Medical Oncology (ESMO).

5.2.1. Infectious Diseases

The identification of the exact etiological agent in microbial infections is important for precision medicine, which has driven the approach of syndromic testing/multiple pathogen testing assays such as BioFire or multiplex PCRs. However, with the limitations of multiplexing, NGS panels are being developed that can detect any pathogen using a shotgun approach or a targeted approach (16S) from various diseased specimens or clinical isolates. These panels can not only pick up causative pathogens but can be used to identify drug-resistant mutations such as antimicrobial drug-resistant mutations and antiviral drug-resistant mutations [180]. The useful data generated through NGS on microbial identification and drug resistance genotyping, e.g., in MTB, HIV, and SARS-CoV-2 [181], have proven important for disease surveillance, disease containment, public health epidemiological studies, policy making, and rapid therapeutic interventions, as evident during the COVID-19 outbreak [182]. However, with the need for fast diagnosis, NGS, in its current form for infectious pathogen detection, cannot replace current standard point-of-care testing such as PCR, multiplex BioFire panel testing, or multiplex QPCR commercial kits.

5.2.2. Inherited Genetic Diseases

The association of multiple genes in multifactorial disorders such as diabetes, hypercholesterolemia, infertility, etc., has been discovered in the rapidly emerging field of genomics. For example, the classical approach to comprehending the genes participating in infertility, gametogenesis, the hormonal cycle, fecundation, and embryo development would have been difficult and time-consuming. Targeted NGS panels have evolved as a result of WGS, and WES has enabled the simultaneous evaluation of multiple genes and their variants explaining the complexity of various disorders, including infertility, inherited genetic diseases, and reproductive genome testing, including NIPT (non-invasive prenatal testing), PGS/PGD (preimplantation genetic disease testing), and pediatric disorders such as developmental delay disorders, metabolic syndromes [183]. This has enabled disease treatment through personalized genome testing for the betterment of human health, preventive testing, and disease management.

5.2.3. HLA Typing

NGS-based HLA typing using WGS or targeted panels over conventional HLA typing methods for organ transplant or HSCT provides more unambiguous, high-throughput, high-resolution typing results from a single platform. This approach provides complete information on all the HLA loci involved in (1) the etiopathogenesis of immune disorders such as coeliac disease, psoriasis, rheumatoid arthritis, type I diabetes, SLE, lung diseases (e.g., asthma or sarcoidosis) [184], infectious disease predispositions (e.g., HIV, hepatitis, leprosy, tuberculosis), and other conditions such as malignancies and neuropathies [185]) generating population/ancestry-based database.

Epigenetics study through methylation profiling was in fact first studied using the HLA gene, which has its epigenetic regulators located in the non-coding region such as enhancers, promoters, and UTR regions that regulate HLA gene expression. Bioinformatically, the sequence data obtained are analyzed using commercial HLA-specific software such as NGSengine or exome-data-based software such as OptiType [186], Polysolver [187], xHLA [188], and HLAminer [189] to determine the HLA types [190].

5.2.4. Cancer

The comprehensive human genome sequencing project, WGS and WES, has identified cancer as the disease of the genome and is a multifactorial disease with non-mendelian (Somatic) origin in the majority of cases and mendelian origin in inherited cancers. Through the efforts of TCGA (The Cancer Genome Atlas) and ICGC (International Cancer Genome Consortium), the understanding of cancer and the comprehensive gene alteration data in protein-coding regions for all types of human cancers are now readily available [191].

Different enterprises, such as FoundationOne by Foundation Medicine (Cambridge, MA, USA), Oncomine by Thermo Fisher (Waltham, MA, USA), CANCERPLEX by KEW (Cambridge, MA, USA), MSK-IMPACT by the Memorial Sloan Kettering Cancer Center (New York, NY, USA), OmniSeq Advance by the Roswell Park Cancer Institute (Buffalo, NY, USA), the CC Onco Panel by Sysmex (Kobe, Japan), and the Todai Onco Panel by Riken Genesis (Tokyo, Japan) have come up with multigene panels using TCGA and ICGC data for different NGS platforms that are now frequently used in cancer prognosis and therapeutics [191]. Figure 4 summarizes the various data integration methods for cancer diagnosis, prognosis, and therapeutics [192]. Though all alterations picked up in NGS may not find immediate application in translation medicine, they help discover the different pathways operating in cancer pathogenesis and build on the cancer genomics database. Lung cancer biomarkers have been developed for almost over a decade for the development of a commercial NGS panel of 15–21 genes for precision oncology in lung cancer, picking up all types of structural variants (SVs) on a single platform [193,194]. This landmark study of precision oncology in lung cancer opened the doors for various solid tumors such as CRC, breast, ovarian, endometrial, pancreatic, and even liquid tumors such as myeloid and lymphoid malignancies to use NGS panels effectively with limited sample requirement, infrastructure, and different technical and analytical expertise [98]. Thus, a comprehensive gene testing approach in cancer provides maximum treatment efficacy and reduces the window period of disease progression in a cancer patient, resulting in improved QOL (quality of life), PFS (progression-free survival), and OS (overall survival).

One important aspect of somatic mutation testing in cancer is tumor heterogeneity. It needs to be clearly and carefully dealt with by setting the variant calling cutoff thresholds to avoid false-positive or false-negative variant calling and reporting [195]. Being the most sensitive method of mutation detection, evolving mutant clones, the allelic burden of mutation and thus the disease progression can be determined through NGS. Liquid biopsy testing in cancer has become a very handy tool in tracking disease progression and treatment monitoring in clinical oncology using the circulating tumor DNA in a metastatic setting [196]. NGS plays a crucial role in identifying biomarkers associated with hereditary/germline cancers. For example, in the case of hereditary breast and ovarian cancer syndrome (HBOC), the understanding of its genetic basis has evolved beyond the BRCA1 and BRCA2 mutations. The inclusion of other genes involved in the homologous recombination repair (HRR) pathway, known as BRACAness genes, has reshaped our understanding of HBOC. These additional genes include CDH1, PTEN, TP53, STK11, PALB2, ATM, CHEK2, MUTYH, BARD1, MRE11A, NBN, RAD50, RAD51C, RAD51D, and NF1, in addition to BRCA1 and BRCA2. NGS has facilitated the identification and characterization of these extended sets of genes associated with HBOC, expanding our knowledge of hereditary cancer predisposition [197].

5.3. NGS in Forensics

Ever since 1984, when Sir Alec Jeffreys first proposed the application of DNA profiling to distinguish between different samples at a crime site, DNA analysis has emerged as a prime investigative tool in forensic science [198]. This field is now being dominated by NGS, keeping behind the old methods of DNA fingerprinting such as restriction fragment length polymorphism (RFLP), mitochondrial DNA, variable number of tandem repeat (VNTR) profiling, and short tandem repeat (STR) typing to solve an array of criminal mysteries [199]. NGS has gained rapid importance in this domain due to its ability to

deliver highly accurate, reproducible, and results of the highest sensitivity from highly contaminated and degraded sample qualities received in forensic labs [200]. NGS is being applied to solve different categories of criminal cases: mtDNA for the investigation of maternal lineage [201], Y chromosome STR analysis for the identification of male DNA in a contaminated sample [202], animal and plant DNA analysis to identify important clues in poisoning cases [203], ancestry tracing [204], predicting phenotypes based on the genes [205], epigenetic analysis to identify the age of the donor DNA [206], and microRNA analysis for identifying body fluids and post-mortem interval [207]. The application of NGS in biodefense and bioterrorism involving the detection of microbial signatures at crime sites is another discipline gaining rapid attraction [208,209]. The major providers of NGS technology dominating the forensic domain are Illumina's MiSeq FGx, Thermo Fisher's Ion Torrent PGM, and Ion S5 [210,211]

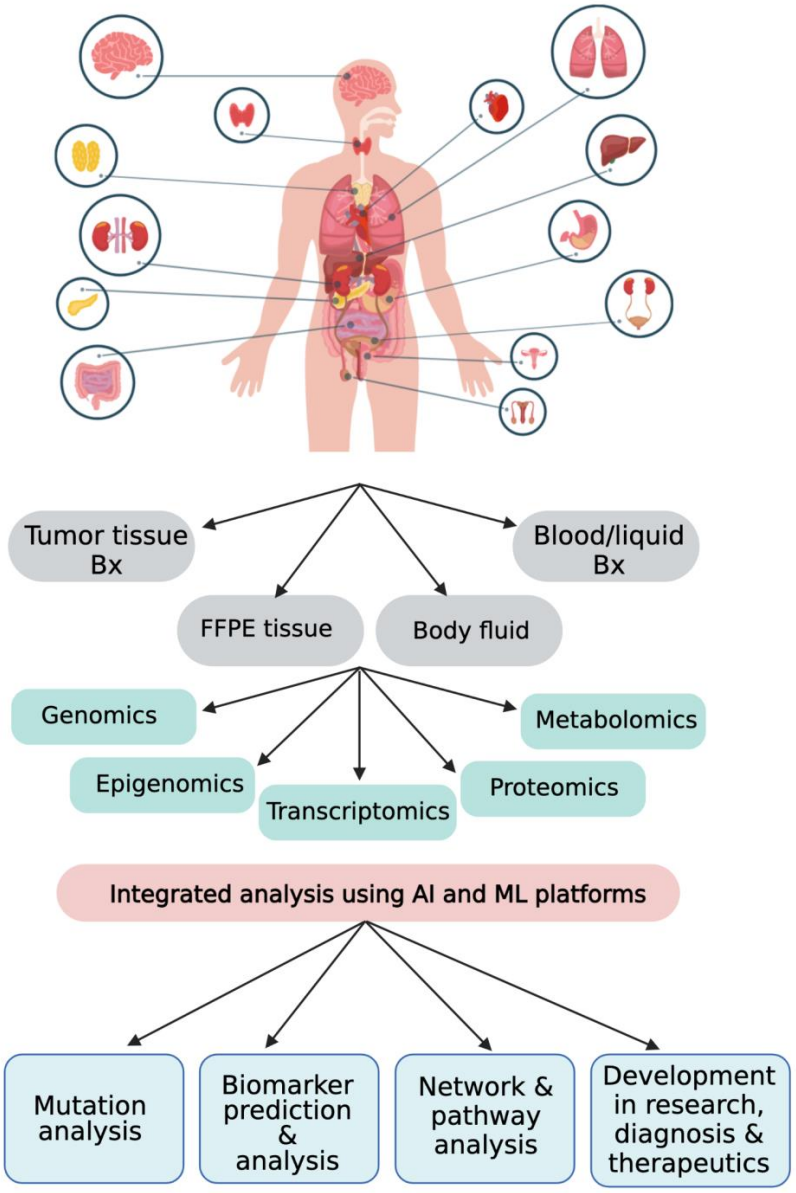

Figure 4. Role of NGS technology in cancer diagnosis, prognosis, and therapeutics using an integrative omics approach. FFPE, formalin-fixed paraffin-embedded; Bx, biopsy; AI, artificial intelligence; Ml, machine learning.

6. Future Prospects and Conclusions

The future scope of NGS holds tremendous potential for advancements and applications in various fields. The progress in bioinformatics, robotics, liquid handling, and nucleic acid preparation will revolutionize NGS sequencing methods, making them faster and more precise. These forthcoming sequencing platforms will necessitate smaller amounts of input DNA and reagents, scaling down to zeptoliters and even a few molecules. Additionally, they will become increasingly portable, enabling their utilization in diagnostic applications across various fields such as medical, agricultural, ecological, and other field-based settings. Taken together, NGS holds immense potential for transformative advancements across multiple domains. NGS has already revolutionized fields such as clinical diagnostics, cancer genomics, and microbial genomics, providing unprecedented insights into the genetic underpinnings of diseases and driving personalized medicine. As technology progresses, NGS is expected to play a pivotal role in areas such as single-cell genomics, long-read sequencing, epigenomics, and multi-omics integration, enabling a deeper understanding of cellular processes, disease mechanisms, and personalized treatment strategies. The development of real-time sequencing and point-of-care applications will further extend the reach of NGS, empowering rapid diagnostics and monitoring in various settings. Additionally, advancements in bioinformatics and data analysis will be crucial for extracting meaningful insights from the vast amount of NGS data generated. The higher order multiplexing will enable more samples to be processed in a shorter time and at a reduced cost supported by the advances in robotics, liquid handling, and sample processing will contribute to these advancements. Equally important will be advanced in faster and more accurate bioinformatic data analysis, as well as data transfer and storage. With ongoing technological improvements and cost reduction, NGS will become more accessible and widespread, facilitating its integration into routine clinical practice, research, agriculture, and environmental studies. The future of NGS is promising, promising to unlock new frontiers of knowledge and catalyze advancements that will have a profound impact on human health, agriculture, environmental conservation, and beyond.

Author Contributions: Conceptualization: H.S., G.D. and S.K.M.; original draft preparation: H.S., K.J., G.D. and S.K.M.; literature search, analysis, writing, review, and editing: H.S., K.J., U.M., S.W., G.Z., S.R., R.P.T., S.B., A.K.M., G.D. and S.K.M. visualization: H.S., S.W. and R.P.T. Supervision: A.K.M., G.D. and S.K.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors express their heartfelt gratitude and tribute to the late Professor Michael Green from the Department of Molecular Cell and Cancer Biology, UMass Chan Medical School, for his invaluable support and remarkable contributions to the field of molecular genetics and cancer genomics.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Goodwin, S.; McPherson, J.D.; McCombie, W.R. Coming of age: Ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **2016**, *17*, 333–351. [CrossRef] [PubMed]
2. Levy, S.E.; Myers, R.M. Advancements in Next-Generation Sequencing. *Annu. Rev. Genom. Hum. Genet.* **2016**, *17*, 95–115. [CrossRef] [PubMed]
3. Rhoads, A.; Au, K.F. PacBio Sequencing and Its Applications. *Genom. Proteom. Bioinform.* **2015**, *13*, 278–289. [CrossRef] [PubMed]
4. Vaser, R.; Sović, I.; Nagarajan, N.; Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **2017**, *27*, 737–746. [CrossRef]
5. Amarasinghe, S.L.; Su, S.; Dong, X.; Zappia, L.; Ritchie, M.E.; Gouil, Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* **2020**, *21*, 30. [CrossRef]

6. Metzker, M.L. Emerging technologies in DNA sequencing. *Genome Res.* **2005**, *15*, 1767–1776. [CrossRef]
7. Kumar, K.R.; Cowley, M.J.; Davis, R.L. Next-Generation Sequencing and Emerging Technologies. *Semin. Thromb. Hemost.* **2019**, *45*, 661–673. [CrossRef]
8. Sakamoto, Y.; Sereewattanawoot, S.; Suzuki, A. A new era of long-read sequencing for cancer genomics. *J. Hum. Genet.* **2020**, *65*, 3–10. [CrossRef]
9. Goto, Y.; Akahori, R.; Yanagi, I.; Takeda, K.-I. Solid-state nanopores towards single-molecule DNA sequencing. *J. Hum. Genet.* **2020**, *65*, 69–77. [CrossRef]
10. Salk, J.J.; Schmitt, M.W.; Loeb, L.A. Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. *Nat. Rev. Genet.* **2018**, *19*, 269–285. [CrossRef]
11. Holley, R.W.; Apgar, J.; Everett, G.A.; Madison, J.T.; Marquisee, M.; Merrill, S.H.; Penswick, J.R.; Zamir, A. Structure of a Ribonucleic Acid. *Science* **1965**, *147*, 1462–1465. [CrossRef]
12. Heather, J.M.; Chain, B. The sequence of sequencers: The history of sequencing DNA. *Genomics* **2016**, *107*, 1–8. [CrossRef]
13. Sanger, F.; Nicklen, S.; Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5463–5467. [CrossRef]
14. Barba, M.; Czosnek, H.; Hadidi, A. Historical Perspective, Development and Applications of Next-Generation Sequencing in Plant Virology. *Viruses* **2013**, *6*, 106–136. [CrossRef]
15. Schuster, S.C. Next-generation sequencing transforms today's biology. *Nat. Methods* **2008**, *5*, 16–18. [CrossRef]
16. Hutchison, C.A. DNA sequencing: Bench to bedside and beyond. *Nucleic Acids Res.* **2007**, *35*, 6227–6237. [CrossRef]
17. Pervez, M.T.; Hasnain, M.J.U.; Abbas, S.H.; Moustafa, M.F.; Aslam, N.; Shah, S.S.M. A Comprehensive Review of Performance of Next-Generation Sequencing Platforms. *BioMed Res. Int.* **2022**. [CrossRef]
18. Ronaghi, M.; Karamohamed, S.; Pettersson, B.; Uhlén, M.; Nyrén, P. Real-Time DNA Sequencing Using Detection of Pyrophosphate Release. *Anal. Biochem.* **1996**, *242*, 84–89. [CrossRef]
19. Slatko, B.E.; Gardner, A.F.; Ausubel, F.M. Overview of Next-Generation Sequencing Technologies. *Curr. Protoc. Mol. Biol.* **2018**, *122*, e59. [CrossRef]
20. Henson, J.; Tischler, G.; Ning, Z. Next-generation sequencing and large genome assemblies. *Pharmacogenomics* **2012**, *13*, 901–915. [CrossRef]
21. Rothberg, J.M.; Hinz, W.; Rearick, T.M.; Schultz, J.; Mileski, W.; Davey, M.; Leamon, J.H.; Johnson, K.; Milgrew, M.J.; Edwards, M.; et al. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **2011**, *475*, 348–352. [CrossRef]
22. Buermans, H.P.J.; Den Dunnen, J.T. Next generation sequencing technology: Advances and applications. *Biochim. Biophys. Acta (BBA)—Mol. Basis Dis.* **2014**, *1842*, 1932–1941. [CrossRef]
23. Shendure, J.; Porreca, G.J.; Reppas, N.B.; Lin, X.; McCutcheon, J.P.; Rosenbaum, A.M.; Wang, M.D.; Zhang, K.; Mitra, R.D.; Church, G.M. Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome. *Science* **2005**, *309*, 1728–1732. [CrossRef] [PubMed]
24. Drmanac, R.; Sparks, A.B.; Callow, M.J.; Halpern, A.L.; Burns, N.L.; Kermani, B.G.; Carnevali, P.; Nazarenko, I.; Nilsen, G.B.; Yeung, G.; et al. Human Genome Sequencing Using Unchained Base Reads on Self-Assembling DNA Nanoarrays. *Science* **2010**, *327*, 78–81. [CrossRef] [PubMed]
25. Xu, Y.; Lin, Z.; Tang, C.; Tang, Y.; Cai, Y.; Zhong, H.; Wang, X.; Zhang, W.; Xu, C.; Wang, J.; et al. A new massively parallel nanoball sequencing platform for whole exome research. *BMC Bioinform.* **2019**, *20*, 153. [CrossRef] [PubMed]
26. Hart, C.; Lipson, D.; Ozsolak, F.; Raz, T.; Steinmann, K.; Thompson, J.; Milos, P.M. Single-Molecule Sequencing. *Methods Enzymol.* **2010**, *472*, 407–430. [CrossRef]
27. Thompson, J.F.; Steinmann, K.E. Single Molecule Sequencing with a HeliScope Genetic Analysis System. *Curr. Protoc. Mol. Biol.* **2010**, *92*, 7.10.1–7.10.14. [CrossRef]
28. Eid, J.; Fehr, A.; Gray, J.; Luong, K.; Lyle, J.; Otto, G.; Peluso, P.; Rank, D.; Baybayan, P.; Bettman, B.; et al. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* **2009**, *323*, 133–138. [CrossRef]
29. Roberts, R.J.; Carneiro, M.O.; Schatz, M.C. The advantages of SMRT sequencing. *Genome Biol.* **2013**, *14*, 405. [CrossRef]
30. Jain, M.; Olsen, H.E.; Paten, B.; Akeson, M. The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community. *Genome Biol.* **2016**, *17*, 239. [CrossRef]
31. Mantere, T.; Kersten, S.; Hoischen, A. Long-Read Sequencing Emerging in Medical Genetics. *Front. Genet.* **2019**, *10*, 426. [CrossRef]
32. Costain, G.; Cohn, R.D.; Scherer, S.W.; Marshall, C.R. Genome sequencing as a diagnostic test. *Can. Med. Assoc. J.* **2021**, *193*, E1626–E1629. [CrossRef]
33. Logsdon, G.A.; Vollger, M.R.; Eichler, E.E. Long-read human genome sequencing and its applications. *Nat. Rev. Genet.* **2020**, *21*, 597–614. [CrossRef]
34. Rabbani, B.; Tekin, M.; Mahdieh, N. The promise of whole-exome sequencing in medical genetics. *J. Hum. Genet.* **2014**, *59*, 5–15. [CrossRef]
35. Iglesias, A.; Anyane-Yeboa, K.; Wynn, J.; Wilson, A.; Cho, M.T.; Guzman, E.; Sisson, R.; Egan, C.; Chung, W.K. The usefulness of whole-exome sequencing in routine clinical practice. *Anesth. Analg.* **2014**, *16*, 922–931. [CrossRef]
36. Van Dijk, E.L.; Auger, H.; Jaszczyszyn, Y.; Thermes, C. Ten years of next-generation sequencing technology. *Trends Genet.* **2014**, *30*, 418–426. [CrossRef]
37. Warr, A.; Robert, C.; Hume, D.; Archibald, A.; Deeb, N.; Watson, M. Exome Sequencing: Current and Future Perspectives. *G3 Genes Genom. Genet.* **2015**, *5*, 1543–1550. [CrossRef]

38. Williams, M.J.; Sottoriva, A.; Graham, T.A. Measuring Clonal Evolution in Cancer with Genomics. *Annu. Rev. Genom. Hum. Genet.* **2019**, *20*, 309–329. [CrossRef]
39. Kim, M. Targeted Panels or Exome—Which Is the Right NGS Approach for Inherited Disease Research? 2017. Available online: https://admin.acceleratingscience.com/behindthebench/targeted-panels-or-exome-which-is-the-right-ngs-approach-for-inherited-disease-research/ (accessed on 10 June 2023).
40. Li, J.; Liu, C. Coding or Noncoding, the Converging Concepts of RNAs. *Front. Genet.* **2019**, *10*, 496. [CrossRef]
41. Lucchinetti, E.; Zaugg, M. RNA Sequencing. *Anesthesiology* **2020**, *133*, 976–978. [CrossRef]
42. Choi, S.-W.; Kim, H.-W.; Nam, J.-W. The small peptide world in long noncoding RNAs. *Brief. Bioinform.* **2019**, *20*, 1853–1864. [CrossRef]
43. Lasda, E.; Parker, R. Circular RNAs: Diversity of form and function. *RNA* **2014**, *20*, 1829–1842. [CrossRef]
44. Chen, J.-W.; Shrestha, L.; Green, G.; Leier, A.; Marquez-Lago, T.T. The hitchhikers' guide to RNA sequencing and functional analysis. *Brief. Bioinform.* **2023**, *24*, bbac529. [CrossRef] [PubMed]
45. Stark, R.; Grzelak, M.; Hadfield, J. RNA sequencing: The teenage years. *Nat. Rev. Genet.* **2019**, *20*, 631–656. [CrossRef] [PubMed]
46. Ura, H.; Togi, S.; Niida, Y. A comparison of mRNA sequencing (RNA-Seq) library preparation methods for transcriptome analysis. *BMC Genom.* **2022**, *23*, 303. [CrossRef] [PubMed]
47. Kolanowska, M.; Kubiak, A.; Jażdżewski, K.; Wójcicka, A. MicroRNA Analysis Using Next-Generation Sequencing. *Methods Mol. Biol.* **2018**, *1823*, 87–101. [CrossRef]
48. Grillone, K.; Riillo, C.; Scionti, F.; Rocca, R.; Tradigo, G.; Guzzi, P.H.; Alcaro, S.; Di Martino, M.T.; Tagliaferri, P.; Tassone, P. Non-coding RNAs in cancer: Platforms and strategies for investigating the genomic "dark matter". *J. Exp. Clin. Cancer Res.* **2020**, *39*, 117. [CrossRef]
49. Atkinson, S.R.; Marguerat, S.; Bähler, J. Exploring long non-coding RNAs through sequencing. *Semin. Cell Dev. Biol.* **2012**, *23*, 200–205. [CrossRef]
50. Wang, Z.; Gerstein, M.; Snyder, M. RNA-Seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **2009**, *10*, 57–63. [CrossRef]
51. Benesova, S.; Kubista, M.; Valihrach, L. Small RNA-Sequencing: Approaches and Considerations for miRNA Analysis. *Diagnostics* **2021**, *11*, 964. [CrossRef]
52. Cao, J. The functional role of long non-coding RNAs and epigenetics. *Biol. Proced. Online* **2014**, *16*, 42. [CrossRef]
53. Kumar, S.; Gonzalez, E.A.; Rameshwar, P.; Etchegaray, J.-P. Non-Coding RNAs as Mediators of Epigenetic Changes in Malignancies. *Cancers* **2020**, *12*, 3657. [CrossRef]
54. Mozdarani, H.; Ezzatizadeh, V.; Parvaneh, R.R. The emerging role of the long non-coding RNA HOTAIR in breast cancer development and treatment. *J. Transl. Med.* **2020**, *18*, 152. [CrossRef]
55. Raghavan, V.; Kraft, L.; Mesny, F.; Rigerte, L. A simple guide to *de novo* transcriptome assembly and annotation. *Brief. Bioinform.* **2022**, *23*, bbab563. [CrossRef]
56. Kulkarni, A.; Anderson, A.G.; Merullo, D.P.; Konopka, G. Beyond bulk: A review of single cell transcriptomics methodologies and applications. *Curr. Opin. Biotechnol.* **2019**, *58*, 129–136. [CrossRef]
57. Adil, A.; Kumar, V.; Jan, A.T.; Asger, M. Single-Cell Transcriptomics: Current Methods and Challenges in Data Acquisition and Analysis. *Front. Neurosci.* **2021**, *15*, 591122. [CrossRef]
58. Wang, J.; Tian, T.; Li, X.; Zhang, Y. Noncoding RNAs Emerging as Drugs or Drug Targets: Their Chemical Modification, Bio-Conjugation and Intracellular Regulation. *Molecules* **2022**, *27*, 6717. [CrossRef]
59. López-Camarillo, C.; Gallardo-Rincón, D.; Álvarez-Sánchez, M.E.; Marchat, L.A. Pharmaco-epigenomics: On the Road of Translation Medicine. In *Translational Research and Onco-Omics Applications in the Era of Cancer Personal Genomics*; Springer: Berlin/Heidelberg, Germany, 2019; Volume 1168, pp. 31–42. [CrossRef]
60. National Human Genoe Research Institute. Epigenomics Fact Sheet. 2020. Available online: https://www.genome.gov/about-genomics/fact-sheets/Epigenomics-Fact-Sheet (accessed on 10 June 2023).
61. Handy, D.E.; Castro, R.; Loscalzo, J. Epigenetic Modifications. *Circulation* **2011**, *123*, 2145–2156. [CrossRef] [PubMed]
62. Fuso, A. Aging and Disease. In *Epigenetics in Human Disease*; Academic Press: Cambridge, MA, USA, 2018; pp. 935–973. [CrossRef]
63. Metere, A.; Graves, C.E. Factors Influencing Epigenetic Mechanisms: Is There a Role for Bariatric Surgery? *Biotech* **2020**, *9*, 6. [CrossRef]
64. Heyn, H.; Esteller, M. DNA methylation profiling in the clinic: Applications and challenges. *Nat. Rev. Genet.* **2012**, *13*, 679–692. [CrossRef]
65. Zhu, H.; Zhu, H.; Tian, M.; Wang, D.; He, J.; Xu, T. DNA Methylation and Hydroxymethylation in Cervical Cancer: Diagnosis, Prognosis and Treatment. *Front. Genet.* **2020**, *11*, 347. [CrossRef] [PubMed]
66. Sarda, S.; Hannenhalli, S. Next-Generation Sequencing and Epigenomics Research: A Hammer in Search of Nails. *Genom. Inform.* **2014**, *12*, 2–11. [CrossRef] [PubMed]
67. Barros-Silva, D.; Marques, C.J.; Henrique, R.; Jerónimo, C. Profiling DNA Methylation Based on Next-Generation Sequencing Approaches: New Insights and Clinical Applications. *Genes* **2018**, *9*, 429. [CrossRef]
68. Wreczycka, K.; Gosdschan, A.; Yusuf, D.; Grüning, B.; Assenov, Y.; Akalin, A. Strategies for analyzing bisulfite sequencing data. *J. Biotechnol.* **2017**, *261*, 105–115. [CrossRef]

69. Frommer, M.; E McDonald, L.; Millar, D.S.; Collis, C.M.; Watt, F.; Grigg, G.W.; Molloy, P.L.; Paul, C.L. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 1827–1831. [CrossRef]
70. Lu, R.J.-H.; Liu, Y.-T.; Huang, C.W.; Yen, M.-R.; Lin, C.-Y.; Chen, P.-Y. ATACgraph: Profiling Genome-Wide Chromatin Accessibility from ATAC-seq. *Front. Genet.* **2021**, *11*, 618478. [CrossRef]
71. Mansisidor, A.R.; Risca, V.I. Chromatin accessibility: Methods, mechanisms, and biological insights. *Nucleus* **2022**, *13*, 238–278. [CrossRef]
72. Liu, E.T.; Pott, S.; Huss, M. Q&A: ChIP-seq technologies and the study of gene regulation. *BMC Biol.* **2010**, *8*, 56. [CrossRef]
73. Furey, T.S. ChIP—Seq and beyond: New and improved methodologies to detect and characterize protein—DNA interactions. *Nat. Rev. Genet.* **2012**, *13*, 840–852. [CrossRef]
74. O'geen, H.; Echipare, L.; Farnham, P.J. Using ChIP-Seq Technology to Generate High-Resolution Profiles of Histone Modifications. *Methods Mol. Biol.* **2011**, *791*, 265–286. [CrossRef]
75. Nakato, R.; Sakata, T. Methods for ChIP-seq analysis: A practical workflow and advanced applications. *Methods* **2021**, *187*, 44–53. [CrossRef] [PubMed]
76. Feng, F.; Yao, Y.; Wang, X.Q.D.; Zhang, X.; Liu, J. Connecting high-resolution 3D chromatin organization with epigenomics. *Nat. Commun.* **2022**, *13*, 2054. [CrossRef]
77. Tang, B.; Cheng, X.; Xi, Y.; Chen, Z.; Zhou, Y.; Jin, V.X. Advances in Genomic Profiling and Analysis of 3D Chromatin Structure and Interaction. *Genes* **2017**, *8*, 223. [CrossRef]
78. Thomas, T.; Gilbert, J.; Meyer, F. Metagenomics—A guide from sampling to data analysis. *Microb. Inform. Exp.* **2012**, *2*, 3. [CrossRef]
79. Bellemain, E.; Carlsen, T.; Brochmann, C.; Coissac, E.; Taberlet, P.; Kauserud, H. ITS as an environmental DNA barcode for fungi: An in silico approach reveals potential PCR biases. *BMC Microbiol.* **2010**, *10*, 189. [CrossRef]
80. Perlejewski, K.; Bukowska-Ośko, I.; Rydzanicz, M.; Pawełczyk, A.; Cortès, K.C.; Osuch, S.; Paciorek, M.; Dzieciątkowski, T.; Radkowski, M.; Laskus, T. Next-generation sequencing in the diagnosis of viral encephalitis: Sensitivity and clinical limitations. *Sci. Rep.* **2020**, *10*, 16173. [CrossRef]
81. Cao, Y.; Fanning, S.; Proos, S.; Jordan, K.; Srikumar, S. A Review on the Applications of Next Generation Sequencing Technologies as Applied to Food-Related Microbiome Studies. *Front. Microbiol.* **2017**, *8*, 1829. [CrossRef]
82. Caporaso, J.G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F.D.; Costello, E.K.; Fierer, N.; Gonzalez Peña, A.; Goodrich, J.K.; Gordon, J.I.; et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **2010**, *7*, 335–336. [CrossRef]
83. Pruitt, K.D. NCBI Reference Sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **2004**, *33*, D501–D504. [CrossRef]
84. Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [CrossRef]
85. Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **2004**, *32*, D258–D261. [CrossRef] [PubMed]
86. Nurk, S.; Meleshko, D.; Korobeynikov, A.; Pevzner, P.A. metaSPAdes: A new versatile metagenomic assembler. *Genome Res.* **2017**, *27*, 824–834. [CrossRef] [PubMed]
87. Peng, Y.; Leung, H.C.M.; Yiu, S.M.; Chin, F.Y.L. Meta-IDBA: A de novo assembler for metagenomic data. *Bioinformatics* **2011**, *27*, i94–i101. [CrossRef] [PubMed]
88. Seemann, T. Prokka: Rapid Prokaryotic Genome Annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [CrossRef] [PubMed]
89. Zhu, W.; Lomsadze, A.; Borodovsky, M. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* **2010**, *38*, e132. [CrossRef] [PubMed]
90. 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).
91. Iyer, R.; Stepanov, V.G.; Iken, B. Isolation and molecular characterization of a novel pseudomonas putida strain capable of degrading organophosphate and aromatic compounds. *Adv. Biol. Chem.* **2013**, *3*, 564–578. [CrossRef]
92. Ewels, P.; Magnusson, M.; Lundin, S.; Käller, M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **2016**, *32*, 3047–3048. [CrossRef]
93. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]
94. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **2011**, *17*, 10–12. [CrossRef]
95. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [CrossRef]
96. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **2013**, arXiv:1303.3997.
97. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef]
98. Caetano-Anolles, D. Functional Equivalence in DRAGEN-GATK. 2022. Available online: https://gatk.broadinstitute.org/hc/en-us/articles/4410456501915 (accessed on 6 July 2023).
99. Broadinstitute. Picard, GitHub. (n.d.). Available online: http://broadinstitute.github.io/picard/ (accessed on 1 July 2023).
100. Tarasov, A.; Vilella, A.J.; Cuppen, E.; Nijman, I.J.; Prins, P. Sambamba: Fast processing of NGS alignment formats. *Bioinformatics* **2015**, *31*, 2032–2034. [CrossRef]

101. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [CrossRef]
102. Garrison, E.; Marth, G. Haplotype-Based Variant Detection from Short-Read Sequencing. *arXiv* **2012**, arXiv:1207.3907.
103. Rimmer, A.; Phan, H.; Mathieson, I.; Iqbal, Z.; Twigg, S.R.F.; Wilkie, A.O.M.; McVean, G.; Lunter, G.; WGS500 Consortium. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat. Genet.* **2014**, *46*, 912–918. [CrossRef]
104. Koboldt, D.C.; Chen, K.; Wylie, T.; Larson, D.E.; McLellan, M.D.; Mardis, E.R.; Weinstock, G.M.; Wilson, R.K.; Ding, L. VarScan: Variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* **2009**, *25*, 2283–2285. [CrossRef]
105. Poplin, R.; Chang, P.-C.; Alexander, D.; Schwartz, S.; Colthurst, T.; Ku, A.; Newburger, D.; Dijamco, J.; Nguyen, N.; Afshar, P.T.; et al. A universal SNP and small-indel variant caller using deep neural networks. *Nat. Biotechnol.* **2018**, *36*, 983–987. [CrossRef]
106. Illumina. DRAGEN Bio-IT Platform, (n.d.). Available online: https://Www.Illumina.Com/Products/by-Type/Informatics-Products/Dragen-Bio-It-Platform.Html (accessed on 15 June 2023).
107. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **2011**, *27*, 2987–2993. [CrossRef]
108. Wang, K.; Li, M.; Hakonarson, H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **2010**, *38*, e164. [CrossRef]
109. McLaren, W.; Gil, L.; Hunt, S.E.; Riat, H.S.; Ritchie, G.R.S.; Thormann, A.; Flicek, P.; Cunningham, F. The Ensembl Variant Effect Predictor. *Genome Biol.* **2016**, *17*, 122. [CrossRef] [PubMed]
110. Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **2012**, *6*, 80–92. [CrossRef] [PubMed]
111. Illumina Inc. Nirvana: Clinical-Grade Variant Annotations. 2023. Available online: https://Illumina.Github.Io/NirvanaDocumentation/ (accessed on 15 June 2023).
112. Rausch, T.; Zichner, T.; Schlattl, A.; Stütz, A.M.; Benes, V.; Korbel, J.O. DELLY: Structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **2012**, *28*, i333–i339. [CrossRef]
113. Layer, R.M.; Chiang, C.; Quinlan, A.R.; Hall, I.M. LUMPY: A probabilistic framework for structural variant discovery. *Genome Biol.* **2014**, *15*, R84. [CrossRef] [PubMed]
114. Chen, X.; Schulz-Trieglaff, O.; Shaw, R.; Barnes, B.; Schlesinger, F.; Källberg, M.; Cox, A.J.; Kruglyak, S.; Saunders, C.T. Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **2016**, *32*, 1220–1222. [CrossRef]
115. Cameron, D.L.; Schröder, J.; Penington, J.S.; Do, H.; Molania, R.; Dobrovic, A.; Speed, T.P.; Papenfuss, A.T. GRIDSS: Sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. *Genome Res.* **2017**, *27*, 2050–2060. [CrossRef]
116. Kronenberg, Z.; Osborne, E.J.; Cone, K.R.; Kennedy, B.J.; Domyan, E.T.; Shapiro, M.D.; Elde, N.C.; Yandell, M. Wham: Identifying Structural Variants of Biological Consequence. *PLoS Comput. Biol.* **2015**, *11*, e1004572. [CrossRef]
117. Ye, K.; Schulz, M.H.; Long, Q.; Apweiler, R.; Ning, Z. Pindel: A pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **2009**, *25*, 2865–2871. [CrossRef]
118. Abyzov, A.; Urban, A.E.; Snyder, M.; Gerstein, M. CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* **2011**, *21*, 974–984. [CrossRef]
119. Babadi, M.; Lee, S.K.; Smirnov, A.; Lichtenstein, L.; Gauthier, L.D.; Howrigan, D.P.; Poterba, T. Abstract 2287: Precise common and rare germline CNV calling with GATK. *Cancer Res.* **2018**, *78*, 2287. [CrossRef]
120. Klambauer, G.; Schwarzbauer, K.; Mayr, A.; Clevert, D.-A.; Mitterecker, A.; Bodenhofer, U.; Hochreiter, S. cn.MOPS: Mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. *Nucleic Acids Res.* **2012**, *40*, e69. [CrossRef]
121. Bellos, E.; Kumar, V.; Lin, C.; Maggi, J.; Phua, Z.Y.; Cheng, C.-Y.; Cheung, C.M.G.; Hibberd, M.L.; Wong, T.Y.; Coin, L.J.M.; et al. cnvCapSeq: Detecting copy number variation in long-range targeted resequencing data. *Nucleic Acids Res.* **2014**, *42*, e158. [CrossRef]
122. Plagnol, V.; Curtis, J.; Epstein, M.; Mok, K.Y.; Stebbings, E.; Grigoriadou, S.; Wood, N.W.; Hambleton, S.; Burns, S.O.; Thrasher, A.J.; et al. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics* **2012**, *28*, 2747–2754. [CrossRef]
123. Kim, D.; Pertea, G.; Trapnell, C.; Pimentel, H.; Kelley, R.; Salzberg, S.L. TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **2013**, *14*, R36. [CrossRef]
124. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **2019**, *37*, 907–915. [CrossRef]
125. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21. [CrossRef]
126. Liao, Y.; Smyth, G.K.; Shi, W. feature Counts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2014**, *30*, 923–930. [CrossRef]
127. Anders, S.; Pyl, P.T.; Huber, W. HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics* **2015**, *31*, 166–169. [CrossRef]

128. Patro, R.; Duggal, G.; Love, M.I.; Irizarry, R.A.; Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **2017**, *14*, 417–419. [CrossRef]
129. Bray, N.L.; Pimentel, H.; Melsted, P.; Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **2016**, *34*, 525–527. [CrossRef]
130. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [CrossRef]
131. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. EdgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, *26*, 139–140. [CrossRef] [PubMed]
132. Dennis, G., Jr.; Sherman, B.T.; A Hosack, D.; Yang, J.; Gao, W.; Lane, H.C.; A Lempicki, R. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **2003**, *4*, R60. [CrossRef]
133. Wu, T.; Hu, E.; Xu, S.; Chen, M.; Guo, P.; Dai, Z.; Feng, T.; Zhou, L.; Tang, W.; Zhan, L.; et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2021**, *2*, 100141. [CrossRef] [PubMed]
134. Chen, E.Y.; Tan, C.M.; Kou, Y.; Duan, Q.; Wang, Z.; Meirelles, G.V.; Clark, N.R.; Ma'Ayan, A. Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinform.* **2013**, *14*, 128. [CrossRef]
135. Pedersen, B.S.; Eyring, K.; De, S.; Yang, I.V.; Schwartz, D.A. Fast and accurate alignment of long bisulfite-seq reads. *arXiv* **2014**, arXiv:1401.1129. [CrossRef]
136. Guo, W.; Fiziev, P.; Yan, W.; Cokus, S.; Sun, X.; Zhang, M.Q.; Chen, P.-Y.; Pellegrini, M. BS-Seeker2: A versatile aligning pipeline for bisulfite sequencing data. *BMC Genom.* **2013**, *14*, 774. [CrossRef]
137. Krueger, F.; Andrews, S.R. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **2011**, *27*, 1571–1572. [CrossRef]
138. Jühling, F.; Kretzmer, H.; Bernhart, S.H.; Otto, C.; Stadler, P.F.; Hoffmann, S. metilene: Fast and sensitive calling of differentially methylated regions from bisulfite sequencing data. *Genome Res.* **2016**, *26*, 256–262. [CrossRef]
139. Hansen, K.D.; Langmead, B.; Irizarry, R.A. BSmooth: From whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol.* **2012**, *13*, R83. [CrossRef]
140. Akalin, A.; Kormaksson, M.; Li, S.; E Garrett-Bakelman, F.; E Figueroa, M.; Melnick, A.; E Mason, C. methylKit: A comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **2012**, *13*, R87. [CrossRef]
141. Zhang, Y.; Liu, T.; Meyer, C.A.; Eeckhoute, J.; Johnson, D.S.; Bernstein, B.E.; Nusbaum, C.; Myers, R.M.; Brown, M.; Li, W.; et al. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **2008**, *9*, R137. [CrossRef]
142. Xu, S.; Grullon, S.; Ge, K.; Peng, W. Spatial Clustering for Identification of ChIP-Enriched Regions (SICER) to Map Regions of Histone Methylation Patterns in Embryonic Stem Cells. *Methods Mol. Biol.* **2014**, *1150*, 97–111. [CrossRef]
143. Ochsner, S.A.; Abraham, D.; Martin, K.; Ding, W.; McOwiti, A.; Kankanamge, W.; Wang, Z.; Andreano, K.; Hamilton, R.A.; Chen, Y.; et al. The Signaling Pathways Project, an integrated 'omics knowledgebase for mammalian cellular signaling pathways. *Sci. Data* **2019**, *6*, 252. [CrossRef]
144. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842. [CrossRef]
145. Carroll, T.S.; Eliang, Z.; Salama, R.; Stark, R.; Santiago, I.E. Impact of artifact removal on ChIP quality metrics in ChIP-seq and ChIP-exo data. *Front. Genet.* **2014**, *5*, 75. [CrossRef]
146. Landt, S.G.; Marinov, G.K.; Kundaje, A.; Kheradpour, P.; Pauli, F.; Batzoglou, S.; Bernstein, B.E.; Bickel, P.; Brown, J.B.; Cayting, P.; et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **2012**, *22*, 1813–1831. [CrossRef]
147. Stark, R.; Brown, G. DiffBind: Differential Binding Analysis of ChIP-Seq Peak Data. 2012. Available online: http://bioconductor.org/packages/release/bioc/html/DiffBind.html (accessed on 5 June 2023).
148. Shao, Z.; Zhang, Y.; Yuan, G.-C.; Orkin, S.H.; Waxman, D.J. MAnorm: A robust model for quantitative comparison of ChIP-Seq data sets. *Genome Biol.* **2012**, *13*, R16. [CrossRef] [PubMed]
149. Schweikert, G.; Cseke, B.; Clouaire, T.; Bird, A.; Sanguinetti, G. MMDiff: Quantitative testing for shape changes in ChIP-Seq data sets. *BMC Genom.* **2013**, *14*, 826. [CrossRef]
150. Bailey, T.L.; Johnson, J.; Grant, C.E.; Noble, W.S. The MEME Suite. *Nucleic Acids Res.* **2015**, *43*, W39–W49. [CrossRef]
151. Heinz, S.; Benner, C.; Spann, N.; Bertolino, E.; Lin, Y.C.; Laslo, P.; Cheng, J.X.; Murre, C.; Singh, H.; Glass, C.K. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* **2010**, *38*, 576–589. [CrossRef] [PubMed]
152. Rivera, A.M.; Defrance, M.; Sand, O.; Herrmann, C.; Castro-Mondragon, J.A.; Delerce, J.; Jaeger, S.; Blanchet, C.; Vincens, P.; Caron, C.; et al. RSAT 2015: Regulatory Sequence Analysis Tools. *Nucleic Acids Res.* **2015**, *43*, W50–W56. [CrossRef] [PubMed]
153. Schloss, P.D.; Westcott, S.L.; Ryabin, T.; Hall, J.R.; Hartmann, M.; Hollister, E.B.; Lesniewski, R.A.; Oakley, B.B.; Parks, D.H.; Robinson, C.J.; et al. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl. Environ. Microbiol.* **2009**, *75*, 7537–7541. [CrossRef]
154. Edgar, R.C. UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **2013**, *10*, 996–998. [CrossRef]

155. McDonald, D.; Price, M.N.; Goodrich, J.; Nawrocki, E.P.; DeSantis, T.Z.; Probst, A.; Andersen, G.L.; Knight, R.; Hugenholtz, P. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* **2012**, *6*, 610–618. [CrossRef]
156. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glöckner, F.O. The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools. *Nucleic Acids Res.* **2012**, *41*, D590–D596. [CrossRef]
157. Cole, J.R.; Wang, Q.; Fish, J.A.; Chai, B.; McGarrell, D.M.; Sun, Y.; Brown, C.T.; Porras-Alfaro, A.; Kuske, C.R.; Tiedje, J.M. Ribosomal Database Project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* **2014**, *42*, D633–D642. [CrossRef]
158. Blanco-Míguez, A.; Beghini, F.; Cumbo, F.; McIver, L.J.; Thompson, K.N.; Zolfo, M.; Manghi, P.; Dubois, L.; Huang, K.D.; Thomas, A.M.; et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species with MetaPhlAn 4. *bioRxiv* **2023**. [CrossRef]
159. Menzel, P.; Ng, K.L.; Krogh, A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* **2016**, *7*, 11257. [CrossRef]
160. Wood, D.E.; Salzberg, S.L. Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **2014**, *15*, R46. [CrossRef]
161. Tatusov, R.L.; Koonin, E.V.; Lipman, D.J. A Genomic Perspective on Protein Families. *Science* **1997**, *278*, 631–637. [CrossRef] [PubMed]
162. Jain, A.; Bhoyar, R.C.; Pandhare, K.; Mishra, A.; Sharma, D.; Imran, M.; Senthivel, V.; Divakar, M.K.; Rophina, M.; Jolly, B.; et al. IndiGenomes: A comprehensive resource of genetic variants from over 1000 Indian genomes. *Nucleic Acids Res.* **2021**, *49*, D1225–D1232. [CrossRef] [PubMed]
163. Karczewski, K.J.; Francioli, L.C.; Tiao, G.; Cummings, B.B.; Alfoldi, J.; Wang, Q.; Collins, R.L.; Laricchia, K.M.; Ganna, A.; Birnbaum, D.P.; et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **2020**, *581*, 434–443. [CrossRef]
164. Lu, Y.; Chan, Y.-T.; Tan, H.-Y.; Li, S.; Wang, N.; Feng, Y. Epigenetic regulation in human cancer: The potential role of epi-drug in cancer therapy. *Mol. Cancer* **2020**, *19*, 79. [CrossRef]
165. Qin, J.; Li, R.; Raes, J.; Arumugam, M.; Burgdorf, K.S.; Manichanh, C.; Nielsen, T.; Pons, N.; Levenez, F.; Yamada, T.; et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **2010**, *464*, 59–65. [CrossRef]
166. Foster, J.A.; McVey Neufeld, K.-A. Gut–brain axis: How the microbiome influences anxiety and depression. *Trends Neurosci.* **2013**, *36*, 305–312. [CrossRef]
167. Scher, J.U.; Abramson, S.B. The microbiome and rheumatoid arthritis. *Nat. Rev. Rheumatol.* **2011**, *7*, 569–578. [CrossRef]
168. Devaraj, S.; Hemarajata, P.; Versalovic, J. The Human Gut Microbiome and Body Metabolism: Implications for Obesity and Diabetes. *Clin. Chem.* **2013**, *59*, 617–628. [CrossRef]
169. Di Iulio, J.; Bartha, I.; Spreafico, R.; Virgin, H.W.; Telenti, A. Transfer transcriptomic signatures for infectious diseases. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2022486118. [CrossRef]
170. Pandey, P.R.; Young, K.H.; Kumar, D.; Jain, N. RNA-mediated immunotherapy regulating tumor immune microenvironment: Next wave of cancer therapeutics. *Mol. Cancer* **2022**, *21*, 58. [CrossRef]
171. Hong, M.; Tao, S.; Zhang, L.; Diao, L.-T.; Huang, X.; Huang, S.; Xie, S.-J.; Xiao, Z.-D.; Zhang, H. RNA sequencing: New technologies and applications in cancer research. *J. Hematol. Oncol.* **2020**, *13*, 166. [CrossRef] [PubMed]
172. Chen, G.; Ning, B.; Shi, T. Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. *Front. Genet.* **2019**, *10*, 317. [CrossRef]
173. Leong, A.Z.-X.; Lee, P.Y.; Mohtar, M.A.; Syafruddin, S.E.; Pung, Y.-F.; Low, T.Y. Short open reading frames (sORFs) and microproteins: An update on their identification and validation measures. *J. Biomed. Sci.* **2022**, *29*, 19. [CrossRef]
174. Ormancey, M.; Thuleau, P.; Combier, J.-P.; Plaza, S. The Essentials on microRNA-Encoded Peptides from Plants to Animals. *Biomolecules* **2023**, *13*, 206. [CrossRef]
175. Berdasco, M.; Esteller, M. Clinical epigenetics: Seizing opportunities for translation. *Nat. Rev. Genet.* **2019**, *20*, 109–127. [CrossRef]
176. Singh, R.; Chandel, S.; Dey, D.; Ghosh, A.; Roy, S.; Ravichandiran, V.; Ghosh, D. Epigenetic modification and therapeutic targets of diabetes mellitus. *Biosci. Rep.* **2020**, *40*, BSR20202160. [CrossRef]
177. Miranda Furtado, C.L.; Dos Santos Luciano, M.C.; da Silva Santos, R.; Furtado, G.P.; Moraes, M.O.; Pessoa, C. Epidrugs: Targeting epigenetic marks in cancer treatment. *Epigenetics* **2019**, *14*, 1164–1176. [CrossRef]
178. Huang, W. MicroRNAs: Biomarkers, Diagnostics, and Therapeutics. *Bioinform. MicroRNA Res.* **2017**, *1617*, 57–67. [CrossRef]
179. Arghiani, N.; Matin, M.M. miR-21: A Key Small Molecule with Great Effects in Combination Cancer Therapy. *Nucleic Acid Ther.* **2021**, *31*, 271–283. [CrossRef]
180. Illumina Inc. Ampliseq for Illumina, (n.d.). 2023. Available online: https://sapac.illumina.com/products/by-brand/ampliseq/community-panels.html (accessed on 5 June 2023).
181. Advani, J.; Verma, R.; Chatterjee, O.; Pachouri, P.K.; Upadhyay, P.; Singh, R.; Yadav, J.; Naaz, F.; Ravikumar, R.; Buggi, S.; et al. Whole Genome Sequencing of Mycobacterium tuberculosis Clinical Isolates from India Reveals Genetic Heterogeneity and Region-Specific Variations That Might Affect Drug Susceptibility. *Front. Microbiol.* **2019**, *10*, 309. [CrossRef]

182. Bhoyar, R.C.; Jain, A.; Sehgal, P.; Divakar, M.K.; Sharma, D.; Imran, M.; Jolly, B.; Ranjan, G.; Rophina, M.; Sharma, S.; et al. High throughput detection and genetic epidemiology of SARS-CoV-2 using COVIDSeq next-generation sequencing. *PLoS ONE* **2021**, *16*, e0247115. [CrossRef] [PubMed]
183. Lorenzi, D.; Fernández, C.; Bilinski, M.; Fabbro, M.; Galain, M.; Menazzi, S.; Miguens, M.; Perassi, P.N.; Fulco, M.F.; Kopelman, S.; et al. First custom next-generation sequencing infertility panel in Latin America: Design and first results. *JBRA Assist. Reprod.* **2020**, *24*, 104–114. [CrossRef] [PubMed]
184. Fiorillo, M.T.; Paladini, F.; Tedeschi, V.; Sorrentino, R. HLA Class I or Class II and Disease Association: Catch the Difference If You Can. *Front. Immunol.* **2017**, *8*, 1475. [CrossRef] [PubMed]
185. Maira, D.; Vansan, A.; Maria, A.; Visentainer, J.E.L.; De Souza, C.A. *HLA and Infectious Diseases*; IntechOpen: Rijeka, Croatia, 2014. [CrossRef]
186. Szolek, A.; Schubert, B.; Mohr, C.; Sturm, M.; Feldhahn, M.; Kohlbacher, O. OptiType: Precision HLA typing from next-generation sequencing data. *Bioinformatics* **2014**, *30*, 3310–3316. [CrossRef] [PubMed]
187. Shukla, S.A.; Rooney, M.S.; Rajasagi, M.; Tiao, G.; Dixon, P.M.; Lawrence, M.S.; Stevens, J.; Lane, W.J.; Dellagatta, J.L.; Steelman, S.; et al. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat. Biotechnol.* **2015**, *33*, 1152–1158. [CrossRef] [PubMed]
188. Xie, C.; Yeo, Z.X.; Wong, M.; Piper, J.; Long, T.; Kirkness, E.F.; Biggs, W.H.; Bloom, K.; Spellman, S.; Vierra-Green, C.; et al. Fast and accurate HLA typing from short-read next-generation sequence data with xHLA. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 8059–8064. [CrossRef]
189. Warren, R.L.; Choe, G.; Freeman, D.J.; Castellarin, M.; Munro, S.; Moore, R.; A Holt, R. Derivation of HLA types from shotgun sequence datasets. *Genome Med.* **2012**, *4*, 95. [CrossRef]
190. Robinson, J. IMGT/HLA and IMGT/MHC: Sequence databases for the study of the major histocompatibility complex. *Nucleic Acids Res.* **2003**, *31*, 311–314. [CrossRef]
191. Nagahashi, M.; Shimada, Y.; Ichikawa, H.; Kameyama, H.; Takabe, K.; Okuda, S.; Wakai, T. Next generation sequencing-based gene panel tests for the management of solid tumors. *Cancer Sci.* **2019**, *110*, 6–15. [CrossRef]
192. Abel, H.J.; Duncavage, E.J. Detection of structural DNA variation from next generation sequencing data: A review of informatic approaches. *Cancer Genet.* **2013**, *206*, 432–440. [CrossRef]
193. Aramini, B.; Masciale, V.; Banchelli, F.; D'amico, R.; Dominici, M.; Haider, K.H. Precision Medicine in Lung Cancer: Challenges and Opportunities in Diagnostic and Therapeutic Purposes. In *Lung Cancer*; IntechOpen: Rijeka, Croatia, 2021. [CrossRef]
194. Lee, C.S.; Song, I.H.; Lee, A.; Kang, J.; Lee, Y.S.; Lee, I.K.; Song, Y.S.; Lee, S.H. Enhancing the landscape of colorectal cancer using targeted deep sequencing. *Sci. Rep.* **2021**, *11*, 8154. [CrossRef]
195. Qin, D. Next-generation sequencing and its clinical application. *Cancer Biol. Med.* **2019**, *16*, 4–10. [CrossRef]
196. Tay, T.K.Y.; Tan, P.H. Liquid Biopsy in Breast Cancer: A Focused Review. *Arch. Pathol. Lab. Med.* **2020**, *145*, 678–686. [CrossRef]
197. Kamps, R.; Brandão, R.D.; van den Bosch, B.J.; Paulussen, A.D.; Xanthoulea, S.; Blok, M.J.; Romano, A. Next-Generation Sequencing in Oncology: Genetic Diagnosis, Risk Prediction and Cancer Classification. *Int. J. Mol. Sci.* **2017**, *18*, 308. [CrossRef]
198. Nic Daeid, N.; Rafferty, A.; Butler, J.; Chalmers, J.; McVean, G.; Tully, G. *Forensic DNA Analysis: A Primer for Courts*; The Royal Society: London, UK, 2017.
199. Jordan, D.; Mills, D. Past, Present, and Future of DNA Typing for Analyzing Human and Non-Human Forensic Samples. *Front. Ecol. Evol.* **2021**, *9*, 646130. [CrossRef]
200. Yang, Y.; Xie, B.; Yan, J. Application of Next-generation Sequencing Technology in Forensic Science. *Genom. Proteom. Bioinform.* **2014**, *12*, 190–197. [CrossRef]
201. Tang, S.; Huang, T. Characterization of mitochondrial DNA heteroplasmy using a parallel sequencing system. *Biotechniques* **2010**, *48*, 287–296. [CrossRef]
202. Van Geystelen, A.; Decorte, R.; Larmuseau, M. Updating the Y-chromosomal phylogenetic tree for forensic applications based on whole genome SNPs. *Forensic Sci. Int. Genet.* **2013**, *7*, 573–580. [CrossRef]
203. Hajibabaei, M.; Shokralla, S.; Zhou, X.; Singer, G.A.C.; Baird, D.J. Environmental Barcoding: A Next-Generation Sequencing Approach for Biomonitoring Applications Using River Benthos. *PLoS ONE* **2011**, *6*, e17497. [CrossRef]
204. Phillips, C.; Prieto, L.; Fondevila, M.; Salas, A.; Gómez-Tato, A.; Álvarez-Dios, J.; Alonso, A.; Blanco-Verea, A.; Brión, M.; Montesino, M.; et al. Ancestry Analysis in the 11-M Madrid Bomb Attack Investigation. *PLoS ONE* **2009**, *4*, e6583. [CrossRef]
205. Han, J.; Kraft, P.; Nan, H.; Guo, Q.; Chen, C.; Qureshi, A.; Hankinson, S.E.; Hu, F.B.; Duffy, D.L.; Zhao, Z.Z.; et al. A Genome-Wide Association Study Identifies Novel Alleles Associated with Hair Color and Skin Pigmentation. *PLoS Genet.* **2008**, *4*, e1000074. [CrossRef] [PubMed]
206. Bocklandt, S.; Lin, W.; Sehl, M.E.; Sánchez, F.J.; Sinsheimer, J.S.; Horvath, S.; Vilain, E. Epigenetic Predictor of Age. *PLoS ONE* **2011**, *6*, e14821. [CrossRef] [PubMed]
207. Courts, C.; Madea, B. Micro-RNA—A potential for forensic science? *Forensic Sci. Int.* **2010**, *203*, 106–111. [CrossRef] [PubMed]
208. Minogue, T.D.; Koehler, J.W.; Stefan, C.P.; Conrad, T.A. Next-Generation Sequencing for Biodefense: Biothreat Detection, Forensics, and the Clinic. *Clin. Chem.* **2019**, *65*, 383–392. [CrossRef]
209. McEwen, S.A.; Wilson, T.M.; Ashford, D.A.; Heegaard, E.D.; Kournikakis, B. Microbial forensics for natural and intentional incidents of infectious disease involving animals. *Rev. Sci. Tech. l'OIE* **2006**, *25*, 329–339. [CrossRef]

210. Jäger, A.C.; Alvarez, M.L.; Davis, C.P.; Guzmán, E.; Han, Y.; Way, L.; Walichiewicz, P.; Silva, D.; Pham, N.; Caves, G.; et al. Developmental validation of the MiSeq FGx Forensic Genomics System for Targeted Next Generation Sequencing in Forensic DNA Casework and Database Laboratories. *Forensic Sci. Int. Genet.* **2017**, *28*, 52–70. [CrossRef]
211. Ballard, D.; Winkler-Galicki, J.; Wesoły, J. Massive parallel sequencing in forensics: Advantages, issues, technicalities, and prospects. *Int. J. Leg. Med.* **2020**, *134*, 1291–1303. [CrossRef]