**NEXT GENERATION SEQUENCING FOR CANCER DIAGNOSIS**

**Shrishti Mishra1 ,Priyanka Vishwakarma2 ,Nashra Iqbal1 ,Vidya Meenakshi\*1**

1-Department of Biotechnology, Isabella Thoburn College, Uttar Pradesh , Lucknow, India

2-Sanjay Gandhi Post Graduate Institute of Medical Sciences, Uttar Pradesh, Lucknow , India

nashraiqbalzaidi@gmail.com

**ABSTRACT**

The development of next-generation sequencing (NGS) technology has facilitated the study of cancer. Massive parallel sequencing made possible by NGS provides for the most thorough genomic analysis of tumors. Different NGS methods focus on DNA and RNA analysis. Whole-genome, whole-exome, and targeted DNA sequencing are types of sequencing that concentrate on a subset of genes that may be relevant to a certain condition. Alternative gene-spliced transcripts, post-transcriptional alterations, gene fusions, mutations/single-nucleotide polymorphisms, small and long noncoding RNAs, and variations in gene expression are all more easily found using RNA sequencing. The majority of NGS applications are in the realm of cancer research, but recently, NGS technology has revolutionized cancer molecular diagnostics because of the many benefits it provides over conventional approaches.

**Keywords:** Hereditary breast cancer, melanoma, prostate cancer, thyroid cancer, lung cancer, colorectal cancer, hematologic cancer

**INTRODUCTION**

The Estimated numbers of new cancer cases and deaths in Cancer, in its many forms, accounted for 1.9million deaths in 2022 [GLOBOCAN 2022] . Our understanding of this extremely complex and varied group of diseases has undergone a revolution thanks to the quick development of DNA sequencing tools [Devita and Rosenberg, 2012].This book chapter discusses current research and clinical applications before summarizing the history of massively parallel next-generation sequencing (NGS) in the context of cancer .Second, we emphasize the significance and possibility of entire or 100% genome sequencing, which is beyond the capabilities of the most recent NGS technology due to highly repetitive non-coding sequences. NGS, which debuted as a cutting-edge research tool a decade ago, is now the preferred technique for concurrent genomic sequencing of numerous cancer indicators. In terms of time savings, tissue preservation, and a precision oncology approach to patient treatment, NGS has undeniable advantages over conventional approaches for hospitals. Next-generation sequencing (NGS) technologies have become increasingly important in recent years for studying the altered genomic pathways connected to human cancer. Compared to older methods of genome-sequencing,[Serrati et al,2016]

This technique is high-throughput in the sense that it enables simultaneous sequencing of numerous targeted genomic areas in numerous samples in order to identify concurrent mutations in the same run. Regular tumors sequencing also has the benefit of faster turnaround times for analysis, which results in quicker clinical reporting. Moreover, compared to conventional sequencing techniques, an analysis in NGS only needs a little amount of input of DNA or RNA. A wide range of genomic aberrations, including single- and multiple-nucleotide variants, minor and large insertions and deletions, copy-number variations (CNVs), and fusion transcripts, can be examined simultaneously with high precision and sensitivity. Compared to Sanger sequencing, NGS has a better sensitivity (detection of 2%–10% versus 15%–25% allele frequency, respectively), allowing for quantitative evaluation of the mutant allele.

NGS has a variety of benefits, including the capacity to thoroughly sequence all types of mutations for many genes (hundreds to thousands) and the sensitivity, speed, and cheap cost compared to other sequencing modalities in a single test. Here, we discussed some of the impending difficulties as well as the technology, methodologies, and applications that can be immediately taken into consideration. [Genome Biol et al. 2013] A comprehensive technology utilized for sequencing (DNA) and gene expression (RNA species) study is next generation sequencing (NGS).

Since DNA, RNA, miRNA, ChIP, and methylation sequencing were the first areas of genomic research to adopt the NGS technique, it was created to address the Sanger sequencing constraint. [Slatko et al, 2018] .Like any technology, NGS offers a number of benefits that have elevated it to the status of being a vital tool in both the lab and the clinic. [Pereira et al,2017] .This technique has certain drawbacks, nevertheless, including the requirement for strong bioinformatics tools and qualified individuals for both experimental and data analysis, even after more than 15 years of development [ Pereira et al,2017]. In order to improve the diagnosis, prognosis, and therapy of numerous diseases, the data offered by NGS have proven beneficial and trustworthy for both research and in the clinic.[ Mellis et al,2018] and are commonly employed in the field of oncology[ Sabour et al,2017]. Routine tumor sequencing also has the advantage of faster turnaround times for analysis, which reduces the time required for clinical reporting. Additionally, compared to traditional sequencing methods, NGS analysis only requires a little amount of input DNA or RNA. It is possible to simultaneously analyze a number of genomic aberrations with exceptional accuracy and sensitivity, including single/multiple nucleotide variants, small and large insertions and deletions, copy-number variations (CNVs), and fusion transcripts. Since NGS has a higher sensitivity than Sanger sequencing (detection of 2%–10% versus 15%–25% allele frequency, respectively), it is possible to evaluate the mutant allele quantitatively.

Sequencing by synthesis, which employs the DNA strand to be sequenced as a template, is used by the majority of NGS platforms.[Li,W,et al 2015]. Optical imaging and four separate fluorescently-tagged nucleotides on IlluminaMiSeq and HiSeq sequencers are used to visualize the complementary strand. The anticipated error rate for Illumina technology is 0.4%.[Quail, Michael A.,et al.2012]

One nucleotide at a time is integrated, and as protons are released, the pH changes, signaling a pH change to the semiconductor chip. The identification of homopolymer stretches is expected to have an error rate of 1.8% to 1.9% with ion torrent technique.[ Ross, Michael G., et al 2013]

Due to this, NGS is being used in highly sensitive studies, such as the analysis of foetal DNA from maternal blood2 [Harris SR, Cartwright EJ, et al.2013] or keeping track of how many tumour cells are present in cancer patients' blood circulation.[ Dawson SJ,Tsui DW MurrtazaM,et al 2013]

1. **APPROACHES OF NGS IN CLINICAL DIAGNOSIS**

Next-Generation Sequencing, or NGS. Chain terminator sequencing, another name for the Sanger DNA sequencing technique, was created in 1997. This approach was the gold standard for sequencing up until the late 2000s, when it was automated and somewhat modified. [Marziali A., et al 2001].

Despite the Sanger technique's significant disadvantages of high cost and time, it was mostly employed as the only sequencing method for three decades. NGS (next-generation sequencing) technologies are quickly becoming one of the most efficient, rapid, and high-throughput ways to sequence DNA.

Various NGS platforms are:

* Roche/454 sequencing.
* Proton/PGM sequencing in an ion torrent
* Sequencing with Solexa by Illumina
* Sequencing with SOLiD.

**Roche/454 sequencing**.- Using the Roche/454 method The first NGS platform, Roche/454 sequencing, was introduced in 2005 by Rothberg and colleagues and is also known as pyrosequencing [Margulies et al. 2005]. The method involves pyrosequencing with pyrophosphate and is based on a sequencing-by-synthesis approach (PPi). The method was created by Ronaghi and associates (1996). Pyrosequencing's chemistry is based on the detection of pyrophosphate, which is released when a nucleotide is incorporated into a freshly created DNA strand. The four enzymes DNA polymerase, ATP sulfurylase, luciferase, and apyrase, as well as the substrates adenosine 5′phosphosulfate (APS) and luciferin, are combined with a sequencing primer that has been hybridised to a single-stranded biotin-labeled template [Gharizadeh et al. 2007; Nyren and Skarpnack 2001]. DNA samples are randomly broken for sequencing, or amplicons of appropriate sizes are generated. The segments are ligated to 454 sequencing adaptors in the following phase of library preparation. The adapter makes it possible to bind DNA fragments to streptavidin beads that have primers on their surface with oligonucleotides that are compatible with the DNA fragments. Each bead is separated into distinct emulsion droplets and is connected to a single fragment. Deoxynucleotide triphosphates (dNTPs) are added to the PCR mixture to begin the first step of the reaction cascade. Emulsion PCR (emPCR) is used to make millions of copies of each DNA fragment on the surface of each bead from these droplets.

One read equals one output from a single bead. The picotiter plate (PTP), a surface with nanometer-sized wells for pyrosequencing, is subsequently used to transfer the template beads. Following the inclusion of a nucleotide, the sequencing reaction advances through a sequence of downstream events to produce light. A pyrosequencing technique will be used to synthesis every DNA fragment in a single well. The use of PTP enables thousands of pyrosequencing processes to be done concurrently, greatly boosting the sequencing throughput [Mardis 2008]. By measuring the light that is released when PPi is converted to ATP and each nucleotide is incorporated by DNA polymerase, the precise nucleotide contributed is identified [Mardis 2008]. A CCD camera records data from PTP points that correspond to each cycle.

**Proton/Personal Genome Machine (PGM) Sequencing -** using Ion Torrent Thermo Fisher's Ion Torrent semiconductor sequencing technology is another sequencing platform for next-generation sequencing. Sequencing-by-synthesis is the name given to the method, which was first created by Toumazou et al. in 2006 [Rothberg et al. 2011]. Here, the synthesis and detection of an electrochemical signal—more specifically, the generation and detection of a hydrogen ion—is used to detect the addition of nucleotides. On a complementary metal-oxide semiconductor (CMOS) chip known as a flow cell, the sequencing reaction takes place. But unlike other next-generation sequencing technologies, it directly monitors the emission of H+ ions during the integration of dNTP to a developing DNA strand rather than employing fluorescently labelled nucleotides or chemiluminescence [Rothberg et al. 2011]. The

Compared to other NGS platforms, the Ion Torrent sequencers produce reads with lengths of 200, 400, and 600 bp. Ion Torrent's platform is distinctive since it only employs natural nucleotides. Therefore, Iron Torrent technology uses unmodified bases rather than fluorescently labelleddNTPs, and the signals are created as H+ ions rather than being identified as particular flourescence. Better enzyme activity, lengthy readings, and affordable consumables are the outcomes of the strategy. Low image scans are not required while doing "base calling" tasks, which are completed significantly more quickly. Ion Torrent devices can process up to 10 Gb of sequencing data each run in about 2.5 hours, according to the most latest standards [Ari and Arkan 2016].

**Solexa/Illumina Sequencing-** Scientists at Cambridge, Shankar Balasubramanian and David Klenerman, created techniques to generate high-quality reads of significantly larger data sizes at a lower cost in the middle of the 1990s. Using fluorescent reversible terminator deoxyribonucleotides, single DNA molecules are bonded to a flat surface, amplified in place, and sequenced. Images are captured from the fluorescent signals created during the reaction. In order to get high-quality sequence data, the surface images are finally processed and evaluated [Bentley et al. 2008]. The method was eventually known as Illumina sequencing technology after researchers formed the Solexa firm, which was later bought by Illumina in 2008. Solexa went on to market these as the Ilumina/Solexa Genome Analyzer (GA) [Balasubramanian 2015; Shendure and Ji 2008]. Currently, the business owns

The original sequencers Illumina/Solexa GA were able to produce paired-end (PE) short reads, which had the advantage of documenting the sequences at both ends of each DNA cluster, and very short reads (35 bp). The current generation of Illumina SBS technology-based instruments, which can produce many terabases (Tb) of data per run, were created through further optimization and refinement. The most recent Illumina sequencers have short read lengths of roughly 125 bp and deliver output data larger than 600 Gb. Illumina systems are said to offer 99.9% accuracy, and 96 samples can be barcoded in a single run using conventional reagents [Morey et al. 2013]. The use of Illumina sequencing technology offers benefits and drawbacks. less time was spent preparing the library

**SOLiD Sequencing** -by ABI Life Technologies also promotes the Supported Oligonucleotide Ligation and Detection (SOLiD) NGS platform. The first ABI/SOLiD sequencing machine was released in 2007, after Applied Biosystems (ABI) bought SOLiD. The sequencing by ligation (SBL) method is the foundation for SOLiD sequencing [Shendure and Ji 2008]. The nucleotide sequence is determined via sequencing by ligation, which makes use of the DNA ligase's mismatch sensitivity [Ho et al. 2011]. Jay Shendure was the first to illustrate the fundamental idea and group to sequence an evolved strain of Escherichia coli with a smaller error rate of 106. The method uses bead-based emulsion PCR of the molecules to create "polonies," sometimes known as polony "sequencing," and immobilisation of the beads onto a surface to create a cell-free, mate-paired library of DNA fragments.

1. **METHODS OF NGS**

Depending on the requirements and the questions to be answered, many strategies might be utilized. Genomic DNA (DNA-seq), messenger or non-coding RNA (RNA-seq), or any nucleic/ribonucleic material acquired following particular methods can be used as the initial input material. Processing of samples or production of libraries. The material is initially physically or enzymatically broken to produce pieces that are compatible with the sequencer (small fragments of 200–300 nucleotides for short-read sequencing, longer for the long-read sequencing). A small number of genomic regions can be examined using this material, such as disease gene panels or microorganisms.[ Duncavage, et al 2011] or the entire human genome's coding exons (from approximately 21,000 genes; Whole-Exome-Sequencing, WES). Whole-Genome-Sequencing, or WGS, is a method for sequencing all of the genomic DNA without the need for an enrichment step (see Section 2.2.3). There are designated areas of interest that will be analyzed (ROIs). In most instances, a PCR amplification step is performed with 4–12 cycles. The DNA fragments are given the appropriate linkers and barcodes at this stage, which are required for the sequencer's subsequent studies. DNA barcodes, which are distinctive nucleotide identifiers (6–8 nt long), enable pooling samples for sequencing processes in a single flow cell. Sequencing. Earlier descriptions of the most popular sequencers[Kamps, et al.2017] [Wright, et al 2017] [Colligs, et al.2017] [Park, et al 2016]. You can find a review of the various sequencing chemistries elsewhere[Goodwin el al.2016].Analysis of the initial quality and raw data.FastQC is mostly used for general quality control of the read quality.[ Andrews, et al.2010]. There are numerous pre-processing tools available for cutting, removing reads of poor quality, etc. Following mapping, specificity—i.e., the percentage (%) of all prescribed ROIs that are appropriately enriched and sequenced—is calculated.Calling variants and interpreting data. The final step depends on the particular application. Some techniques and bioinformatics tools pertinent to oncogenomics data interpretation will be provided in this review.

**Targeted sequencing: gene panels**

Because of its cost-effectiveness, high ROI coverage, and simplicity in the initial and subsequent data analyses, the use of gene-panels to sequence only a defined number of genes of interest has been the method of choice for the majority of clinical applications. Targeted re-sequencing is the term typically used when the number of genes analysed is limited to the handful that have already been examined in earlier diagnostic testing using conventional procedures.

There are various procedures available to create and capture gene panels and other ROIs. Most of the time, businesses that sell library preparation kits also offer user-friendly online tools for designing PCR oligos or hybridization probes that will enrich the appropriate ROIs. A form of enrichment [Kamps,et al .2017].

The latter offers the benefit of amplifying numerous targeted regions simultaneously into individual micro drops, maintaining each amplification distinct from the others and reducing the disturbance caused by primer pair interactions. The Molecular Inversion Probe is an affordable and adaptable technique for capturing tiny portions of the genome for NGS analysis (MIP)[Niedzicka,et al.2016].

After sequencing, raw data analysis is a fairly straightforward process. Due to the high specificity and low coverage per nucleotide, only specific DNA sections may be ineffectively collected (due to high GC nucleotide content, for instance). For the most part, germline variations that are present in homo- or heterozygous state can be found with a sequencing depth of about 80. Since somatic mutations typically occur at sub-clonal levels and result in low percentages, a higher coverage is necessary (>500).

**WES—Whole-Exome-Sequencing-**

The same or comparable methods as those discussed for the enrichment of gene-panels are used in the protocols/kits that are available from various firms to enrich the library for all exons. After sequencing, raw data analysis is important to assess the efficacy of the studies and look for potential issues that may have arisen during library creation and/or sequencing. For high-quality data to be obtained, both stages are essential. For sufficient specificity and sensitivity in mutation detection, a high sequence-on-target yield of more than 90% of the ROIs and coverage higher than 20 per nucleotide are required. Sample processing was typically subpar when less than 90% of the ROIs are sequenced but the coverage is high; when the ROIs are suitably sequenced (>90%) but the coverage is low, [BaoR ., et al 2014].

**WGS—Whole-Genome-Sequencing-**

When genetic tests based on WES produced negative results in families with a high likelihood of carrying a genetic mutation, WGS may be helpful in clinical diagnostics to uncover familial germline mutations.

The primary technological benefit of WGS is that it does not require any enrichment or amplification for the library preparation, and it achieves a theoretical specificity of 100% (really, it is closer to 95%–98%, with almost no missing gaps) with uniform ROI coverage over the input material. As a result, WGS does not have the potential to overlook a disease-causing variant (or any other information) due to technical issues like ineffective probe focused enrichment, ineffective amplification of a particular ROI, or PCR amplification artefacts. [Leggett R.M ., et al 2013] [EkblomR.,et al 2014] [Chrystoja C C ., et al 2014].

1. **DIAGNOSTIC POTENTIAL OF NGS IN VARIOUS CANCERS**

**Hematologic cancer and NGS-**

Genetic abnormalities, particularly significant mutations that underlie the many phenotypes in the range of hematopoietic tumors, constitute the basis for hematological malignancies. In a variety of contexts, NGS technologies have been used to study hematological disorders, including guiding diagnosis (TCR gene rearrangement to establish T-cell clonality), subclassification (recurrent cytogenetic translocations in acute myeloid leukemia), prognosis (Philadelphia chromosome-positive in acute lymphoblastic leukemia), and minimal residual disease (MRD) testing (BCR-ABL transcripts in chronic myelogen. [Hussaini M .et al 2015] [Black JS ,et al . 2015]

The accurate identification of more frequent mutations that may have significant predictive value and clinical significance is one aspect of the ongoing evolution in the characterisation of leukemias, lymphomas, and myelomas.

Critical cancer-gene mutations can be found in solid-tumor samples, which helps to clarify patient diagnosis and prognosis and to determine which targeted medicines can be used to enhance care for specific cancer patients. Here, NGS studies on solid cancer are described. These studies demonstrate advantages over conventional diagnostic techniques and provide a fundamental overview of how the cancer molecular approach is evolving.

**Familial breast cancer-**

About 30% of cases of hereditary breast cancer (HBC), which make up 5%–10% of all BCs, are brought on by BRCA1 and BRCA2 mutations. Tumor-suppressor proteins, which are crucial for DNA repair and genomic stability, are encoded by the BRCA1 and BRCA2 genes. For BC patients with early onset or a significant family history, genetic counselling and a BRCA gene test are advised due to the increased lifetime risk of developing HBC caused by the existence of these mutations.

Because the BRCA1 and BRCA2 genes include 23 and 27 exons, respectively, respectively, conventional DNA sequencing, such as direct Sanger sequencing, necessitates lengthy processing periods and expensive costs. To speed up the molecular analysis, prescreening techniques such denaturing high-performance liquid chromatography have been proposed.

Our lab's work and a number of recent publications have shown how NGS techniques work well to identify point mutations and indels in the BRCA1/BRCA2 genes, revolutionizing this genetic study and cutting down on time and expense.[PilatoB,et al 2016]

Since it is quicker and more sensitive than denaturing high-performance liquid chromatography/Sanger sequencing technologies, this methodology is actually appropriate for routine diagnostic procedure. Participation in international quality programmes on BRCA1/BRCA2 testing with the NGS method, such as the European Molecular Genetics Quality Network, which also allows for the acquisition of specific certification on correct results, sensitivity, specificity, and variant calling interpretation, ensures the quality of the data.

Today, it has been established that genes other than BRCA1/BRCA2 impart a high chance of developing BC. NGS technologies provide gene panel customization, giving patients more opportunities to assess their BC risk. [Lhota F, et al.2016]

**Lung cancer-**

The most common type of cancer-related death in developed nations is lung cancer (LC), which is frequently discovered at an advanced stage. The choice of LC patients for the use of tyrosine-kinase inhibitors has been made possible by full understanding of predictive indicators (TKIs). To treat patients for TKI medication appropriately in clinical practise, EGFR mutations must be assessed. However, other TKI-sensitive EGFR mutations can occur in exons 18–21. The majority (80%–90%) of EGFR mutations are either small exon 19 deletions or the L858R mutation in exon 21. Because it is linked to first-generation TKI resistance but third-generation TKI sensitivity, the mutation T790M in exon 20 has to be researched. [Cross DA, et al 2014].

ALK rearrangement is yet another indicator of TKI resistance. In fact, the FDA has only licensed medications for the treatment of LC for the two actionable genes EGFR and ALK to date.

Tissue that has been formalin-fixed and paraffin-embedded is regarded as the best specimen for molecular examination. Sanger sequencing served as the industry standard method for years to find EGFR mutations, but more recently, other techniques have been used in molecular diagnostics (high resolution melting, restriction fragment-length polymorphism, mutant allele-specific PCR, peptide nucleic acid-mediated PCR, pyrosequencing, immunohistochemistry with specific EGFR antibodies, and the Scorpion Amplification Refractory Mutation System). The gold standard for research into ALK rearrangements is still immunohistochemistry or fluorescence in situ hybridization.

Numerous studies have shown that the use of NGS into routine clinical practise for LC molecular diagnostics has resulted in a number of modifications.

As a matter of fact, Lim et al. recently reported that 58% of patients with wild type by routine testing for EGFR/KRAS/ALK had changes discovered by NGS, offering these patients a treatment chance. [Lim SM et al. 2016].

1. **OPPORTUNITIES FOR CANCER DIAGNOSIS**

NGS deployment in a clinical setting offers a number of intriguing possibilities, therefore it's not unexpected that there is a great deal of interest and engagement in this area. Some of these techniques will merely replace current Sanger sequencing or PCR-based assays for genetic testing within genes associated with familial cancer syndromes or for spotting mutations in therapeutically significant genes within cancer cells or tissues. The capacity to quickly and cheaply screen a large number of gene targets will be a key driver for many applications. Despite ongoing legal disputes over infringements of gene patents, this should result in more people having access to genetic testing and a successful conclusion for patients and their families. The appearance

It's not surprising that there is a great deal of interest and engagement in this field given that NGS deployment in a clinical environment offers a number of exciting possibilities.

Some of these methods will merely replace the Sanger sequencing or PCR-based tests that are currently used to identify mutations in therapeutically important genes within cancer cells or tissues or to investigate the genetics of genes linked to familial cancer syndromes.

For many applications, the ability to quickly and affordably screen a large number of gene targets will be crucial.

This should lead to more people having access to genetic testing and a happy ending for patients and their families, notwithstanding ongoing legal fights over gene patent infringements.

Small chemical inhibitors and antibodies targeting druggable gene targets are revolutionizing the treatment of cancer. Many of these drugs are regarded to function best in conjunction with further diagnostic tests.

One of the many therapeutic applications that are now being developed is the monitoring of disease burden through the least invasive detection of tumour DNA in the peripheral blood of cancer patients.

1. **NGS TECHNOLOGY: OPTIONS AND CHALLENGES**

Unlike conventional sequencing methods, NGS can sequence the genome in a great deal of parallel. However, the bulk of NGS platforms use sequencing-by-synthesis (SBS) for sequencing, much like the older methods. [Metzker, ML et al . 2010].

This method obtains the sequence of the template strand by synthesizing a complementary strand using the DNA strand that needs to be sequenced as a template. The most popular technique makes use of optical imaging and four different fluorescently tagged nucleotides to see the complementary strand developing (as in the case of MiSeq and HiSeqIllumina Sequencers). As all fluorescent tags must be scanned for sequencing, this is known as 4-channel sequencing. [Pareek, C.S ., et al 2011].

Three fluorescently labelled nucleotides (C, T, and A) and one unlabeled nucleotide (G) are used for sequencing using a newly released NGS sequencer from Illumina (NextSeq). For 2-channel imaging, filters are employed to detect C bases (red fluorescence) and T bases separately (green fluorescence). The green and red fluorescent tags used to label the Abases appear as yellow fluorescence with both filters. A G base inclusion is believed to be the absence of any fluorescence signal. The sequencer from Pacific Biosciences also employs optical detection, which involves imaging the clearly labelled nucleotides that the DNA polymerase holds prior to integration during SBS in order to determine the sequence of the template. [Chaisson, MJ.,et al 2015] [Quail, et al.2012].

In addition to these technologies, a number of non-optical genome sequencing methods have also been reported. One such method is the semiconductor-based Ion Torrent sequencing method, which has garnered a lot of popularity.[ Rothberg, et al 2011].

[SBS is another technique used in the semiconductor-based Ion Torrent technology, where sequencing is carried out in microscopic wells connected to a semiconductor chip. On microscopic beads, the DNA of interest is clonally amplified before unlabeled nucleotides are added one at a time in a specified order. The pH changes after integration as a result of the protons released from the 31-OH group during the creation of the phosphodiester bonds, and the semiconductor chip measures this change. There have been multiple reports of this technology's validation and use in research and therapeutic settings. [Singh, et al 2014] [Singh,et al.2013] [Merriman, et al 2012].

Additionally, the Nanopore technology (Oxford Nanopore Technologies, Oxford, UK), a non-SBS and non-optical method of sequencing, has been described. It involves applying an electric field to a protein nanopore to move single strands of nucleic acids through it. Each nucleotide's characteristic modification of the electric field, which is brought about by the transit of the nucleotides through the pore, aids in deducing the sequence. The Nanopore technique has significant advantages over the aforementioned sequencing methods, including as minimal pre-sequencing, a tiny footprint, adjustable run lengths, and lengthy reads, but it is constrained by high error rates. [Kilianski, et al .2015] [Vercaoutere, et al.2001].

1. **CLINICAL REPORTING AND NGS DATA ANALYSIS**

Every NGS run typically includes the parallel sequencing of numerous barcoded and multiplexed samples from various genomic regions. Base-calling, which assigns the base sequence using the signal read-out, is typically the first step in the NGS data analysis pipeline. This could involve pH change monitoring or optical measurement of fluorescent tags on nucleotides (Illumina and PacBio) (Ion Torrent). Following base-calling, the sequence reads are aligned to a reference genome. Typically, to make the alignment process simpler for targeted sequencing, the specific regions of interest in the genome are defined. Different platforms favour different alignment strategies depending on the outcome of their sequencing. For instance, the Burrows-Wheeler Aligner is used by the MiSeq Reporter software (BWA).[ Li H., et al.2010].

[Using sequence alignment, it is possible to identify genomic abnormalities by filtering and removing off-target reads and comparing on-target reads. The process of calling variants is known as variant calling, and the aberrations may be as straightforward as single or multiple nucleotide variants (SNVs and MNVs) or as complex as small- and large-scale insertions and deletions, gene copy number variants (CNVs), or gene fusions. It may be difficult to accurately identify the complicated genomic variants, making the selection of the best variant-calling algorithm essential. Multiple variant-calling techniques may be utilized to cover the complete range, particularly for complex insertion-deletions, CNVs, and gene fusions. [Ye, K., et al.2009] [McPherson, et al 201q] [McPherson, et al 2011].

Clinical laboratories, which are often habituated to low- or medium-throughput procedures, find it difficult to manage the large volumes of data provided by NGS. To provide complete results traceability, NGS data must be archived, along with records of multiple software and algorithm versions that adhere to regulatory agency requirements. Numerous file formats, including FASTQ (base calling and quality scores), BAM, SAM (post-alignment information), and VCF, are produced as a result of the numerous phases in the processing of NGS data (variant calls). It is impractical to keep all of the information created during NGSanalysis because it might involve everything from tiny gene panels to big gene panels, as well as whole exome to whole genome analysis. In general, files that contain sufficient information to repeat the analysis are sufficient in accordance with the recommendations from the[Aziz,et al.2015] [Rehm,et al.2013].

In general, variant calling produces a large number of variant calls, including both real and erroneous calls (sequencing artifacts, errors caused by repetitive sequences such as tandem repeats and homopolymers). A significant interpretation difficulty is removing erroneous mutations for clinical interpretation and reporting. The variant calls are filtered for this purpose based on a variety of standards provided by the laboratory during assay validation. Generally speaking, this entails taking into account sufficient sequencing quality, sequencing depth, allelic frequency, connection with tumour percentage, presence of the variant in both forward and reverse sequencing reads, or lack of strand bias. Utilizing applications like Integrated Genome Viewer, directly visualizing the sequence readings (IGV) [Thorvaldsdottir, et al.2013].

such as UCS genome browser[Kent, et al. 2002].is crucial in determining the legitimacy of the variant before reporting and plays a big role in this process.

The natural growth of platforms toward higher levels of multiplexing, involving the screening of numerous markers in many samples, is a common trend seen in all molecular technologies. Improved technologies are desirable because of these clear benefits, but they come with difficulties when it comes to evaluating performance and interpreting more complicated data. This pattern has characterised the development and use of the ground-breaking NGS technologies, which have replaced the conventional low- and medium-throughput sequencing techniques. Due to its inherent advantages, NGS over previous orthogonal sequencing technologies is preferred for mutation screening of malignancies for both research and clinical diagnostics. Given the fast rate of new marker discovery, the utility of NGS for routine diagnostics is only going to increase. This will call for frequent testing of several indicators,

Recent years have seen tremendous advancements in target capture, library preparation, and sequencing technologies, which have fueled a rise in NGS adoption and use. Additionally, a number of studies that thoroughly analysed every component of NGS technology, including wet bench procedures and platforms, informatics for data processing, and testing prices, have made it clearer which choices are best. The validation and application of NGS-based assays in a clinical diagnostic environment has also been clarified by recommendations and guidelines created by several regulatory authorities. The anticipated introduction of cutting-edge sequencing technologies, such Nanoporetechnology, has the potential to completely transform parallel sequencing and further solidify NGS as the dominant clinical sequencing technology. These cutting-edge technologies could possibly replace or supplement existing ones.

1. **CONCLUSION**

In this article, we discussed the past, present, and future of next generation sequencing in cancer diagnosis. As advancements in sequencing technology enabled investigators to study cancer diagnosis in greater breadth and depth, the field has produced novel insights into tumor pathogenesis, identified clinically useful biomarkers, and developed increasingly precise diagnostics and targeted therapeutics. The clinical usefulness of next generation sequencing requires advancement in our knowledge of the genome and bioinformatic systems to process genetic data—advancements that would be built on a foundation of 100% next generation sequencing.

NGS methods, which represent a significant revolution in genome sequencing, are able to handle the difficulties brought on by the growing need for routine mutation profiling of malignancies. However, due to this technology's high level of complexity and performance validation, successful adaption in the context of clinical diagnosis presents significant problems. The implementation of NGS technologies is now more practical, making them the most preferred large-scale genome sequencing technologies. This is due to increased clarity regarding the validation and implementation of NGS tests by several regulatory agencies, published reports from several clinical laboratories, and technological advancements.

**REFERENCE**

1. Serratì, Simona, et al. "Next-generation sequencing: advances and applications in cancer diagnosis." OncoTargets and therapy 9 (2016): 7355.
2. Slatko, Barton E., Andrew F. Gardner, and Frederick M. Ausubel. "Overview of next‐generation sequencing technologies." Current protocols in molecular biology 122.1 (2018): e59.
3. Pereira, M., et al. "Application of next-generation sequencing in the era of precision medicine." Applications of RNA-Seq and Omics Strategies: From Microorganisms to Human Health. London, England, UK: InTechOpen (2017): 293-318.
4. Mellis, R.; Chandler, N.; Chitty, L.S. Next-generation sequencing and the impact on prenatal diagnosis. Expert Rev. Mol. Diagn. 2018.
5. Sabour, L.; Sabour, M.; Ghorbian, S. Clinical Applications of Next-Generation Sequencing in Cancer Diagnosis. Pathol. Oncol. Res. POR 2017,
6. Li W, Zhao K, Kirberger M, Liao W, Yan Y. Next generation sequencing technologies in cancer diagnostics and therapeutics: a mini review. Cell MolBiol (Noisy-le-grand) 2015;61(5):91–102.
7. Quail MA, Smith M, Coupland P, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and IlluminaMiSeq sequencers. BMC Genomics. 2012;13:341.
8. Ross MG, Russ C, Costello M, et al. Characterizing and measuring bias in sequence data. Genome Biol. 2013;14(5):R51
9. Dawson SJ, Tsui DW, Murtaza M, et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer. N Engl J Med 2013;368:1199–209.
10. Pilato B, Pinto R, De Summa S, et al. BRCA1-2 diagnostic workflow from next-generation sequencing technologies to variant identification and final report. Genes Chromosomes Cancer. 2016;55(10):803–813. [PubMed] [Google Scholar] [Ref list]
11. Lhota F, Zemankova P, Kleiblova P, et al. Hereditary truncating mutations of DNA repair and other genes in BRCA1/BRCA2/PALB2-negatively tested breast cancer patients. Clin Genet. 2016;90(4):324–333.
12. Cross DA, Ashton SE, Ghiorghiu S, et al. AZD9291, an irreversible EGFR TKI, overcomes T790M-mediated resistance to EGFR inhibitors in lung cancer. Cancer Discov. 2014;4(9):1046–1061.
13. Lim SM, Kim EY, Kim HR, et al. Genomic profiling of lung adenocarcinoma patients reveals therapeutic targets and confers clinical benefit when standard molecular testing is negative. Oncotarget. 2016;7(17):24172–24178.
14. Marziali A., Akeson M. New DNA sequencing methods. Annu. Rev. Biomed. Eng. 2001;3:195–223. doi: 10.1146/annurev.bioeng.3.1.195.
15. Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 17(6):333
16. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA et  al (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376
17. Gharizadeh B, Ghaderi M, Nyrén P (2007) Pyrosequencing technology for short DNA sequencing and whole genome sequencing. SeibutsuButsuri 47:129. <https://doi.org/10.2142/biophys.47.129>
18. Nyren P, Skarpnack SE (2001) Method of sequencing DNA based on the detection of the release of pyrophosphate and enzymatic nucleotide degradation. Retrieved from <http://ip.com/patent/US6258568>
19. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M et al (2011) An integrated semiconductor device enabling non-optical genome sequencing. Nature 475:348
20. Ari S, Arıkan M (2016) Next-generation sequencing: advantages, disadvantages, and future. In: Plant omics: trends and applications. Springer, Cham, pp  109–135. <https://doi.org/10.1007/978-3-319-31703-8_5>
21. Balasubramanian S (2015) Solexa sequencing: decoding genomes on a population scale. ClinChem 61(1):21–24
22. Shendure J, Ji H (2008) Next-generation DNA sequencing. Nat Biotechnol 26(10):1135–1145. <https://doi.org/10.1038/nbt1486>
23. Morey M, Fernandez-Marmiesse A, Castineiras D, Fraga JM, Couce ML, Cocho JA (2013) A glimpse into past, present, and future DNA sequencing. Mol Genet Metab 110(1–2):3–24. <https://doi.org/10.1016/j.ymgme.2013.04.024>
24. Ho A, Murphy M, Wilson S, Atlas SR, Edwards JS (2011) Sequencing by ligation variation with endonuclease V digestion and deoxyinosine-containing query oligonucleotides. BMC Genomics 12(1):598
25. Duncavage, Eric J., et al. "Hybrid capture and next-generation sequencing identify viral integration sites from formalin-fixed, paraffin-embedded tissue." The Journal of molecular diagnostics 13.3 (2011): 325-333.
26. Kamps, Rick, et al. "Next-generation sequencing in oncology: genetic diagnosis, risk prediction and cancer classification." International journal of molecular sciences 18.2 (2017): 308.
27. Wright, Marvin N., Damian Gola, and Andreas Ziegler. "Preprocessing and quality control for whole-genome sequences from the IlluminaHiSeq X platform." Statistical human genetics. Humana Press, New York, NY, 2017. 629-647.
28. Colling, Richard, Lai Mun Wang, and Elizabeth Soilleux. "Validating a fully automated real-time PCR-based system for use in the molecular diagnostic analysis of colorectal carcinoma: a comparison with NGS and IHC." Journal of Clinical Pathology 70.7 (2017): 610-614.
29. Park, Sang Tae, and Jayoung Kim. "Trends in next-generation sequencing and a new era for whole genome sequencing." International neurourology journal 20.Suppl 2 (2016): S76.
30. Kosack, Cara S., Anne-Laure Page, and Paul R. Klatser. "A guide to aid the selection of diagnostic tests." Bulletin of the World Health Organization 95.9 (2017): 639.
31. Fürjes, Péter. "Controlled focused ion beam milling of composite solid state nanopore arrays for molecule sensing." Micromachines 10.11 (2019): 774.
32. Goodwin, Sara, John D. McPherson, and W. Richard McCombie. "Coming of age: ten years of next-generation sequencing technologies." Nature Reviews Genetics 17.6 (2016): 333-351.
33. Andrews, S. "FastQC: a quality control tool for high throughput sequence data. Available online." Retrieved May 17 (2010): 2018.
34. Kamps, Rick, et al. "Next-generation sequencing in oncology: genetic diagnosis, risk prediction and cancer classification." International journal of molecular sciences 18.2 (2017): 308.
35. Niedzicka, Marta, et al. "Molecular Inversion Probes for targeted resequencing in non-model organisms." Scientific Reports 6.1 (2016): 1-9.
36. Leggett R.M., Ramirez-Gonzalez R.H., Clavijo B.J., Waite D., Davey R.P. Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics. Front. Genet. 2013;4:288. doi: 10.3389/fgene.2013.00288.
37. Ekblom R., Wolf J.B. A field guide to whole-genome sequencing, assembly and annotation. Evol. Appl. 2014;7:1026–1042. doi: 10.1111/eva.12178.
38. Chrystoja C.C., Diamandis E.P. Whole genome sequencing as a diagnostic test: Challenges and opportunities. Clin. Chem. 2014;60:724–733. doi: 10.1373/clinchem.2013.209213.
39. Hussaini M. Biomarkers in hematological malignancies: a review of molecular testing in hematopathology. Cancer Control. 2015;22(2):158–166.
40. Black JS, Salto-Tellez M, Mills KI, Catherwood MA. The impact of next generation sequencing technologies on haematological research: a review. Pathogenesis. 2015;2(1–2):9–16.
41. Metzker, M.L. Sequencing technologies—The next generation. Nat. Rev. Genet. 2010, 11, 31–46. [CrossRef][PubMed]
42. . Pareek, C.S.; Smoczynski, R.; Tretyn, A. Sequencing technologies and genome sequencing. J. Appl. Genet.2011, 52, 413–435. [CrossRef] [PubMed]
43. Chaisson, M.J.; Huddleston, J.; Dennis, M.Y.; Sudmant, P.H.; Malig, M.; Hormozdiari, F.; Antonacci, F.;Surti, U.; Sandstrom, R.; Boitano, M.; et al. Resolving the complexity of the human genome usingsingle-molecule sequencing. Nature 2015, 517, 608–611. [CrossRef] [PubMed]
44. Quail, M.A.; Smith, M.; Coupland, P.; Otto, T.D.; Harris, S.R.; Connor, T.R.; Bertoni, A.; Swerdlow, H.P.;Gu, Y. A tale of three next generation sequencing platforms: Comparison of Ion Torrent, Pacific BiosciencesandIlluminaMiSeq sequencers. BMC Genomics 2012, 13, 341. [CrossRef] [PubMed]
45. 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 genomesequencing. Nature 2011, 475, 348–352. [CrossRef] [PubMed]
46. Singh, R.R.; Patel, K.P.; Routbort, M.J.; Aldape, K.; Lu, X.; Manekia, J.; Abraham, R.; Reddy, N.G.;Barkoh, B.A.; Veliyathu, J.; et al. Clinical massively parallel next-generation sequencing analysis of 409cancer-related genes for mutations and copy number variations in solid tumours. Br. J. Cancer 2014, 111,2014–2023. [CrossRef] [PubMed]
47. Singh, R.R.; Patel, K.P.; Routbort, M.J.; Reddy, N.G.; Barkoh, B.A.; Handal, B.; Kanagal-Shamanna, R.;Greaves, W.O.; Medeiros, L.J.; Aldape, K.D.; et al. Clinical validation of a next-generation sequencing screenfor mutational hotspots in 46 cancer-related genes. J. Mol. Diagn. 2013, 15, 607–622. [CrossRef] [PubMed]
48. Merriman, B.; Rothberg, J.M. Progress in ion torrent semiconductor chip based sequencing. Electrophoresis2012, 33, 3397–3417. [CrossRef] [PubMed]
49. Kilianski, A.; Haas, J.L.; Corriveau, E.J.; Liem, A.T.; Willis, K.L.; Kadavy, D.R.; Rosenzweig, C.N.; Minot, S.S.Bacterial and viral identification and differentiation by amplicon sequencing on the MinIONnanoporesequencer. Gigascience 2015, 4, 12. [CrossRef] [PubMed]
50. Vercoutere, W.; Winters-Hilt, S.; Olsen, H.; Deamer, D.; Haussler, D.; Akeson, M. Rapid discriminationamong individual DNA hairpin molecules at single-nucleotide resolution using an ion channel.Nat. Biotechnol. 2001, 19, 248–252. [CrossRef] [PubMed]
51. Li, H.; Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics2010, 26, 589–595. [CrossRef] [PubMed]
52. ] and Torrent Suite software on Ion Torrent–PGMand Ion Proton uses the Torrent Mapping Alignment Program (TMAP)
53. Merriman, B.; Rothberg, J.M. Progress in ion torrent semiconductor chip based sequencing. Electrophoresis2012, 33, 3397–3417. [CrossRef] [PubMed]
54. Ye, K.; Schulz, M.H.; Long, Q.; Apweiler, R.; Ning, Z. Pindel: A pattern growth approach to detect breakpoints of large deletions and medium sized insertions from paired-end short reads. Bioinformatics 2009, 25,2865–2871. [CrossRef] [PubMed]
55. McPherson, A.; Hormozdiari, F.; Zayed, A.; Giuliany, R.; Ha, G.; Sun, M.G.; Griffith, M.; HeraviMoussavi, A.; Senz, J.; Melnyk, N.; et al. deFuse: An algorithm for gene fusion discovery in tumor RNA-Seqdata. PLoSComput. Biol. 2011, 7, e1001138. [CrossRef] [PubMed]
56. Aziz, N.; Zhao, Q.; Bry, L.; Driscoll, D.K.; Funke, B.; Gibson, J.S.; Grody, W.W.; Hegde, M.R.; Hoeltge, G.A.;Leonard, D.G.; et al. College of American Pathologists’ Laboratory Standards for Next-GenerationSequencing Clinical Tests. Arch. Pathol. Lab. Med. 2015, 139, 481–493. [CrossRef] [PubMed]
57. Rehm, H.L.; Bale, S.J.; Bayrak-Toydemir, P.; Berg, J.S.; Brown, K.K.; Deignan, J.L.; Friez, M.J.; Funke, B.H.;Hegde, M.R.; Lyon, E.; et al. ACMG clinical laboratory standards for next-generation sequencing.Genet. Med. 2013, 15, 733–747. [CrossRef] [PubMed]
58. Thorvaldsdottir, H.; Robinson, J.T.; Mesirov, J.P. Integrative Genomics Viewer (IGV): High-performancegenomics data visualization and exploration. Brief. Bioinform. 2013, 14, 178–192. [CrossRef] [PubMed]
59. Kent, W.J.; Sugnet, C.W.; Furey, T.S.; Roskin, K.M.; Pringle, T.H.; Zahler, A.M.; Haussler, D. The humangenome browser at UCSC. Genome Res. 2002, 12, 996–1006. [CrossRef] [PubMed]