

# NEXT GENERATION SEQUENCING FOR CANCER DIAGNOSIS

## Abstract

The development of sequencing of the next generation (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. Genome on the whole part, whole-exome, and targeted DNA sequencing are several classes of sequencing that concentrate on a subdivision of genetic factors that may be related to a certain condition. Alternative gene-spliced transcripts, small and long non-coding RNAs, mutations/single-nucleotide polymorphisms, post-transcriptional alterations, gene fusions and variations in genetic expression are all more easily found using RNA sequencing. The majority of NGS implementations are in the realm of cancer research, but recently, NGS technology has revolutionized molecular diagnosis of cancer because of the many benefits it provides over conventional approaches.

**Keywords:** cancer, Next Generation Sequencing, NGS, mutation, breast cancer, melanoma, prostate cancer, thyroid cancer, lung cancer, hematologic cancer

## Authors

### **Nashra Iqbal**

Department of Biotechnology  
Isabella Thoburn College  
Lucknow, Uttar Pradesh, India  
nashraiqbalzaidi@gmail.com

### **Priyanka Vishwakarma**

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

### **Vidya Meenakshi**

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

## I. INTRODUCTION

The Estimated numbers of new cancer cases and deaths in Cancer, in its many subdivisions, reported for 1.9million demises in 2022 [GLOBOCAN 2022]. Our comprehension 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 chronicles of largely side by side next-generation sequencing (NGS) in the context of cancer .Secondly, we emphasize the significance and possibility of entire or 100% sequencing of the genome, 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 rapidly, augmenting vital in recent years for analyzing the changed pathwaysof the genome related to human cancer. Compared to older methods of genome-sequencing [Serrati et al,2016]

This approach is high bit rate in the manner that it facilitates concurrent sequencing of numerous aimed genomic spheres in numerous specimens so as to figure out simultaneous 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 insertion of DNA or RNA. A wide variety of genomic abnormalities which includes single nucleotide variants and multiple nucleotide variants, minor insertions, large insertions and deletions, copy-number variations (CNVs), and fusion-transcripts. These can be inspected simultaneously with high sensitivity and precision. Compared to the Sanger sequencing, NGS has a better sensitivity (detection of 2% to 10% versus 15% to 25% allele frequency, respectively).

NGS has a variety of advantages including the ability to do indepth sequencing of all types of mutations for many genes (hundreds to thousands) and the sensitiveness, pace, and low rate costs compared to another sequencing manners in a single test. Here, we discussed some of the impending difficulties as well as the science, methodologies, and advantages that can be taken at the moment into consideration. [Genome Biol et al. 2013] A comprehensive technology utilized for sequencing (DNA) and expression of the gene (species of RNA) study is next generation sequencing (NGS).

DNA sequencing, RNA sequencing, miRNA sequencing, ChIP sequencing, and methylation sequencing adopted the NGS technique [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 disadvantages, but, taking the requirement into consideration for powerful bioinformatics tools and eligible personalities for both experimental and data analysis, furthermore, in 15 years of development [ Pereira et al,2017]. In order to upgrade the prognosis and diagnosis and therapy of numerous illness, the information offered by NGS have demonstrated beneficial and trustworthy in the clinical reasearches. 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 quantity of inserted DNA or RNA. It is possible to simultaneously analyze a number of genomic abnormalities with exceptional accuracy and sensitivity, including single nucleotide variants, multiple nucleotide variants, small/large insertions, deletions, copy-number variations and fusion transcripts. Since NGS has a higher sensitivity when compared to Sanger sequencing (detection of 2%–10% opposite to 15 percent –25 percent frequency of allele, respectively), it is possible to estimate the value and the mutant allele quantitatively.

Sequencing by combination, which employs the DNA strand to be sequenced as a template, is employed by the majority of NGS platforms [Li, W, et al 2015]. Optical imaging and four separate fluorescently-tagged nucleotides on Illumina MiSeq and HiSeq sequencers are used to envision the reciprocating strand. The anticipated inaccuracy figure 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 blood [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]

## II. APPROACHES OF NGS IN CLINICAL DIAGNOSIS

Chain terminator sequencing, or the Sanger DNA sequencing, was established in 1997. This approach was the best benchmark for sequencing up until the time behind 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 sole sequencing method for three decennium. NGS technologies are extremely becoming one of the most efficient, rapid, and high-bit rate 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.

**1. Roche/454 Sequencing:** Using the Roche/454 method The first NGS programme, Roche/454 sequencing, inaugurated in the year 2005 via Rothberg et al. and is also known as pyro-sequencing [Margulies et al. 2005]. The method involves pyrosequencing with pyrophosphate and is dependent 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 namely luciferase, DNA polymerase, ATP sulfurylase, and also apyrase with the reactants luciferina and APS (adenosine 5'phosphosulfate) are fused with primer used in sequencing which has been cross-bred to a single-stranded template containing biotin-labeled [Gharizadeh et al. 2007; Nyren and Skarpnack 2001]. DNA samples are randomly broken for sequencing, or amplicons of appropriate dimensions are created. The segments are joined to adaptors of 454 sequencing in the following phase of library preparation. The adapter makes it possible to bind DNA snippets to beads containing streptavidin 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 mix to begin the first step of the reaction cascade. Emulsion PCR (emPCR) is used to create several thousands of copies of every DNA fragment on the exterior top of each bead from these droplets.

One read equals one output from a single bead. The picotiter plate (PTP), an exterior with wells for pyrosequencing that are upto the size of a nanometer- 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 transformed to ATP and each nucleotide is encompassed by DNA polymerase, the precise nucleotide contributed is identified [Mardis 2008]. A CCD camera records data from PTP points that correspond to each cycle.

- 2. Proton/Personal Genome Machine (Pgm) Sequencing:** Torrent semiconductor sequencing technology is one more sequencing objectives for NGS. 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 recognition of signal of electrochemical nature—more categorically, the generation and noticing of a ion of hydrogen—is used to detect the addition of nucleotides. On a (CMOS) Complementary Metal-oxide Semiconductor Chip known as a flow cell, the sequencing reaction takes place. But unlike various NGS technologies, it directly monitors the release of Hydrogen 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 sequencers of Ion Torrent generates data with lengths of 200, 400, and 600 bp. Ion Torrent's platform is distinctive since it only employs natural nucleotides. Therefore, technology of Iron Torrent uses unchanged bases rather than fluorescently labelled dNTPs, and the signals are created as Hydrogen ions rather than being identified as particular flourescence. Enhanced enzyme activity, lengthy readings, and affordable comestibles 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].

- 3. Solexa/Illumina Sequencing:** Cambridge scientists, namely David Klenerman and Shankar Balasubramanian, created techniques to generate high-performance reads of significantly larger data sizes at a lower cost in the middle of the 1990s. Using fluorescent reversible terminator deoxyribonucleotides, solo molecules of DNA are bonded to a even surface, intensify 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 (GA) Illumina/Solexa Genome Analyzer [Balasubramanian 2015; Shendure and Ji 2008].

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 terminals of each DNA group, and very short reads (35 bp). The current generation of instruments of Illumina SBS technology, which can produce many terabases (Tb) of data/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 specimens can be barcoded in a one take using conventional reagents [Morey et al. 2013]. The use of Illumina sequencing technology offers benefits and drawbacks. less time was spent preparing the library

- 4. Solid Sequencing:** by ABI Life Technologies is seen to encourage the Supported Oligonucleotide Ligation and Detection (SOLiD) NGS platform. The initial ABI/SOLiD sequencing machine was released in 2007, after Applied Biosystems (ABI) bought SOLiD. The sequencing by joining called 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 adapted specimen of *Escherichia coli* with a smaller inaccuracy figure of 106. The method employment of bead-dependent emulsion PCR of the molecules to generate "colonies," sometimes known as polony "sequencing," and immobilisation of the beads on a surface to create a library of DNA fragments that is mate-paired and cell-free

### III.METHODS OF NGS

Depending on the requirements and the questions to be answered, many strategies might be utilized. Genomic DNA (DNA-seq), non-coding RNA (RNA-seq), miRNA, any nucleic/ribonucleic material acquired following particular methods can be used as the starting insertion substance. Processing of samples or production of libraries. The substance is starting enzymatically or physically disintegrated to give rise to fragments that are ok with the sequencer (compact particles ranging from 200 to 300 nucleotides for short-read sequencing and 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 units called exons (from about 21,000 genes in number; Whole-Exome-Sequencing, WES). Whole-Genome-Sequencing is a procedure for sequencing all of the genomic DNA unaccompanied by the need for an enhancement procedures (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 needed for the sequencer's related studies. DNA barcodes, which are distinctive nucleotide identifiers (6–8 nt long), enable samples of pooling for sequencing processes in a single flow cell. 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 et al.2016]. Analysis of the initial quality and raw data. FastQC is mostly used for common standard control of the observed 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 rich and sequenced—is calculated. Calling variants and interpreting data. The final step depends on the particular application. Some techniques and tools related to bioinformatics pertinent to oncogenomics data elucidations will be provided in this review.

**1. Targeted Sequencing- Gene Panels:** Since of its cost-effectiveness, elevated ROI description, and simplicity in the initial and subsequent data analyses, the employment of gene-panels to only sequence a definite number of genes of target has been the method of choice for the majority of clinical applications. Targeted re-sequencing is the term typically used when the count of genes analysed is limited to the handful that have already been examined in earlier diagnostic testing making use of conventional procedures.

There are various procedures accessible to create and capture gene panels and other ROIs. Most of the time, businesses that sell kits for library preparation also offer easy 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 selected area simultaneously into individual drops of micro size, maintaining each intensification distinct from the rest and reducing the interruption caused by interactions of primer pair. 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. Because of the elevated specificity and low coverage/nucleotide, only specific DNA sections may be ineffectively collected (because of high content of GC nucleotide, for example). For the most part, germline variations that are known to be found in homogenous or heterozygous state can be found with a sequencing depth of about 80. Since somatic mutations typically occur at certain sub-clonal levels and result in lower percentages, a higher coverage is necessary (>500).

**2. WES—Whole-Exome-Sequencing:** The same or comparable methods as those discussed for the improvement of panels of genes are used in the protocols/kits that are available from various firms to make better the library for every exons. After sequencing, unprocessed information analysis is important to assess the efficacy of the studies and look for potential concerns that may have arisen during library creation and/or sequencing. For high-quality data to be obtained, both stages are essential. For abundant specificity or sensitivity in detections of mutations, an increased sequence yield on target

of 90% extra of the ROIs and higher coverage than 20/nucleotide are needed. Sample processing was typically subpar when less than 90 percent of the ROIs are sequenced but their is high coverage; when the ROIs are satisfactory sequenced (greater than 90%) but is low coverage. [BaoR ., et al 2014].

- 3. WGS—Whole-Genome-Sequencing:** When tests of genetics dependent on WES produced negative results in species or genera with a high likelihood of conveying a genetic mutation, WGS may be helpful in clinical diagnostics to uncover familiar mutations of germline.

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

#### IV. DIAGNOSTIC POTENTIAL OF NGS IN VARIOUS CANCERS

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

The accurate recognition of more recurrent mutations that may have important prognostic figure 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 detection and prediction and to determine which chosen medicines can be used to enhance care for specific cancer patients. Here, NGS studies that had put light on solid cancer are elaborated. These studies demonstrate advantages over conventional diagnostic techniques and provide a basic outline of how the approach towards cancer molecular studies is evolving.

- 2. 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. Proteins of Tumor-suppressor nature, which are crucial for repair of the DNA and stability of genes, are encoded by the BRCA1 and BRCA2 genes. For BC patients with initial induction or a significant family history, genetic counselling and a BRCA gene test are

advised due to the elevated lifetime chance of developing HBC caused by the existence of such mutations.

Because the BRCA1 and BRCA2 genes include 23 and 27 exons, respectively, respectively, standard sequencing of DNA for example direct Sanger sequencing, necessitates lengthy processing periods and expensive costs. To speed up the molecular analysis, prescreening techniques such liquid chromatography of denaturing high-performance 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]

- 3. Lung Cancer:** The most usual type of cancer-related demise in developed nations is lung cancer (LC), which is frequently discovered at an advanced stage. The choice of LC sufferers for the utilization of inhibitors that are tyrosine-kinase based 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 EGFR mutations that are TKI-sensitive can happen in 18–21 exons. The (80%–90%) majority of EGFR mutations are each of the L858R mutation in exon 21 or small 19 exon deletions. The reason that it is linked to first-generation TKI resistance but 3<sup>rd</sup> generation TKI sensitivity, the mutation T790M in exon number 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-firmed and paraffin-implanted 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 lately, more techniques have been used in diagnostics at molecular level (peptide nucleic acid-mediated PCR, restriction fragment-length polymorphism, high resolution melting, mutant allele-specific PCR, immune histochemistry with specific EGFR antibodies, pyrosequencing and the Scorpion Amplification Refractory Mutation System). The gold standard for research into ALK rearrangements is still fluorescence in situ hybridization or immunohistochemistry.



Numerous researches have shown that the use of NGS into routine clinical practise for LC diagnostics at molecular level has resulted in a number of modifications. As a matter of fact, Lim et al. lately mentioned that patients (around 58%) with wild type by regular testing for EGFR/KRAS/ALK had alterations came across by NGS, offering these sufferers a treatment option. [Lim SM et al. 2016].

## **V. 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 engrossment and engagement in this area. A few of these techniques will merely return current PCR-based test procedures and Sanger sequencing or for testing based on genetics within genes associated with related syndromes related to cancer or for spotting mutations in therapeutically significant genes found in 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 approach to genetic testing and a successful conclusion for sufferers and their pedigree. 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.

A few of such techniques will merely return the PCR-based or Sanger sequencing 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.

## **VI. 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 HiSeq Illumina 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 unlabelled nucleotide (G) are employed for sequencing using a newly released NGS sequencer from Illumina (NextSeq). For 2-channel imaging, filters are employed to detect C bases that shows red fluorescence and T bases separately giving green fluorescence. The green and red fluorescent tags used to label the A bases 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 before hand to integration during SBS in order to determine the sequence of the template. [Chaisson, MJ., et al 2015] [Quail, et al.2012].

Moreover, except a these technologies, a number of methods of non-optical genome sequencing have also been found out. 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 wells of microscopic capacity connected to a semiconductor chip. On microscopic beads, the DNA of our target is clonally amplified before unlabeled nucleotides are added one by one in a specified order. The pH changes after integration as a result of the protons set free from the 3'-OH group during the creation of the bonds of phosphate, 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 methods (such as those called non optical) of sequencing, has been described. It involves applying an electric field to a protein nanopore to proceed 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] [Vercautere, et al.2001].

## **VII. 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 measurement of fluorescent tags optically on nucleotides (Illumina and PacBio) (Ion Torrent). Following base-calling, the sequence reads are assembled to a control 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 untargeted reads and comparing on-target reads. The process of calling variants is known as calling of the variant, and the abnormalities may be as straightforward as single or multiple variants based on nucleotides (SNVs and MNVs respectively) or as complex as small- and large-scale deletions and insertions, gene fusions and gene copy number variants (CNVs), 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 high level insertion and 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 a range from low level to mediocre level of high bit rate procedures, find it difficult to manage the huge content 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, enlisting - FASTQ (quality scores and base calling ), BAM, SAM (after 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 NGS analysis 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 data to duplicate the observation are enough 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, malinformation sourced by repetitive sequences such as homopolymers tandem repeats). A significant interpretation difficulty is removing erroneous mutations for clinical understanding and reporting. The variant calls are filtered for this purpose based on a variety of standards provided by the labs during assay authentication. Generally speaking, this entails taking into account ample sequencing standards, depth in sequencing , frequency of allele, linking with tumour%, variant presence found in both forward and reverse sequencing reads, or absence 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 genuine growth of platforms with regard to higher levels of multiplexing, including the screening of numerous markers in many specimens, 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. Mentioned the rapid evaluation of

new marker discovery, the utilization of NGS for regular diagnostics is only going to elevate. This will call for frequent testing of several indicators

Recent years have seen tremendous advancements in capture of the target, preparation library and technologies based on sequencing, 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 informatics for data processing and platforms, and testing prices, have made it clearer which choices are best. The validation and application of NGS-dependent on assays in an environment of clinical diagnostics has also been clarified by directions and instructions created by several regulatory authorities. The anticipated introduction of cutting-edge sequencing technologies, such as Nanopore technology, has the capability to completely transform parallel sequencing and additionally solidify NGS as the dominant clinical sequencing technology. These cutting-edge innovations could possibly replace or supplement existing ones.

## VIII. CONCLUSION

In this write up, we talked through the evolution of all times of next generation sequencing in illness caused by cancer and its diagnosis. Further discoveries in sequencing spheres gave rise to many researchers to develop in depth knowledge of cancer diagnosis in greater breadth, this particular sphere of innovation has paved way for novel perceptions into pathogenesis of tumor, discovered clinically biomarkers that are useful, and made risingly accurate diagnostics and targeted therapeutics. The clinical usefulness of next generation sequencing calls for high tech approaches in our data of the genome and bioinformatic systems of bioinformatics to operate genetic data—developments that would be developed on a base of 100% next generation sequencing.

NGS methods, which reflect a significant uprising in sequencing of the genome, are able to handle the difficulties brought on by the growing necessity for procedural 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 enactment of NGS technologies is now evermore practical, making them the most recommended extensive genome sequencing technologies. This is due to elevated clearance referring the authentication and usage 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, Arikan 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](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 neurouology 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 using single-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 Biosciences and Illumina MiSeq 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 genome sequencing. *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 409 cancer-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 screen for 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. *Electrophoresis* 2012, 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 MinION nanopore sequencer. *Gigascience* 2015, 4, 12. [CrossRef] [PubMed]
- [50] Vercoutere, W.; Winters-Hilt, S.; Olsen, H.; Deamer, D.; Haussler, D.; Akeson, M. Rapid discrimination among 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. *Bioinformatics* 2010, 26, 589–595. [CrossRef] [PubMed]
- [52] ] and Torrent Suite software on Ion Torrent–PGM and Ion Proton uses the Torrent Mapping Alignment Program (TMAP)
- [53] Merriman, B.; Rothberg, J.M. Progress in ion torrent semiconductor chip based sequencing. *Electrophoresis* 2012, 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-Seq data. *PLoS Comput. 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-Generation Sequencing 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-performance genomics 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 human genome browser at UCSC. *Genome Res.* 2002, 12, 996–1006. [CrossRef] [PubMed]