International Journal of Innovation and Scientific Research
ISSN 2351-8014 Vol. 70 No. 2 Jan. 2024, pp. 72-80
© 2024 Innovative Space of Scientific Research Journals
http://www.ijisr.issr-journals.org/

Next Generation Sequencing and its Future

Suchindra Suchindra[1], Preetam Nagaraj[2], and Asanka Sudeshini Hewage[3]

[1]Department of Engineering, National Institute of Mental Health and Neurosciences, Bangalore, India, India

[2]Department of Engineering, IBM, Bangalore, India, India

[3]Institute of Biochemistry, Molecular Biology and Biotechnology, University of Colombo, Colombo, Sri Lanka

Copyright © 2024 ISSR Journals. This is an open access article distributed under the ***Creative Commons Attribution License***, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT: Next-generation sequencing (NGS) has the capability to produce hundreds and thousands of DNA / RNA sequences per analysis. This ability to sequence thousands of sequences at a time has brought about profound changes in genomic research. The primary aim is to see regions of similarity or variation / mutation amongst many sequences at hand. Today, NSG has reduced the time and is cost effective. This reduction of cost and time to obtain genomic data quickly has provided information on gene structure, protein structure, protein expression, genetic variations, gene expression profiles and others. This range of information has now expanded genomics research from analyzing and studying rare genetic disorders, cancer, infectious diseases like COVID, Phylogenetic trees, animal research, disease and control in husbandry, oral diseases, and metagenomic studies.

The analyzed information has enabled diseases diagnosis, prognosis, targeted therapies, precision medication and other methods. The scope is not only restricted to human genome analysis but is used in disciplines like animals, and agriculture in detecting pathogens and developing new diseases resistant crops and animal breeds. Moreover, the NGS capacity to use massive parallel processing to sequence has had a major impact in pharmaceutical industry. New drugs discovery uses DNA / RNA sequencing methods to first identify mutations, to develop targeted medicine / drug development. NSG has a bright future in unlocking various mysteries and providing insights into genomic information which we have still not been able to uncover. This paper first talks about NGS techniques, challenges, techniques used, NGS uses currently, future development directions in sequenced data quality, cost, and analysis techniques.

KEYWORDS: Next-generation sequencing; Genome; Exon; Gene; Phylogenetic tree; Metagenomic.

1 INTRODUCTION

Over the last 60 - 70 years, researchers have tolled deep into developing better newer techniques to understand the DNA residue in the samples either by sequencing entire DNA or RNA molecules. It all began with Watson and Crick famously soling the structure of DNA in 1953 [1]. This created an overall conceptual framework for DNA replication and encoding proteins in nucleic acids [1]. However, scientists were still not able to read the sequences as today to see any mutations or variations. Strategies to study the chain of proteins could not be used for DNA sequences as they were too long and made of fewer amino acids (A, C, G, T), which presented confusion, misunderstanding and time consumption.

At the same time, there was another school of scientist who were interested in sequencing either the whole sequence or wanted to find regions which were like two sequences, giving birth to global and local sequence pairwise alignment algorithms. Of these, the first edition of these algorithms was all based on Dynamic Programming methods and were very sensitive but time consuming. These are Needleman-Wunsch and Smith Waterman algorithms [2], [3]. As time progresses, heuristic algorithms which were faster but less sensitive were developed for local and global sequence alignment. The new bioinformatics algorithms varied in finding smaller regions of similarity, to finding Maximal match subsequence where were employed in LAGAN and MASAA family of algorithms. BLAST, LAGAN and MASAA family continue to dominate this spectrum

Corresponding Author: Suchindra Suchindra

of whole genome pairwise sequencing today as they are faster and cheaper (BLAST family) or more sensitive (LAGAN and MASAA family) [4 – 17], on this topic we have extensive research done and are presenting the findings. The widely used and effective ones are BLAST, LAGAN and MASAA as they are faster and sensitive at the same time.

While technologies were still in their infancy in 'reading' DNA sequences, it was the RNA - tRNA molecule which was first sequenced by Robert Holley in 1965, quickly after RNA of bacteriophage MS2 by [18, 19]. The techniques used in sequencing RNA sequence was quickly adopted to sequence DNA with the breakthrough by Federick Sanger and colleagues, who developed chain termination method [20]. There after automated DNA sequencing method was developed by [21, 22]. From the mid 80's there has been an explosion of sequencing methods especially in the last few decades [23 – 27, 59, 60]. The development of which has led to many pathbreaking discoveries in the field of DNA sequencing.

In the next section we will first discuss 3 generations of NGS algorithms starting from the Sanger discovery to third generation, techniques used, and uses of NGS at the end. The evolution of sequencing is shown in Fig 1. Looking at Fig 1, one can see how far we have come, today we are able to sequence 1000Gb of data in a single run which was impossible to fathom few decades ago. In the last years, the reduction of time and cost to obtain genetic information (genome, exome, and genes) have led to serious discoveries in other fields like animals, insects, birds, genetic diseases, protein structure, microbiology, cancer and metagenomic studies.

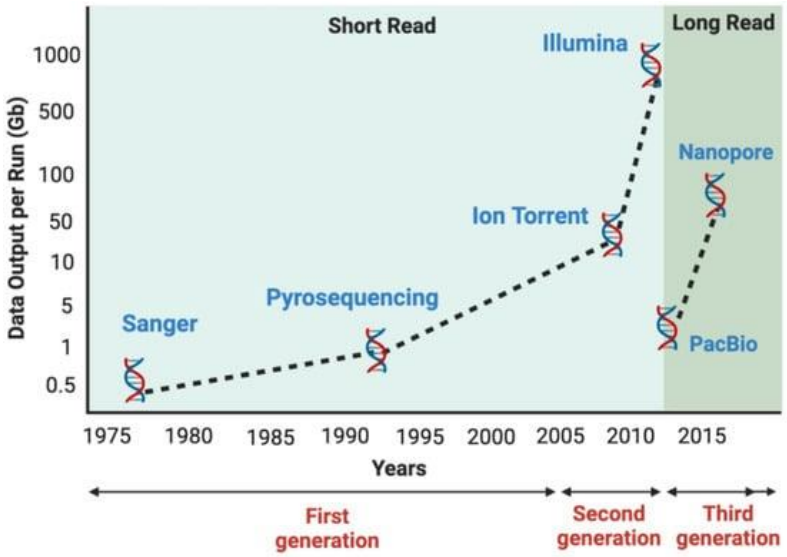

Fig. 1. Evolution of Sequencing Techniques [28]

1.1 FIRST GENERATION SEQUENCERS

The first generation is widely recognized as a period between year 1964 to 2001. The first attempts were made at sequencing DNA and RNA by Robert Holey who first sequenced a nucleic acid molecule, alanine tRNA using ribonuclease from S. cerevisiae [29]. However, a breakthrough came about in 1977, when Frederick Sanger developed a new sequencing method based off chain-termination method (Which later came to be known as Sanger Sequencing). At the same time, Walter Gilbert developed a method based on chemical degradation of DNA and its cleavage at specific base pair in the same year [31]. Sanger method used dideoxy nucleotides, which terminate the chain elongation of DNA stands during replication and allowed sequence reading up too few hundred bp in length. At the time, DNA sequencing was manual, time consuming and used radioactive materials. Sangers' methods were highly efficient with low radioactivity, hence was quickly adopted as the primary technique in laboratory and commercial application [30].

The first commercial automated sequencing machine, ABI 370 from Applied Biosystems was launched in 1987. ABI 370 used dideoxy nucleotides and capillary electrophoresis to mimic and automate sangers sequencing, which increased its efficiency in terms of speed and accuracy [32, 33]. ABI 370 was able to detect 96 base pairs at a given time, 500K per day and could read bases as long as 600 bases. 600 bases length is considered short read today and we will talk about these at the end of this section. ABI 370 gained fame quickly and adopted and many improvements were made to it over time, which led to higher-throughput sequencers capable of producing longer read up to few thousands [34]. This sums up the generation one sequencers which were later replaced by higher-through put sequencers while reducing manpower time and cost employing

massive parallel analysis techniques. Nonetheless the first generation made historical milestones which revolutionized many areas of biology, botany, pharma, and medical industry.

1.2 SECOND GENERATION SEQUENCERS

Second generation NGS sequencing techniques which emerged from 2001 - 2008. These systems were now capable of simultaneously sequencing thousands of millions of DNA markers. These methods differed from the first generation as these were able to perform parallel sequencing. One of the popular methods to emerge in this period is the Roche 454 method, which relied on pyrosequencing [28], and could produce 20,000 reads and an output of 20Mb in a single run. Later improvements were made to the 454 methods, to develop 454 GS FLX titanium which could produce 700 bp reads and could produce 14 Gb on a single run. The commercial 454 GS FLX system was available in 2009.

Ion torrent genome machine was another method like Roche 454, employed changes to pH, by detecting release of hydrogen ions during DNA synthesis to determine the sequence. Illumina sequencing method utilizes a sequence by synthesis which is based on the reversible dye terminators [28]. This system used sequencing by synthesis method (SBS) and the amplification bridge [28]. The analyzer can output 1GB / run which was later updated to 600 Gb/run and 200 bp reads which is by far the lengthiest sequence read of this generation. Sequence by Oligonucleotide ligation and detection (SOLiD), as the name employs, uses ligation method using reversible termination to determine the DNA sequence. All these second-generation methods increased the throughput in this short period of time at the same time increasing the speed of DNA sequencing [35]. This led to the gene therapy first used in breast cancer, targeted medication, and care. Various NGS technologies are shown in Figure 2.

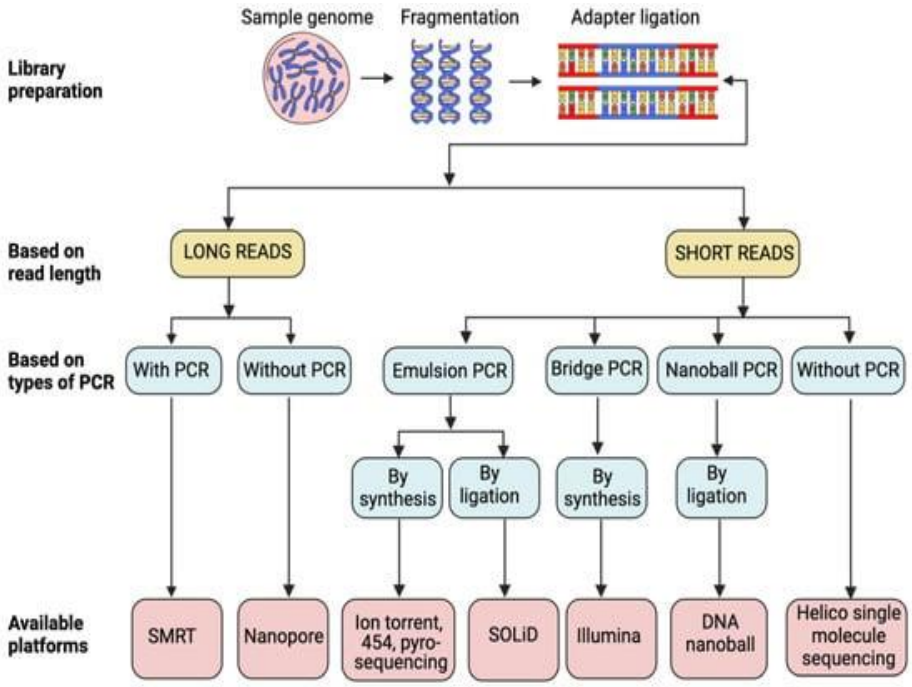

Fig. 2. Various NGS technologies [28]

Oral health is often forgotten as being part of the overall 'well-being' of an individual. More so, in India and south Asia in general. Th word, "well-being", means the state of mind, physical, social well-being, or lifestyle in totality [1]. A good oral health leads to good dental functionality as well as plays a vital role in the aesthetic part of an individual, making one pretty or handsome or looked upon as so. A poor oral health on the other hand would result in diminished confidence, shyness and a conscious awareness that certain activity is not well looked upon by others when an individual exercises it. E.g., Smiling with crooked teeth. Hence oral health plays a role in both functional as well as social lifestyle.

1.3 THIRD GENERATION SEQUENCERS

Third-generation sequencing methods are techniques used from the year 2009 - till date. Top of the list is the Helicos single molecule sequencing which is highly accurate. This method can output nearly 28 Gb, the only disadvantage being its limited capacity to correctly identify indels. Indels are regions of mutation, which can see as insertion or a deletion, symbolized by '-' in the DNA sequence. Another disadvantage is the short reads which are well below 50 bp and at the time when other methods produce 100k bp, 50 bp looks rather too small.

Other third generation sequencing methods provided long-read sequencing capabilities, enabling for much large DNA fragments, one of which is PacBio sequencing, which used single-molecule real time (SMRT) approach, which enabled for long reads up to tens of kbs. One of the most popular one in this generation is the Oxford Nanopore Minion technology, which as the name gives away, uses nanopore sequencing technologies, which is based on DNA molecules, when an electric potential is applied, through an array of nanopore in a high-salt buffer. Bases are identified based off electric change to determine the DNA sequence. Oxford Nanopore provides long reads and real-time analysis [28]. With the advent of third generation sequencing methods, scientist today can identify infectious diseases in real time and are able to analyze quickly without any historical data related to these diseases.

A short-read sequencing is basically sequencing by synthesis based upon other enrichment methods like fragmentation, amplification, and hybridization [28]. Long-read, on the other hand are based on detection either by synthesis or electrical voltage change/ variation to generate single base when it is passed through a membrane or a buffer. Long read tend to produce 10kb read while short reads produce around 400 - 600 bp. Long-reads are error prone when compared to short-reads. Also, the quality of data produced in short reads are far more accurate than long-reads. Thus, short reads find its usefulness in counting of specific sequences, profiling, identification of variants and others. Long reads find their usefulness in identifying large insertions / deletions, duplications, and others [36].

2 AREAS WHERE NGS IS USED AND DATA ANALYSIS

NSG is used in whole Genome Sequencing research. The well-known genome project from USA, Europe and Japan used NGS, all these projects were mainly focused on the analysis of the entire nucleotide sequence of the genome of an organism, initially they started with drosophila and human genome project elevated the project to use it for human genome analysis [37]. Genome analysis is quite complex, and voluminous as there are 3 billion characters to analyze from the resultant data. Being complex, and voluminous, it was costly in the beginning, although costs have dramatically come down over the last few years to few thousands of dollars, its data analysis still presents a challenge, since it is genome analysis is interested in knowing the entire complete picture of organism in interest [38], see Fig 3.

NGS is also used in Whole exome sequencing (WES), which is more interested in the protein coding exons (stretch of bps). Since the protein coding exons makes as little as 2% of the 3 billion length genome sequence, it offers opportunity for larger wider study on these exons [39]. Since the exons lengths do not run into thousands of bps in length, the study is quite cheaper, less voluminous, and less complex when compared to (WGS).Since WES is interested in sequencing smaller section of the whole genome, knowledge of smaller intricacies between exon and non-exon regions might be missed to garner and one might miss the complete picture of the information flow between exon and non-exon region as a whole [40].

NGS is well known in in Targeted sequencing. targeted sequencing as the name gives away, targets specific regions of a gene, and can identify various types of variation from small gene deletion (whole gene), gene duplications, insertion (Mutated gene) or gene rearrangement (Gene Mutated), which are all associated with diseases in one way or the other. In some cases, abnormalities like clefts, abnormalities during birth, (extra digits), missing nose and others are all attributed to gene mutations. The main advantage of Targeted sequence is that it is less expensive and clinical establishments can manage this small data, making laboratory decision making quicker and easier. Over the course of time targeted sequencing has been able to throw light on the mutant clones which are due to tumour heterogeneity or metastases nature of cancerous tumour [41]

Next-generation Sequencing has had a profound impact on transcriptomics [28], by improving the study of RNA molecules in an organism. Since NGS offers better output, and is cost inexpensive in profiling and analyzing molecules, scientists or laboratories have been able to gain much better knowledge on the gene expressions, location where these genes are expressed inside the cell, splicing, non-coding regions of RNA and various other biological mechanism which are one or the other way tied to our understanding of the diseases and drug discovery [42 - 45]. mRNA sequencing is the most common used in transcriptomics, mainly for getting a snapshot of expressed gene in a sample [28]. Since NGS provides short reads and long reads, short reads here are helpful in identifying expressed genes levels (in percentage) in different conditions, newer unknown transcripts if any, and to gain knowledge on expression mechanism on different cell types and their gene expression levels at

the same time. NGS is used in alternative splicing too, which is basically a technique where a single gene generates mRNA isoforms (same form). It does this by aligning RNA-seq reads to the reference genome (NSG generated), thereby abetting scientist to locate splicing junctions [46].

NGS is helpful to identifying non-coding RNAs, apparently these regions play an important role in gene regulations [28]. Small-RNA and long non-coding RNA sequencing enables the scientist to classify these regions of non-coding RNAs. The classification allows for profiling regulatory RNAs, microRNAs, PiRNAs and snoRNAs for their role in gene expression [28], [12]. Other long RNA-seq reads can inform the connectivity between other exons and reveal the information which is otherwise thought to be 'junk' [47]. NSG is also used nowadays in the reconstruction and annotation of transcriptome of organisms by aligning RNA-seq reads to a reference sequence. NGS as we can see can play an integrated role in various molecular fields and help us understand the cell regulation mechanisms and other biological processes necessary for a healthy cell and its reproduction.

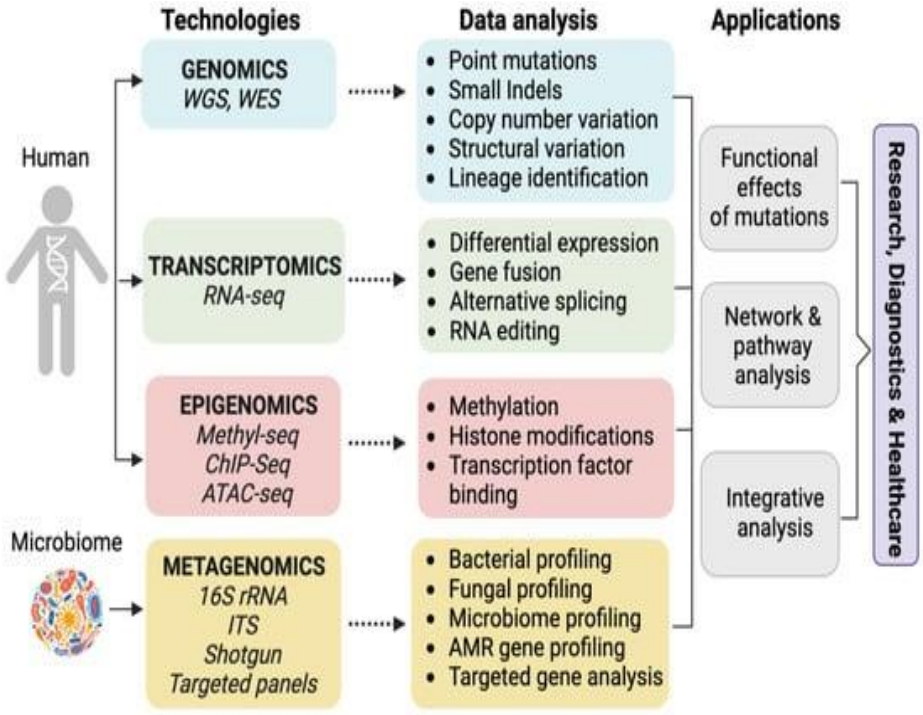

Fig. 3. Areas NGS is used [28]

Epigenomics is other field where NSG is used actively. It is a study where hereditary gene expression patterns have changed but the DNA sequences or the genome has not mutated at all [48]. The most common fields under this category are DNA methylation, histone modification and RNA methylation [49]. The chemicals tags associated with the organism alter the DNA accessibility, chromatin remodeling and nucleosome positioning [50, 51]. All this modification, it is attributed to environmental factors such as diet, pollutants, lifestyle, toxic intakes, exposure to radiation and information professions [52]. The knowledge and insights garnered by scientist through whole genome projects has helped them attain good knowledge into these gene alterations. This unusual changes in the gene expression patterns have bewildered scientist and has infused more interest in understanding the disorders like behavior disorders, memory loss, cancer, immune system depletion, addiction, lethargy, aggression, neurodegeneration, and other psychological disorder attributed to this change in gene expression patterns [28].

Next-generation sequencing generates vast amount of data which is then used by other bioinformatics / computational method to analyze and interpret these data. This raw data produced by NGS must be processed, analyzed, and interpreted to make good biological sense / insights and based on these insights some knowledge is gained or reaffirmed [28]. The computational method ranges from DNA / RNA sequence alignment both in pairwise and multiple sequence alignment, variant determination, gene expression, expression analysis, protein structure development, and other specialized analyses. There are other bioinformatics tools like de novo assembly, reference-based mapping, and transcriptome analysis, which are used to gain more biological information from the sequences generated by NGS [28]. Other tools are used for identification of gene variation, including both single-nucleotide polymorphisms and copy number variations (CNV) [28]. Other combine the NGS

data with other genomic and functional data readily available and annotated to explore gene expression and regulatory networks at the molecular level [28].

3 NGS AND RESEARCH

Fgdfd NGS has revolutionized the field of microbiology for good, especially the scientific community is able to synthesize genomic large data because of its parallelized millions of sequences high-throughput in a single producing high-throughput. NGS is looked upon as a weapon in translation medicine because of this high throughput but also because of the vast bioinformatic tools that can now be used upon which help researchers, laboratory scientist, drug companies understand the genetic nature of the disease and thereby aid in the new drug discoveries.

NSG is used in the Microbiomes research. Microbes are everywhere, their relationship with humans at times pathogenic and at times symbiotic leads to greater importance in understanding their ecosystem, their survival characteristics through their and our evolution [53]. The understanding the host (Human, plant etc.) and host-microbiome interactions are important for the host survival and reproduction. These include pathogen surveillance, therapeutic and functional dysbiosis [28]. There are studies made where the gut microbes are linked to obesity and other mental disorders. Experiments are successfully conducted where gut microbes are switched between healthy humans to see the effects and the results are stunning [54]. Designing panels to identify mutation or mutated genes, followed by usage of bioinformatics tools to see the sequence alignment, genes would now help in designing drugs for the pathogens. targeted medication would then ensure at least a recovery from these pathogen and aid in the immune system building resistance to these pathogens in the future [28].

Today, NGS research is focused more on the study of transcriptome and epigenome [28]. The human genome project and other genome projects through WGS and WES have provided scientists with knowledge of the microbiological process as minute as a protein synthesis and its location inside the cell today. Thus, the research has moved to other areas like wellness check, agriculture for diseases and pest resistance crops, and food research. The NGS research has also widen into genetic variation, phylogenetic trees, ancestral linkages, and cancerous cell through the mutations biologists can now discover. This need in detecting the mutational landscape where a benign tumor would lead to malignant through gene variation finders like 261 gene panel, 400 gene panel, 500 panel by Illumina, IST, Agilent and Thermo Fisher has flooded the market with new devices and procedures [28]. The panels can discover various cancers like lung (cigarette smoking led research), breast cancer (female), ovarian (female), pancreatic and renal cancers [28].

NGS is useful in finding the exact microbes which indirectly helps in developing the correct medicine and its dosage to help an individual. The panels discussed in the previous paragraph help detect the pathogens from various diseased reference or clinical annotated specimen. The panels also help in identifying antimicrobial drug-resistant mutations and antiviral drug-resistance mutations [55]. The data from these identified mutations are later used for surveillance, containment, drug discovery, policy changes, health interventions and vaccine discovery to stem the epidemic like we saw in COVID [56]. NSG is activity used in the gene mutations where scientist find a correlation with diseases like diabetes, infertility, immune system changes, hormonal changes, embryo development and others.

On the cancer diagnostic front, NGS main contribution has been that all cancers without exception have been a disease of the genome and not necessarily mendelian in their origin [28]. By mandelian, we mean, diseases inherited with single gene mutation or gene pattern mutations. There are many efforts being made on the cancer cure direction, TCGA (The Cancer Genome Atlas) and ICGC (International Cancer Genome Consortium), our understanding of cancer and its subsequent gene alteration data in protein-coding regions inside the cell for all cancer types are available today [57]. Then there are enterprises, like such as FoundationOne by Foundation Medicine (Cambridge, MA, USA), Oncomine by Thermo Fisher (Waltham, MA, USA), CANCERPLEX by KEW (Cambridge, MA, USA), MSK-IMPACT by the Memorial Sloan Kettering Cancer Center (New York, NY, USA), OmniSeq Advance by the Roswell Park Cancer Institute (Buffalo, NY, USA), the CC Onco Panel by Sysmex (Kobe, Japan), and the Todai Onco Panel by Riken Genesis (Tokyo, Japan) who have developed different gene panels using various technologies that are now used for cancer detection, prognosis and therapy [58].

4 CONCLUSION

The future of NGS is very bright. Today, NSG is steppingstone on which other fields like Bioinformatics, bio robotics, Nucleic acid preparation are based off. With the advancement in computers and other technologies, the sequencing method would only get faster and accurate with time. We see, that in the future, Sequence would be handled by a smaller amounts of input DNA and would be simpler test results in the laboratory settings. This would eventually help other diagnostic tools ranging from health, dental, and other botanical fields especially agriculture where there is a need for diseases and pest resistant crops.

NSG holds immense prospects in the cancer diagnostics and gene expression study. We have still a long way to know the minute chemical reactions that go inside the cell and an understanding of those would help in various fields especially cancer. More time and money in spent each year across the world in different well know laboratories to find these intricacies leading to cancer cure. The development of real-time sequencing and point-of-care application [28] would help monitor health, spread of diseases, and help develop faster and better drugs to counter these diseases. The vast data that NGS produces also need faster, better bioinformatics tools to analyse and find better understanding of the biomechanism at microbiology cells throwing new insights about the cell. We are confident that NGS would open new frontiers in both faster disease diagnostics tools, better and accurate bioinformatic tools, drug discovery and perhaps a cure for cancer.

REFERENCES

[1] Watson J., Crick F. Molecular structure of nucleic acids. *Nature.* 1953; 171: 709–756.

[2] Needleman, S.B. and Wunsch, C.D. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48: 443-453.

[3] Smith, T.F. and Waterman, M.S. 1981. Identification of common molecular subsequences. J. Mol. Biol. 147: 195-197.

[4] Reddy, Bharath Govinda. *Multiple Anchor Staged Local Sequence Alignment Algorithm-MASAA*. Diss. The University of Northern British Columbia, 2009.

[5] Waqar Haque, Alex Aravind, and Bharath Reddy. 2009. Pairwise sequence alignment algorithms: a survey. In Proceedings of the 2009 conference on Information Science, Technology and Applications (ISTA '09). ACM, New York, NY, USA, 96-103. DOI=http://dx.doi.org/10.1145/1551950.1551980.

[6] Haque, W., Aravind, A., & Reddy, B. (2008, August). An efficient algorithm for local sequence alignment. In *2008 30th annual international conference of the IEEE engineering in medicine and biology society* (pp. 1367-1372). IEEE.

[7] D. J. Lipman and W. R. Pearson, «Rapid and Sensitive Protein Similarity Searches,» *Science,* vol. 227, pp. 1435-1441, 1985.

[8] S. F. Itschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, «Basic Local Alignment Search Tool,» *J. Molecular Biology,* vol. 215, pp. 403-410, 1990.

[9] S. F. Altschul, et al., «Gapped BLAST and PSI-BLAST: a new generation of protein database search programs,» *Nucleic Acids Res.,* vol. 25, pp. 3389-3402, Sep. 1997.

[10] W. J. Kent, «BLAT—The BLAST-Like Alignment Tool,» *Genome Research,* vol. 12, pp. 656-664, 2002.

[11] S. Schwartz, et al., «Human–Mouse Alignments with BLASTZ,» *Genome Research,* vol. 13, pp. 103-107, 2003.

[12] L. Batzoglou, J. Pachter, B. Mesirov, B. Berger, and E. S. Lander, «Human and mouse gene structure: comparative analysis and application to exon prediction,» in *RECOMB '00: Proc of the 4th Int'l Conference on Computational Molecular Biology,* 2000, pp. 46-53.

[13] M. Brudno and B. Morgenstern, «Fast and sensitive alignment of large genomic sequences,» in *Proc of IEEE Computer Science Bioinformatics Conference,* 2002, pp. 138-147.

[14] Bharath Reddy and Richard Fields, «Techniques for Reader-Writer Lock Synchronization,» International Journal of Electronics and Electrical Engineering, Vol. 8, No. 4, pp. 63-73, December 2020. doi: 10.18178/ijeee.8.4.63-73.

[15] Bharath Reddy and Richard Fields. 2022. From past to present: a comprehensive technical review of rule-based expert systems from 1980 -- 2021. In Proceedings of the 2022 ACM Southeast Conference (ACM SE '22). Association for Computing Machinery, New York, NY, USA, 167–172. https://doi.org/10.1145/3476883.3520211.

[16] Reddy, Bharath. (2020). BIOINFORMATICS & PAIRWISE SEQUENCE ALIGNMENT: Local and Global Sequence Alignment Algorithms. Amazon publications, May 2020, https://www.amazon.com/BIOINFORMATICS-PAIRWISE-SEQUENCE-ALIGNMENT-Algorithms-ebook/dp/B0879F5B5T/ref=sr_1_1?crid=2WEBAOL36MKNC&keywords=sequence+alignment+bharath&qid=1696737104&sprefix=sequence+alignment+bharath+%2Caps%2C295&sr=8-1.

[17] Reddy, B., Fields, R. (2022). Multiple Sequence Alignment Algorithms in Bioinformatics. In: Zhang, YD., Senjyu, T., So-In, C., Joshi, A. (eds) Smart Trends in Computing and Communications. Lecture Notes in Networks and Systems, vol 286. Springer, Singapore. https://doi.org/10.1007/978-981-16-4016-2_9.

[18] Holley RW, Apgar J, Everett GA, et al. Structure of a ribonucleic acid. *Science (80-).* 1965; 147 (3664): 1462-1465. doi: 10.1126/science.147.3664.1462.

[19] Fiers W, Contreras R, Duerinck F, et al. Complete nucleotide sequence of bacteriophage MS2 RNA: Primary and secondary structure of the replicase gene. *Nature*. 1976; 260 (5551): 500-507. doi: 10.1038/260500a0.

[20] Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*. 1977; 74 (12): 5463-5467. doi: 10.1073/pnas.74.12.5463.

[21] Smith LM, Sanders JZ, Kaiser RJ, et al. Fluorescence detection in automated DNA sequence analysis. *Nature*. 1986; 321 (6071): 674-679. doi: 10.1038/321674a0.

[22] Hood LE, Hunkapiller MW, Smith LM. Automated DNA sequencing and analysis of the human genome. *Genomics*. 1987; 1 (3): 201-212. doi: 10.1016/0888-7543 (87) 90046-2.
[23] Metzker, M.L. Emerging technologies in DNA sequencing. Genome Res. 2005, 15, 1767–1776.
[24] Kumar, K.R.; Cowley, M.J.; Davis, R.L. Next-Generation Sequencing and Emerging Technologies. Semin. Thromb. Hemost. 2019, 45, 661–673.
[25] Sakamoto, Y.; Sereewattanawoot, S.; Suzuki, A. A new era of long-read sequencing for cancer genomics. J. Hum. Genet. 2020, 65, 3–10.
[26] Goto, Y.; Akahori, R.; Yanagi, I.; Takeda, K.-I. Solid-state nanopores towards single-molecule DNA sequencing. J. Hum. Genet. 2020, 65, 69–77.
[27] Salk, J.J.; Schmitt, M.W.; Loeb, L.A. Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. Nat. Rev. Genet. 2018, 19, 269–285.
[28] Satam H, Joshi K, Mangrolia U, Waghoo S, Zaidi G, Rawool S, Thakare RP, Banday S, Mishra AK, Das G, et al. Next-Generation Sequencing Technology: Current Trends and Advancements. *Biology*. 2023; 12 (7): 997. https://doi.org/10.3390/biology12070997.
[29] Holley, R.W.; Apgar, J.; Everett, G.A.; Madison, J.T.; Marquisee, M.; Merrill, S.H.; Penswick, J.R.; Zamir, A. Structure of a Ribonucleic Acid. Science 1965, 147, 1462–1465.
[30] Heather, J.M.; Chain, B. The sequence of sequencers: The history of sequencing DNA. Genomics 2016, 107, 1–8.
[31] Sanger, F.; Nicklen, S.; Coulson, A.R. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 1977, 74, 5463–5467.
[32] Barba, M.; Czosnek, H.; Hadidi, A. Historical Perspective, Development and Applications of Next-Generation Sequencing in Plant Virology. Viruses 2013, 6, 106–136.
[33] Schuster, S.C. Next-generation sequencing transforms today's biology. Nat. Methods 2008, 5, 16–18.
[34] Hutchison, C.A. DNA sequencing: Bench to bedside and beyond. Nucleic Acids Res. 2007, 35, 6227–6237.
[35] Pervez, M.T.; Hasnain, M.J.U.; Abbas, S.H.; Moustafa, M.F.; Aslam, N.; Shah, S.S.M. A Comprehensive Review of Performance of Next-Generation Sequencing Platforms. *BioMed Res. Int.* 2022.
[36] Mantere, T.; Kersten, S.; Hoischen, A. Long-Read Sequencing Emerging in Medical Genetics. *Front. Genet.* 2019, *10,* 426.
[37] Costain, G.; Cohn, R.D.; Scherer, S.W.; Marshall, C.R. Genome sequencing as a diagnostic test. Can. Med. Assoc. J. 2021, 193, E1626–E1629.
[38] Logsdon, G.A.; Vollger, M.R.; Eichler, E.E. Long-read human genome sequencing and its applications. Nat. Rev. Genet. 2020, 21, 597–614.
[39] Rabbani, B.; Tekin, M.; Mahdieh, N. The promise of whole-exome sequencing in medical genetics. J. Hum. Genet. 2014, 59, 5–15.
[40] Iglesias, A.; Anyane-Yeboa, K.; Wynn, J.; Wilson, A.; Cho, M.T.; Guzman, E.; Sisson, R.; Egan, C.; Chung, W.K. The usefulness of whole-exome sequencing in routine clinical practice. Anesth. Analg. 2014, 16, 922–931.
[41] Williams, M.J.; Sottoriva, A.; Graham, T.A. Measuring Clonal Evolution in Cancer with Genomics. *Annu. Rev. Genom. Hum. Genet.* 2019, *20,* 309–329.
[42] Li, J.; Liu, C. Coding or Noncoding, the Converging Concepts of RNAs. Front. Genet. 2019, 10, 496.
[43] Lucchinetti, E.; Zaugg, M. RNA Sequencing. Anesthesiology 2020, 133, 976–978.
[44] Choi, S.-W.; Kim, H.-W.; Nam, J.-W. The small peptide world in long noncoding RNAs. Brief. Bioinform. 2019, 20, 1853–1864.
[45] Lasda, E.; Parker, R. Circular RNAs: Diversity of form and function. RNA 2014, 20, 1829–1842.
[46] Ura, H.; Togi, S.; Niida, Y. A comparison of mRNA sequencing (RNA-Seq) library preparation methods for transcriptome analysis. BMC Genom. 2022, 23, 303.
[47] Kolanowska, M.; Kubiak, A.; Jażdżewski, K.; Wójcicka, A. MicroRNA Analysis Using Next-Generation Sequencing. Methods Mol. Biol. 2018, 1823, 87–101.
[48] Wang, J.; Tian, T.; Li, X.; Zhang, Y. Noncoding RNAs Emerging as Drugs or Drug Targets: Their Chemical Modification, Bio-Conjugation and Intracellular Regulation. Molecules 2022, 27, 6717.
[49] López-Camarillo, C.; Gallardo-Rincón, D.; Álvarez-Sánchez, M.E.; Marchat, L.A. Pharmaco-epigenomics: On the Road of Translation Medicine. In Translational Research and Onco-Omics Applications in the Era of Cancer Personal Genomics; Springer: Berlin/Heidelberg, Germany, 2019; Volume 1168, pp. 31–42.
[50] National Human Genoe Research Institute. Epigenomics Fact Sheet. 2020. Available online: https://www.genome.gov/about-genomics/fact-sheets/Epigenomics-Fact-Sheet (accessed on 10 November 2023).
[51] Handy, D.E.; Castro, R.; Loscalzo, J. Epigenetic Modifications. Circulation 2011, 123, 2145–2156.
[52] Fuso, A. Aging and Disease. In Epigenetics in Human Disease; Academic Press: Cambridge, MA, USA, 2018; pp. 935–973.
[53] Qin, J.; Li, R.; Raes, J.; Arumugam, M.; Burgdorf, K.S.; Manichanh, C.; Nielsen, T.; Pons, N.; Levenez, F.; Yamada, T.; et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature 2010, 464, 59–65.

[54] Foster, J.A.; McVey Neufeld, K.-A. Gut–brain axis: How the microbiome influences anxiety and depression. Trends Neurosci. 2013, 36, 305–312.
[55] Illumina Inc. Ampliseq for Illumina, (n.d.). 2023. Available online: https://sapac.illumina.com/products/by-brand/ampliseq/community-panels.html (accessed on November 09, 2023).
[56] Bhoyar, R.C.; Jain, A.; Sehgal, P.; Divakar, M.K.; Sharma, D.; Imran, M.; Jolly, B.; Ranjan, G.; Rophina, M.; Sharma, S.; et al. High throughput detection and genetic epidemiology of SARS-CoV-2 using COVIDSeq next-generation sequencing. PLoS ONE 2021, 16, e0247115.
[57] Nagahashi, M.; Shimada, Y.; Ichikawa, H.; Kameyama, H.; Takabe, K.; Okuda, S.; Wakai, T. Next generation sequencing-based gene panel tests for the management of solid tumors. *Cancer Sci.* 2019, *110,* 6–15.
[58] Nagahashi, M.; Shimada, Y.; Ichikawa, H.; Kameyama, H.; Takabe, K.; Okuda, S.; Wakai, T. Next generation sequencing-based gene panel tests for the management of solid tumors. *Cancer Sci.* 2019, *110,* 6–15.
[59] Reddy, B., & Fields, R. (2023). Maximum Match Subsequence Alignment Algorithm Finely Grained (MMSAA FG). arXiv e-prints, arXiv-2305.
[60] B. Reddy and R. Fields, «Multiple Anchor Staged Alignment Algorithm – Sensitive (MASAA –S),» *2020 3rd International Conference on Information and Computer Technologies (ICICT),* San Jose, CA, USA, 2020, pp. 361-365. doi: 10.1109/ICICT50521.2020.00064.