

## Four Generations of Sequencing- Is it ready for the Clinic Yet?

Srilakshmi Srinivasan and Jyotsna Batra\*

Australian Prostate Cancer Centre- Queensland, Translational Research Institute, Queensland University of Technology, Woolloongabba, Australia

\*Corresponding author: Batra J, Australian Prostate Cancer Centre- Queensland, Translational Research Institute, Queensland University of Technology, Woolloongabba, Queensland, Australia Tel: +61 7 3443 7336; Fax: +61 7 34437336; E-mail: [jyotsna.batra@qut.edu.au](mailto:jyotsna.batra@qut.edu.au)

Rec date: ; Acc date: July 11, 2014; Pub date: July 17, 2014

Copyright: ©2014 Batra J, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### Abstract

Next-generation sequencing techniques have revolutionized over the last decade providing researchers with low cost, high-throughput alternatives compared to the traditional Sanger sequencing methods. These sequencing techniques have rapidly evolved from first-generation to fourth-generation with very broad applications such as unraveling the complexity of the genome, in terms of genetic variations, and having a high impact on the biological field. In this review, we discuss the transition of sequencing from the second-generation to the third- and fourth-generations, and describe some of their novel biological applications. With the advancement in technology, the earlier challenges of minimal size of the instrument, flexibility of throughput, ease of data analysis and short run times are being addressed. However, the need for prospective analysis and effectiveness to test whether the knowledge of any given new variants identified has an effect on clinical outcome may need improvement.

**Keywords:** NGS; TGS; Genome sequencing; Amplification; Read-length

### Abbreviations

NGS: Next-generation sequencing; SNPs: Single nucleotide polymorphisms; CNVs: Copy number variants; CLIP-seq: Cross-linking immunoprecipitation sequencing; ICGC: International Genome Consortium; TCGA: The Cancer Gene Atlas; TGS-3G: Third-generation sequencing; 4G: Fourth-generation sequencing; APS: Adenosine 5' phosphosulfate; PPI: Pyrophosphate; Mb: Mega bases; G-Giga SBS: Sequencing by synthesis; SNVs: Single nucleotide variants; MDA: Multiple displacement amplification; SLAF-seq: Specific-locus amplified fragment sequencing; SMS: Single-molecule sequencing; PGM: Personal genome machine; RRL: Reduced representation library; SMS: Single molecule sequencing; SMRTTM: Single molecule real-time; tSMS: True single molecule sequencing; ZMW: Zero-mode waveguide; FRET: Fluorescence resonance energy transfer; TEM: Transmission electron microscopy; Ms: Milliseconds; ELLIDA: Enzymatic luminometric inorganic pyrophosphate detection assay; FDA: Food and drug administration; CANCP: Solid Tumor Targeted Cancer Gene Panel by Next-Generation Sequencing; VUS: variants of unknown clinical significance; IVD: In vitro diagnostic; NIST: National Institute of Standards and Technology

### Introduction

The past decade has experienced advancement in the study of genetics, molecular diagnostics and personal medicine through the discovery and improvement of next-generation sequencing (NGS) techniques. Continuous improvement in NGS technologies imply, that the whole genome can be sequenced faster, easier and with a higher accuracy since the advent of Sanger sequencing in 1977. These new techniques are highly poised to disclose the complexity between genetic variants such as single nucleotide polymorphisms (SNPs), copy number variations (CNVs), structural variants, including gene fusions with diseases. Insights into epigenetics, protein-DNA interactions of

microbial genomes and various Mendelian disorders previously thought to be inaccessible, are now possible through these new high-throughput techniques.

NGS encompasses several technologies utilizing distinct sequencing biochemistry approaches and is mainly emphasized by its ability to simultaneously perform millions of sequencing reactions. The most widely used applications of NGS include targeted sequencing and whole transcriptome sequencing [1,2] while CLIP-seq (Cross-linking immunoprecipitation sequencing) a technique to map RNA binding sites for a protein of interest on a genome-wide scale is less frequently used. NGS enables worldwide collaborative efforts like the International Genome Consortium (ICGC) [3] and the Cancer Genome Atlas (TCGA) projects [4] to catalogue many thousands of cancer genomes for several disease types. Research discoveries from these projects have been published and have been non-trivial in improving our knowledge of disease pathogenesis thereby bridging the molecular pathology and personalized medicine [5]. These discoveries have the potential to benefit patients, clinicians and health-care systems to suggest changes of lifestyle behavior and/or in medical interventions to reduce disease related morbidity and mortality [6].

Although NGS have a wide range of biological outcomes, the costs per sample analyses often limit the use of these techniques. Fortunately, recently developed high-throughput techniques reduce the burden of the costs of sequencing. For example, sequencing costs have massively reduced from \$5,292.39/ Mb in 2001 to \$0.06/Mb by April 2013 [7]. It is estimated that the sequencing costs will further reduce with precipitous dropping per-base cost with advancing techniques.

In this review, we discuss common methods and challenges of NGS sequencing, and also their advancement from second-generation to the next two levels, third-generation (3G) and fourth-generation (4G). These techniques have allowed genomics to move from platforms that require PCR amplification of the template prior to sequencing, to single DNA molecule sequencing without a prior amplification step as in third-generation sequencing techniques, and to a more refined level of the fourth-generation.

## Second-Generation Sequencing

NGS techniques are quite diverse but conceptually similar. The preparation of library includes random shearing of DNA followed by ligation with common adaptors. The method used for determining template DNA sequence can be used to classify NGS techniques as follows:

### Cyclic array sequencing

This technique comprises the early methods of NGS techniques and includes pyrosequencing that relies on the detection of released pyrophosphates as they are incorporated during amplification. The Roche 454 GS20 sequencer introduced the first NGS platform using this technique. The library DNA's attached to 454-specific adaptors are denatured into single strands and captured by amplification beads followed by emulsion PCR [8]. On a picotiter plate, one of dNTPs (dATP, dCTP, dTTP, dGTP) complementary to the template are incorporated by ATP sulfurylase, luciferase, luciferin, DNA polymerase, and adenosine 5' phosphosulfate (APS) and pyrophosphate (PPi) equal to the number of incorporated nucleotides. The incorporated pyrophosphates can be traced by pyrogram™ or enzymatic luminometric inorganic pyrophosphate detection assay (ELLIDA), which corresponds to the order of correct nucleotides that has been incorporated [9]. The unincorporated nucleotides are degraded by apyrase (Figure 1). The read length initially for Roche 454 was 100-250 bp in 2005, generating 200,000+ reads and 20 Mb output per run. In 2008 the Roche 454 was replaced by the 454 GS FLX Titanium system which is capable of generating 700 megabase (Mb) of sequence in 700 bp reads in a 23 hr. run with an accuracy of 99.9% after filtering [10]. In 2009, Roche combined the GS Junior, a bench top system, into the 454 sequencing system and the output was upgraded to 14 gigabases (G) per run [11]. The specific advantage with these systems is the speed which only takes 10 hours to sequence. However, the technique has the disadvantage of having high reagent costs and high error rates for poly-bases longer than 6 bp (Table 1).

Method	Advantages	Disadvantages	Ref	Comments
Second-generation sequencing techniques				
454 sequencing	Generates long read lengths and relatively fast run times of the instrument	Poor interpretation of homopolymers leading to errors	[49]	First introduced NGS technique
Illumina (Solexa) Genome Analyzer	Short read length approach and is the most widely used analyzer	Aberrant incorporation of incorrect dNTPs by polymerases	[50]	Low multiplexing ability
HiSeq 2000 (Illumina, CA, USA)	Requires less sample < 1 µg	75 (35-100) bp read lengths. More false positives	[11]	Addition of fluorescent-labeled nucleotides
ABI SOLiD system	Reduction in error rates relative to Illumina NGS system	Have long run times and need for 2-20 µg DNA	[16]	Driven by DNA ligase than polymerase
Polonator 007	Decode the base by single-base probe in nonamers	In adequate coverage, false-positive SNP selection rate	[18]	Ligation based sequencer

Ion Torrent Sequencing	First platform to eliminate cost and complexity with 4-color optical detection used by other NGS platforms	High accuracy and short run time	[27]	Non-optical DNA sequencing
SLAF-seq	De novo SNP discovery with reduced cost and high accuracy	Needs complex instrument	[20]	Double barcode system ensures simultaneous genotyping of large populations
Third-generation sequencing techniques				
PacBio RS (Pacific Biosciences, CA, USA)	No amplification of template DNA required, real-time monitoring of nucleotide incorporation,	High error rates and low reads	[25]	Generates long-read lengths 800-1000 bp
Heliscope™ Sequencer	Nonbiased DNA sequence	High NTP incorporation error rates	[51]	Single molecule sequencing
Fourth-generation techniques				
Oxford Nanopore	Fastest sequencer whole-genome scan within 15 min	Not much data available, high cost per Mb	[32]	Expanding technique

**Table 1:** Summary of high-throughput sequencing methods

The cyclic array sequencing technique also involves fluorescent in situ sequencing by synthesis that can determine the template DNA sequence by detecting the exact nucleotide extended by its tagged fluorescent moiety as the sequencing proceeds. The sample preparation involves denaturing of the DNA libraries with adaptors to single strands, grafted to the flow-cell and bridge amplified to form millions of spatially separated clonal DNA fragments. Illumina (Solexa) Genome Analyzer adopts this technique and is the current dominating NGS technique in the market [12]. The first Solexa Genome Analyzer had an output of 1 G/run in 2006 and has gradually increased to 85 G/run within three years by the GAIIx series. The HiSeq 2000 released after 4 years of the first release of the Genome Analyzer and has an output of 600 G/run with an error rate of <2% in average after filtering. The HiSeq2000 is cheaper than the 454 and SOLiD systems, with an ability to handle thousands of samples simultaneously. MiSeq from Illumina also uses this sequencing by synthesis (SBS) technique that combines cluster generation, synthesis and data analysis in single equipment. It can complete sample preparation and analysis within a single day [11]. Efficient mapping of the short reads generated by these sequencers to the reference genome is challenging using traditional methods like BLAST [13] and BLAT [14]. Therefore, recent alignment algorithms such as BWA, which enables variant detection, and Bowtie, that can identify indels and are less prone to false-positive single nucleotide variants (SNVs) around indels, have been developed. These two algorithms and additional novel sequencing pipelines have emerged to help researchers in taking full advantage of NGS technologies [15]. Recently Qiagen (Venlo, Netherlands) have released

the whole genome amplification technique that can amplify complex genomes with less bias called the REPLI-g kit. REPLI-g uses Multiple Displacement Amplification (MDA) that involves binding of random hexamers and strand displacement synthesis by the enzyme Phi 29 polymerase. The Phi 29 polymerase can perform uniform amplification with minimized mutation rates and is compatible with many sequencing systems such as the Illumina HiSeq, Roche 454 and many other NGS instruments.

With the aim of commercial use of NGS in clinics, Qiagen is planning for commercialization of a new push-button automated "sample-to-insight workflow" to target clinical research and the diagnostics market in 2014. This new SBS sequencing system, Gene Reader bench-top sequencer integrates the Qia Cube instrument for DNA extraction, target selection, and sample preparation; Qia Cube NGS for library amplification preparation and bead enrichment; and the Gene Reader. The Gene Reader performs primary analysis - base calling, and generates a FASTQ file; secondary analysis to perform read alignment, variant calling using the CLC Bio software; and tertiary analysis - biological and clinical interpretation using the Ingenuity software. This high-grade integration would likely have a high impact on bridging the gaps between the laboratory based research and clinical application of NGS.

### Hybridization using oligonucleotide probes

This method originated from Jay Shendure and colleagues in 2005 [16], and is utilized by the ABI SOLiD system since 2006. Repeated cycles of hybridization with oligonucleotide probes allow sequencing of libraries on a SOLiD flow cell. These probes harbor a ligation site (1st base), a cleavage site (5th base), and 4 different fluorescent dyes linked to the last base [17]. The fluorescent signal is recorded as the complementary primer is hybridized to the template strand and the fluorescent signal is lost by the cleavage of the last 3 bases of the probe. The sequence of the DNA template can be deduced after 5 rounds of sequencing using ladder primer sets. The SOLiD system can be used for whole genome sequencing, transcriptome research, targeted sequencing, and epigenome sequencing. The latest SOLiD 5500x1 system has an output of 85 bp, and a high accuracy of 99.99% and 30 G/run.

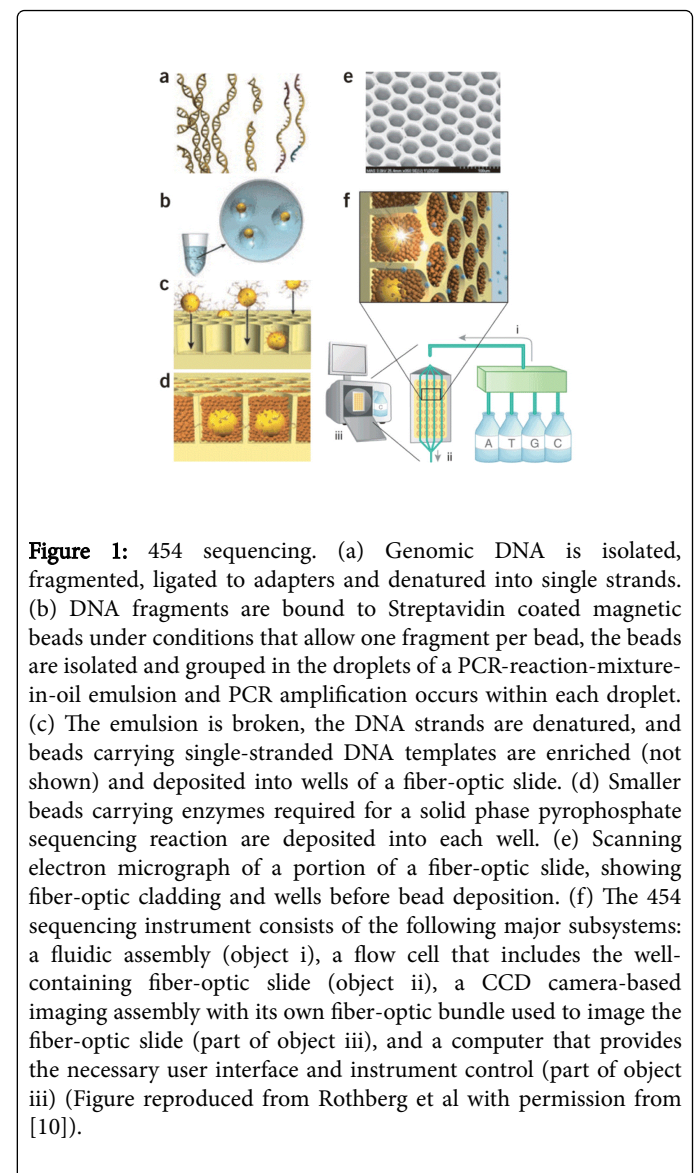
Another sequencer from complete genomics based on the Polonator G.007 is a ligation-based sequencer. The sequencer can decode the base by single-base probe in nanonucleotides (nanomers) [18]. The primary challenge of the sequencing by hybridization technique involves the probe design and the challenge to avoid cross-hybridization of the probe to the incorrect target due to repetitive elements. This may lead to inadequate coverage of the genome and false-positive SNP detection rates [19]. Also, the technique is time consuming, expensive and needs high infrastructure for analysis.

### Specific-locus amplified fragment sequencing (SLAF-seq)

SLAF-seq is a low cost but efficient high-throughput sequence based technique and reduces the complexity of high-quality reference

genome libraries that are required for other NGS techniques, and uses the strategy of a reduced representation library (RRL) method [20]. As the method does not require a reference genome sequence, it uses barcode multiplexed sequencing for multiple loci simultaneously, and combines locus-specific amplification and high-throughput sequencing for de novo SNP detection. The double barcode system distinguishes individuals in large populations of about 10,000 samples (Figure 1). A study testing the efficiency of the SLAF-seq on rice and soybean genomes have observed the genotyping data to be accurate and the density of the genetic map to be high compared to all the genome data available by other methods so far. Thus, SLAF-seq represents a low-cost large-scale genotyping technique with an important role in genetic association studies.

Although all the above 2G platforms may be suitable for clinical application, second-generation sequencing machines involve PCR amplification of sheared DNA fragments of about 35-400 bp, which are easily error-prone due to clonal amplification.

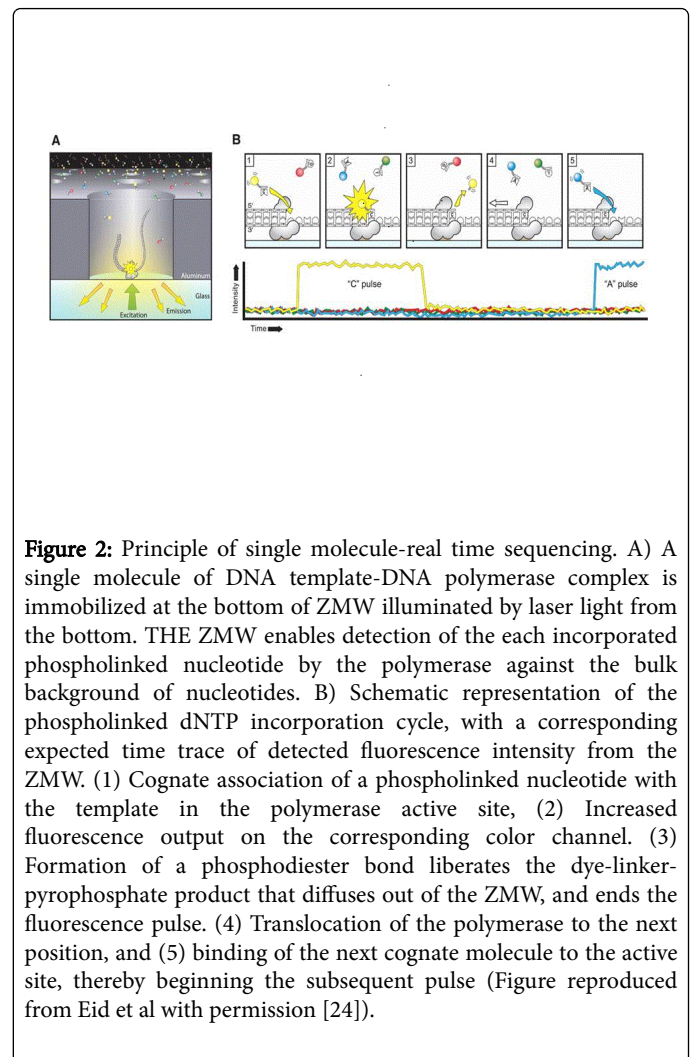


### Third-Generation Sequencing:

Despite the technological differences, the three categories of second-generation sequencing techniques have similar workflows for the sample preparation and analysis [21]. They also harness the need of an amplification step for template DNA as the NGS techniques are not designed to detect single fluorescent events [12]. PCR amplification is associated with PCR bias and has the possibility of base sequence errors or favoring certain sequences over the others. These biases can be avoided if a single molecule is used for sequencing without a prior amplification step. Also, the data generated by NGS techniques is massive, approximately 300 G bases or more by Illumina HiSeq 2000 instrument. The time-to-result is also long (many days) due to the large number of scanning and washing cycles required. Loss of synchronicity with addition of each nucleotide is another disadvantage of the technique and may lead to noise and second-generation sequencing errors [22] and short read lengths. The sequencing biochemistry, configuration and generation of array varies for second and third-generation sequencing techniques, few of which are discussed below.

#### Single-molecule sequencing (SMS)

SMS also known as single-template technology provides several advantages over second-generation sequencing. Two devices, Pacific Biosciences Single Molecule Real-time (SMRT) sequencing and the Helicos Biosciences true Single Molecule Sequencing (tSMS) were the providers of the first commercial 3G instruments. The techniques utilize the sequencing by synthesis approach, similar to a few of the NGS techniques but differ by not requiring amplification and hence, reduce the sequencing errors due to amplification as in NGS, reducing compositional bias [18] producing long sequences and supporting a short run-around time. The technique uses a DNA polymerase to drive the reaction and is based on real-time imaging of fluorescent labeled nucleotides as they are synthesized along the template DNA molecules [23]. This imaging is performed by dense array of zero-mode waveguide (ZMW) nanostructures that allow optical interrogation of single fluorescent molecules (Figure 2). Single functioning DNA polymerase is immobilized at the bottom of each ZMW to process fluorescently labeled nucleotide substrates. The four bases (G, C, T, A) are differentially labeled with deoxyribonucleoside pentaphosphate (DN5Ps) substrates. The fluorescent substrate is linked to the phosphate chain rather than the base and therefore the phosphate chain is cleaved when the nucleotide is incorporated into the DNA strand. Thus, on the incorporation of the phospholinked nucleotide, DNA polymerase frees the substrate molecule from the nucleotide when it cleaves the phosphate chain. The label is quickly removed and does not halt the DNA polymerase activity; which halts after incorporating few base-labeled nucleotides. This sequencing technique generated long read lengths of ~1000 bp in 2009. The recent PacBio®RSII sequencer (Pacific Biosciences, California) released in 2013 can generate 8.5 kb reads by combining the P5 DNA polymerase with C3 chemistry (P5-C3), with the longest reads exceeding 30,000 bases. This SMRT sequencer was reported to be least biased and good coverage in extreme GC content (both GC-rich and GC-poor) compared to Illumina and Ion Torrent sequencers [24]. Although the SMRT technique have many advantages over the NGS techniques, a number of challenges still remain such as excess of 5% error rates by insertions and deletions when assembling genomes [25].



**Figure 2:** Principle of single molecule-real time sequencing. A) A single molecule of DNA template-DNA polymerase complex is immobilized at the bottom of ZMW illuminated by laser light from the bottom. THE ZMW enables detection of the each incorporated phospholinked nucleotide by the polymerase against the bulk background of nucleotides. B) Schematic representation of the phospholinked dNTP incorporation cycle, with a corresponding expected time trace of detected fluorescence intensity from the ZMW. (1) Cognate association of a phospholinked nucleotide with the template in the polymerase active site, (2) Increased fluorescence output on the corresponding color channel. (3) Formation of a phosphodiester bond liberates the dye-linker-pyrophosphate product that diffuses out of the ZMW, and ends the fluorescence pulse. (4) Translocation of the polymerase to the next position, and (5) binding of the next cognate molecule to the active site, thereby beginning the subsequent pulse (Figure reproduced from Eid et al with permission [24]).

The tSMS technique is slightly different from the SMRT by using a fluorophore tagged DNA polymerase which in proximity to a nucleotide, tagged with an acceptor fluorophore, emits a fluorescence resonance energy transfer (FRET) signal. The fluorophore label is released after incorporation. These two 3G techniques, SMRT and tSMS have a series of advantages compared to the second-generation such as higher throughput, longer read lengths to enhance de novo assembly, direct detection of haplotype, whole-chromosome phasing, higher consensus accuracy of identifying rare variants, less sample requirement, making it useful for clinical application. The tSMS still retains many characteristics of second-generation sequencers such as the sequencing approach and chemistry, but the ability to perform direct RNA-sequencing signifies its clear improvement over second-generation techniques.

**Transmission electron microscopy (TEM) offered by Halcyon Molecular is another SMS based technique that images and chemically detects atoms comprising DNA templates [26]. Scanning tunneling microscopy can detect DNA bases according to specific electronic differences among the four bases. Although these ideas appear to be straightforward, it has yet to make a long journey due to its challenges to**

**prepare stretched single-stranded DNA on the surface, high cost of microscopes, and its requirement of more specific equipment.**

### Non-optical semi-conductor sequencing technique

The Ion Personal Genome Machine (PGM) launched by Ion torrent, utilizes the power of semiconductor technology to detect protons released as nucleotides are incorporated during synthesis. The system is unique from other techniques by its ability to detect amplification by measuring pH rather than light. The library preparation is similar to 454 sequencing [27]. Sheared DNA strands are linked to adapter sequences; a single DNA template is affixed to a bead (Ion Sphere Particle) and clonally amplified by emulsion PCR. The beads are loaded onto the chip and the dNTPs are flowed over the surface of these beads in a predetermined sequence with zero or more dNTPs ligating during each flow. The 454 sequencing system can introduce 4 nucleotides sequentially, whereas the ion torrent can include 32 nucleotides. This complex flow cycle referred to as Samba improves the synchronicity of clonal templates on the bead at the cost of a flow-sequence not optimized for read length. The protons released for every nucleotide-incorporated decrease the net pH in the surrounding solution that can be measured by an ionic sensor and then converted to a flow value. A base-caller corrects these flow-values for phase and signal loss, normalizes to the key and generates corrected base calls for each flow in each well to produce sequencing reads. Each read is sequentially passed between two signal-based filters to exclude less accurate reads. Per-base quality values are predicted by the Phred method [28] that quantifies the similarity between the phasing model predictions and the observed signal. The Ion PGM is the first system that does not require fluorescence and camera scanning and therefore enables higher output within less time compared to other NGS systems. This technique is also thought to be between the second and third-generation categories [29] because of its wash-and-scan technology. However, its output, speed and use for single-molecule analysis make it a third-generation sequencer.

### Fourth-Generation Sequencing

The 4G platforms evolved fast, combining single-molecule sequencing of 3G and integrating nanopore technology. Nanopore techniques achieve sequencing without requiring amplification, real-time sequencing without repeated cycles and elimination of synthesis, and so are classified as 4G sequencing techniques.

### Nanopore sequencing

Nanopore sequencing is based on the concept that single DNA molecules can be identified when passing through a tiny nanopore chamber. The nanopores are tiny biopores with nanoscale diameter and can be categorized into biological and solid-state [30].

The biological nanopore is formed by a pore forming protein in the membrane (lipid bilayer) and synthetic materials such as silicon nitride form the solid-state. Oxford Nanopore technology has recently released the commercial nanopore platform that can achieve long-read lengths and have sequenced a phage genome using this technique. The company predicts it can decipher a billion DNA bases in 6 h and sell for US \$900 [31]. A single protein nanopore is incorporated in a lipid bilayer across a microwell equipped with electrodes. Sample preparation, detection and analysis are performed in microwells incorporated onto an array chip. The nanopore sequencer has wider

applications for protein, RNA and DNA and is mainly adapted for DNA. The sequencing methodologies developed by Oxford include exonuclease and strand sequencing. A cyclodextrin adapter inside the protein nanopore serves as a DNA binding site, and the exonuclease coupled with the nanopore cleaves the individual bases from the DNA strand. The cleaved bases attached to the cyclodextrin are detected (20 ms/base) based on their differences in magnitude of current disruption. Single stranded DNA is passed through the pore in strand sequencing technique and is potentially faster and more accurate than exonuclease sequencing (Table 1). The solid-state nanopore is reliably more stable than biological and could be multiplexed to work in parallel on a single device. Many companies are now using the solid-state nanopore technique to perform whole genome sequencing and achieve higher readout within a short time [32].

Complete Genomics have also developed the nanoarray platform where genomic DNA is sonicated to prepare sequencing fragments and adapter sites are inserted followed by circularization of the template (400 bp) and restriction enzyme excision. The circularized templates are amplified by a Phi 29 polymerase and are called a DNA nanoball (DNB) [33]. Up to three billion DNBs are then selectively attached to a silicon chip. This helps in minimizing the reagent costs and increased throughput compared to NGS and earlier TGS instruments. Common probes with standard and extended anchors are hybridized to the DNB array chip and generate read lengths of 5 – 10 bases, resulting in 62 to 70 bases sequences per DNB. Although this method has a higher coverage of the genome compared to Sanger and NGS techniques, the size of the circularized fragments (400 bases), and very short read lengths (6 – 10 bases), prevents complete and accurate genome assembly. In June 2014, Roche announced its intention to purchase Genia, a single molecule tag-based nanopore NGS company. The technique is similar to Oxford Nanopore except that instead of extending the growing amplified DNA product, the Geniananopore allows a cleavable label to enter the pore. The labels induce a change in current through the nanopore, which is read by the instrument. Genia tag-based nanopore technique aims for \$100 genomes, and is aiming to release the first platform later this year.

### NGS in clinics

The concept of personalized medicine relies on our access to information on an individual's unique genetic makeup to administer customized therapy. Although the era of NGS has begun, traditional techniques such as PCR and Sanger sequencing are still used for detecting mutations for cancers and other infectious diseases. Few tests such as BRAC Analysis® [34], Melaris® [35] are examples. Initially karyotyping and fluorescent in situ hybridization were used to identify genetic abnormalities, and recently microarrays are successfully used to detect CNVs and SNPs in some Mendelian disorders [36], cancers [37] and common disorders. CNV microarrays suffer the limitation in their ability to assess only large aberrations and therefore cannot identify small genetic abnormalities as observed for many genetic disorders. The SNP microarrays with millions of probes can provide genotype information of common variants as the probes are built only for known information of the genome and cannot assess large aberrations. However, rare variants that may contribute to pathogenesis are often missed. Also, these classical methods do not allow multiplexing of many samples in one lane and targeted or comprehensive coverage of the genome.

The current NGS platforms address the some of the aforementioned limitations of traditional methods, and the clinical community is

rapidly embracing NGS techniques to allow treatment based on genetic fingerprints of an individual. NGS techniques use as little as a few picograms of DNA [38] compared to few micrograms for Sanger sequencing [39]. This offers a distinct advantage for use of NGS in the clinic where the amount of available sample is always limiting. In-depth coverage by advanced NGS techniques help in providing information on every nucleotide of the region of interest and therefore novel rare variants of clinical importance can be identified. For example, a rare missense variant not identified by SNP microarrays in the SLC26A3 gene helped in classification of a patient with congenital chloride diarrhea rather than the Bartter syndrome [40]. This example demonstrates the clinical utility of whole-exome sequencing in pinpointing the disease-causing genes in genomic regions identified by GWAS and thus helps in accurate clinical diagnosis.

Further, small bench-top NGS sequencers offer the best use as a clinical diagnostic device as they take less space. The Ion Proton™, a subsequent generation of Torrent sequencing technology is a bench-top machine for small sample sizes, and can perform whole exome and whole genome sequencing [41]. The current leading platforms include MiSeq from Illumina, and the PGM from Life Technologies which compromise >85% of market share earlier this year. Verinata Health, a non-invasive prenatal testing company leveraged the MiSeq NGS platform to test for fetal chromosomal aneuploidies in high risk pregnancies. This test performed better than standard screening methods for aneuploidies such as Down syndrome, Edwards syndrome, Patau syndrome and other common sex chromosome aneuploidies like Turner syndrome [42]. A bench-top machine from Qiagen, the Gene Reader™ is released this year and is utilized for clinical use and not just for the research setting.

As of late 2013, more than 50 NGS-based diagnostic tests including single-gene assays, multi-gene or multi-transcript panels are in use [43]. Some of the NGS panels employ targeted sequencing methods e.g. include tests available for non-syndromic deafness [44], cancer spinocerebellar ataxias or cardiomyopathies. In April 2014, the Mayo Clinic (<http://www.mayoclinic.org/>) launched a new gene panel cancer test called Solid Tumor Targeted Cancer Gene Panel by Next-Generation Sequencing (CANCP) which is a single assay performed that uses formalin-fixed paraffin-embedded tissue to assess common mutations in 50 genes known to affect tumor growth and response to chemotherapy. This test performed using Roche 454 sequencer can be useful for assessing prognosis and guiding treatment of individuals with solid tumors. The data is being used to help determine the clinical trial eligibility for patients with somatic mutations in genes not amenable to current Food and drug administration (FDA)-approved targeted therapies.

Illumina has released the first and only in vitro diagnostic (IVD) NGS platform for accurate and comprehensive cystic fibrosis testing. This system includes two assays: the Cystic Fibrosis 139-Variant Assay that detects 130 clinically relevant CFTR gene variants, and the cystic Fibrosis Clinical Sequencing Assay that accurately captures all variants in the protein coding regions and intron/exon boundaries of CFTR.

One of the major bottlenecks for the adoption of NGS into clinics includes the variability in the performance of variant calling softwares [45,46]. Recently, the National Institute of Standards and Technology (NIST) developed a highly confident variant caller by encouraging the NGS community to share sequencing data of their reference material. This will aid in the standardization of analysis methodologies and better quality checks for assessing false positives and negatives [47,48]. CLC Cancer Research Workbench launched in April 2014, offers

cancer-specific, ready-to-use analysis workflows that can be modified and personalized by clinicians and scientists. Ingenuity® Clinical, a new web-based solution is promising to deliver faster, easier-to-use, and high-confidence clinical interpretation of NGS-based tests.

Analysis of variants of unknown clinical significance (VUS) is another major challenge in clinical diagnosis as resources for functional testing on a patient basis is limited. Improving high-throughput models of disease using patient-derived samples such as tumor tissues can help to clarify the consequence of the VUS. The commercial success of NGS in the diagnostic era also needs to address other challenges beyond the technology itself, including regulatory approval (such as FDA), ethical concerns, insurance reimbursement, and the ease of analyzing large datasets.

An ethical issue arising from NGS techniques is whether or not to disclose incidental findings. For example, a patient undergoing tests for BRCA1 and 2 gene mutations for breast cancer may show mutations leading to Lynch disease. Whether or not disclosure of this information to that patient is beneficial or harmful has been a dispute. Advancing genomic technologies generate large amounts of sequence data, thus, questioning the right of privacy and confidentiality for an individual. This is because storing large data generated in databases may lead to accidental or unauthorized release of a patient's unique genomic information. Data sharing across national borders to researchers in other countries also need to be carefully performed as the flaws persist in the current policies as to the handling of biological data.

## Conclusion

Recently, high-throughput sequencing techniques have progressed vastly, and are in continuous development and improvement. They have a wider application and have a common feature of high-throughput data generation. Second-generation sequencing techniques generate massive output but still have limitations of requiring exclusive bioinformatics alignment software for analyses. Further advanced single-molecule sequencing (third-generation) techniques and fourth-generation sequencers have the advantages of accurate and long read lengths without homo-polymer tailing errors, and are moving forward fast to achieve the goal of the \$100 genome.

NGS-based diagnostics constitute the major part of personalized medicine and the clinical diagnostic industry. The current target enrichment techniques are more feasible in a clinical diagnostic context. While whole-genome sequencing can identify disease-related genes in clinically well-defined Mendelian diseases (biased for recessive disorders), the application of WGS for heterogenic more complex diseases like cancers need more improvement. Although, the application of NGS with its unique technology is encouraging, several technical and ethical challenges are yet to be addressed. For wider and easier application in clinics, further reduction of the cost of data storage, flexibility in throughput, data analysis, interpretation and reporting are needed. A bench-top sequencer such as the Gene Reader (Qiagen), which can be used with just a push-button option, could be more beneficial in a clinical setting for genomic and molecular diagnostics. Such systems would be easier and more amenable for translating research from “bench to bedside.”

We envisage that in 10 years, we may reach a point where analytic validity of sequencing technologies are high, with ease of access to clinical interpretation of genomic data and knowledge about patient responses to genomic information.

## References

- Mardis ER, Wilson RK (2009) Cancer genome sequencing: a review. *Hum Mol Genet* 18: 163-168.
- Bras JR, Guerreiro R, Hardy J (2012) Use of next-generation sequencing and other whole-genome strategies to dissect neurological disease. *Nat Rev Neurosci* 13:453-464.
- Guyer M, Hemsley FM, Jennings JL, Kerr D, Klatt P, et al (2010) International network of cancer genome projects. *Nature* 464 : 993-998  
[www.http://cancergenome.nih.gov/](http://cancergenome.nih.gov/)
- Harris TJ, McCormick F (2010) The molecular pathology of cancer. *Nat Rev Clin Oncol* 7:251-265.
- Mwenifumbo JC, Marra MA (2013) Cancer genome-sequencing study design. *Nat Rev Genet* 14: 321-332.
- Spurdle AB, Lakhani SR, Da Silva LM, Balleine RL, kConFab Investigators, et al (2010) Bayes analysis provides evidence of pathogenicity for the BRCA1 c.135-1G>T (IVS3-1) and BRCA2 c.7977-1G>C (IVS17-1) variants displaying in vitro splicing results of equivocal clinical significance. *Hum Mutat* 31: 1141-1145.
- Dressman D, Yan H, Traverso G, Kinzler KW, Vogelstein B, et al (2003) Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc Natl Acad Sci USA* 100: 8817-8822.
- Ronaghi M, Karamohamed S, Pettersson B, Uhlen M, Nyren P (1996) Real-time DNA sequencing using detection of pyrophosphate release. *Analytical biochemistry* 242: 84-89.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376-380.
- Liu L, Li Y, Li S, Hu N, He Y, et al (2012) Comparison of next-generation sequencing systems. *J Biomed Biotechnol* 2012:251364.
- Metzker ML (2010) Sequencing technologies - the next generation. *Nat Rev Genet* 11:31-46.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of molecular biology* 215: 403-410.
- Kent WJ (2002) BLAT--the BLAST-like alignment tool. *Genome research* 12: 656-664.
- Cox DG, Simard J, Sinnett D, Hamdi Y, Soucy P et al (2011) Common variants of the BRCA1 wild-type allele modify the risk of breast cancer in BRCA1 mutation carriers. *Hum Mol Genet* 20: 4732-4747.
- Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, et al (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 309:1728-1732.
- Mardis ER (2008) The impact of next-generation sequencing technology on genetics. *Trends Genet* 24: 133-141.
- Timp W, Mirsaidov UM, Wang D, Comer J, Aksimentiev A, et al (2010) Nanopore Sequencing: Electrical Measurements of the Code of Life. *IEEE Trans Nanotechnol* 9: 281-294.
- Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, et al (2001) Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* 294:1719-1723.
- Sun X, Liu D, Zhang X, Li W, Liu H, et al (2013) SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. *PLoS One*, 8: 58700.
- Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nat Biotechnol* 26: 1135-1145.
- Whiteford N, Skelly T, Curtis C, Ritchie ME, Löhr A, et al (2009) Swift: primary data analysis for the Illumina Solexa sequencing platform. *Bioinformatics* 25: 2194-2199.
- Eid J, Fehr A, Gray J, Gray J, Lyle J, et al (2009) Real-time DNA sequencing from single polymerase molecules. *Science* 323:133-138.
- Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, et al (2013) Characterizing and measuring bias in sequence data. *Genome Biol* 14: R51.
- Roberts RJ, Carneiro MO, Schatz MC (2013) The advantages of SMRT sequencing. *Genome Biol* 14: 405.
- Krivanek OL, Chisholm MF, Nicolosi V, Pennycook TJ, Corbin GJ, et al (2010) Atom-by-atom structural and chemical analysis by annular dark-field electron microscopy. *Nature* 464: 571-574.
- Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, et al (2011) An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475: 348-352.
- Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 8:186-194.
- Barzon L, Lavezzo E, Militello V, Toppo S, Palù G (2011) Applications of next-generation sequencing technologies to diagnostic virology. *Int J Mol Sci* 12: 7861-7884.
- Niedringhaus TP, Milanova D, Kerby MB, Snyder MP, Barron AE (2011) Landscape of next-generation sequencing technologies. *Anal Chem* 83: 4327-4341.
- Pennisi E (2012) Genome sequencing. Going solid-state. *Science* 336: 536.
- Thomas CS, Glassman MJ, Olsen (2011) BDSolid-state nanostructured materials from self-assembly of a globular protein-polymer diblock copolymer. *ACS Nano* 5: 5697-5707.
- Anderson JP, Reynolds BL, Baum K, Williams JG (2010) Fluorescent structural DNA nanoballs functionalized with phosphate-linked nucleotide triphosphates. *Nano Lett* 10: 788-792
- Frank TS, Manley SA, Olopade OI, Cummings S, Garber JE, et al (1998) Sequence analysis of BRCA1 and BRCA2: correlation of mutations with family history and ovarian cancer risk. *J Clin Oncol*, 16: 2417-2425.
- Glendening JM, Flores JF, Walker GJ, Stone S, Albino AP, et al (1995) Homozygous loss of the p15INK4B gene (and not the p16INK4 gene) during tumor progression in a sporadic melanoma patient. *Cancer Res* 55: 5531-5535.
- Park JJ, Oh BR, Kim JY, Park JA, Kim C, et al (2011) Copy number variation of beta-defensin genes in Behcet's disease. *Clin Exp Rheumatol* 29: S 20-23.
- Pandita A, Bayani J, Paderova J, Marrano P, Graham C, et al (2011) Integrated cytogenetic and high-resolution array CGH analysis of genomic alterations associated with MYCN amplification. *Cytogenet Genome Res* 134: 27-39.
- Parkinson NJ, Maslau S, Ferneyhough B, Zhang G, Gregory L, et al (2012) Preparation of high-quality next-generation sequencing libraries from picogram quantities of target DNA. *Genome Res* 22: 125-133.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.
- Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, et al (2009) Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci USA* 106: 19096-19101.
- Merriman B, Rothberg JM et al (2012) Progress in ion torrent semiconductor chip based sequencing. *Electrophoresis* 33: 3397-3417.
- Bianchi DW, Parker RL, Wentworth J, Madankumar R, Saffer C, et al (2014) DNA sequencing versus standard prenatal aneuploidy screening. *N Engl J Med* 370: 799-808.
- Rehm HL (2013) Disease-targeted sequencing: a cornerstone in the clinic. *Nat Rev Genet* 14: 295-300.
- Shearer AE, DeLuca AP, Hildebrand MS, Taylor KR, Gurrola J, et al (2010) Comprehensive genetic testing for hereditary hearing loss using massively parallel sequencing. *Proc Natl Acad Sci USA* 107: 21104-21109.
- Gullapalli RR, Desai KV, Santana-Santos L, Kant JA, Becich MJ (2012) Next generation sequencing in clinical medicine: Challenges and lessons for pathology and biomedical informatics. *J Pathol Inform* 3:40.
- Gullapalli RR, Lyons-Weiler M, Petrosko P, Dhir R, Becich MJ, et al (2012) Clinical integration of next-generation sequencing technology. *Clin Lab Med* 32: 585-599.
- Zook JM, Samarov D, McDaniel J, Sen SK, Salit M (2012) Synthetic spike-in standards improve run-specific systematic error analysis for DNA and RNA sequencing. *PLoS One* 7: e41356

- 
48. Zook JM, Chapman B, Wang J, Mittelman D, Hofmann O, et.al (2014) Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. *Nat Biotechnol*32: 246-251
49. Wang J, Lin M, Crenshaw A, Hutchinson A, Yeager M, et.al (2009) High-throughput single nucleotide polymorphism genotyping using nanofluidic Dynamic Arrays. *BMC genomics* 10: 561
50. Peters EJ, McLeod HL (2008) Ability of whole-genome SNP arrays to capture 'must have' pharmacogenomic variants. *Pharmacogenomics*9: 1573-1577.
51. Harris TD, Buzby PR, Babcock H, Beer E, Bowers J,et.al (2008) Single-molecule DNA sequencing of a viral genome. *Science*320:106-109.