“Next Generation Sequencing Methods and Its Impacts on Genomics and Clinical Applications”

BIOC-258, Genomics, Bioinformatics and Medicine

“Final Project”

by: Roxana Jalili

December 6, 2013

Introduction

DNA sequencing is one of the main concerns of medical research nowadays. Union of chain termination sequencing by Sanger *et al.* in 1977 and the polymerase chain reaction (PCR) by Mullis *et al.* in 1983 established many marked events such as the completion of the Human Genome Project, providing a barely sufficient reference to investigate the genetic alterations in the associated phenotypes [1].

DNA sequencing technologies ideally should be fast, accurate, easy-to-operate, and cheap [2]. The NGS technologies are different from the Sanger method in aspects of massively parallel analysis, high throughput, and reduced cost. Although NGS makes genome sequences handy, the followed data analysis and biological explanations are still the bottleneck in understanding genomes [2].

Following the human genome project, 454 was launched by 454 in 2005, and Solexa released Genome Analyzer the next year, followed by (Sequencing by Oligo Ligation Detection) SOLiD provided from Agencourt, which are three most typical massively parallel sequencing systems in the next- generation sequencing (NGS) that shared good performance on throughput, accuracy, and cost compared with Sanger sequencing. These founder companies were then purchased by other companies: in 2006 Agencourt was purchased by Applied Biosystems, and in 2007, 454 was purchased by Roche, while Solexa was purchased by Illumina [2]. After years of evolution, these three systems exhibit better performance and their own advantages in terms of read length, accuracy, applications, consumables, man power requirement and informatics infrastructure, and so forth [2]. In October 2013, Roche announced that 454 will be phased out in mid 2016 as in recent years, with the ascent of other sequencing technologies; the 454 instruments were pushed to the research margins only.

Roche 454 System

This sequencer uses pyrosequencing technology, which relies on the detection of pyrophosphate released during nucleotide incorporation. The library DNAs with 454-specific adaptors are denatured into single strand and captured by amplification beads followed by emulsion PCR [2]. Then on a picotiter plate, one of dNTP (ddATP, ddGTP, ddCTP, ddTTP) will complement to the bases of the template strand with the help of ATP sulfurylase, luciferase, luciferin, DNA polymerase, and adenosine 5′ phosphosulfate (APS) and release pyrophosphate (PPi) which equals the amount of incorporated nucleotide [2]. The ATP transformed from PPi drives the luciferin into oxyluciferin and generates visible light. At the same time, the unmatched bases are degraded by apyrase. Then another dNTP is added into the reaction system and the pyrosequencing reaction is repeated [2]. Figure 1 shows an outline of the sequencer workflow.

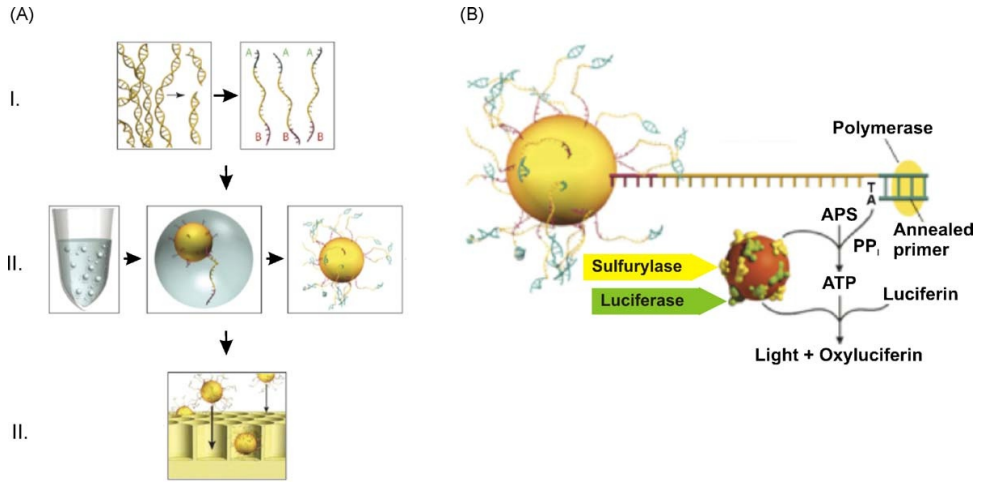

Figure 1. (A) 454 workflow. Library construction (I) ligates 454-specific adapters to DNA fragments (indicated as A and B) and couples amplification beads with DNA in an emulsion PCR to amplify fragments before sequencing (II). The beads are loaded into the picotiter plate (III). (B) Schematic illustration of the pyrosequencing reaction, which occurs on nucleotide incorporation to report sequencing-by-synthesis [3].

AB SOLiD System

AB sequencer adopts the technology of two-base sequencing based on ligation sequencing. On a SOLiD flowcell, the libraries can be sequenced by 8 base-probe ligation, which contains ligation site (the first base), cleavage site (the fifth base), and 4 different fluorescent dyes (linked to the last base) [2]. The fluorescent signal will be recorded during the probes complementary to the template strand and vanished by the cleavage of probes' last 3 bases. And the sequence of the fragment can be deduced after 5 round of sequencing using ladder primer sets [2]. Figure 2 shows a schematic of this technology.

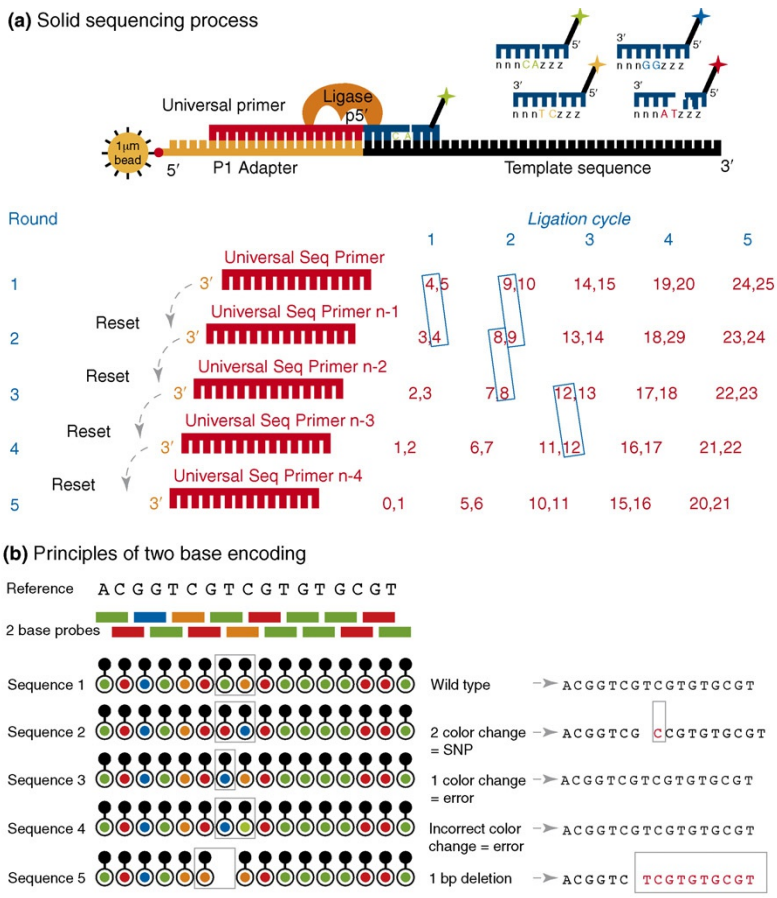

Figure 2. AB SOLiD sequencing. (a) AB SOLiD sequencing by ligation first anneals a universal sequencing primer then goes through subsequent ligation of the appropriate labeled 8mer, followed by detection at each cycle. (b) Two base encoding of the AB SOLiD data greatly facilitates the discrimination of base calling errors from true polymorphisms or indel events [4].

Illumina GA/HiSeq System

Illumina sequencer adopts the technology of sequencing by synthesis (SBS). The library with fixed adaptors is denatured to single strands and grafted to the flowcell, followed by bridge amplification to form clusters, which contains clonal DNA fragments [2]. Before sequencing, the library splices into single strands with the help of linearization enzyme, and then four kinds of nucleotides (ddATP, ddGTP, ddCTP, ddTTP) which contain different cleavable fluorescent dye and a removable blocking group would complement the template one base at a time, and the signal could be captured by a (charge-coupled device) CCD [2].

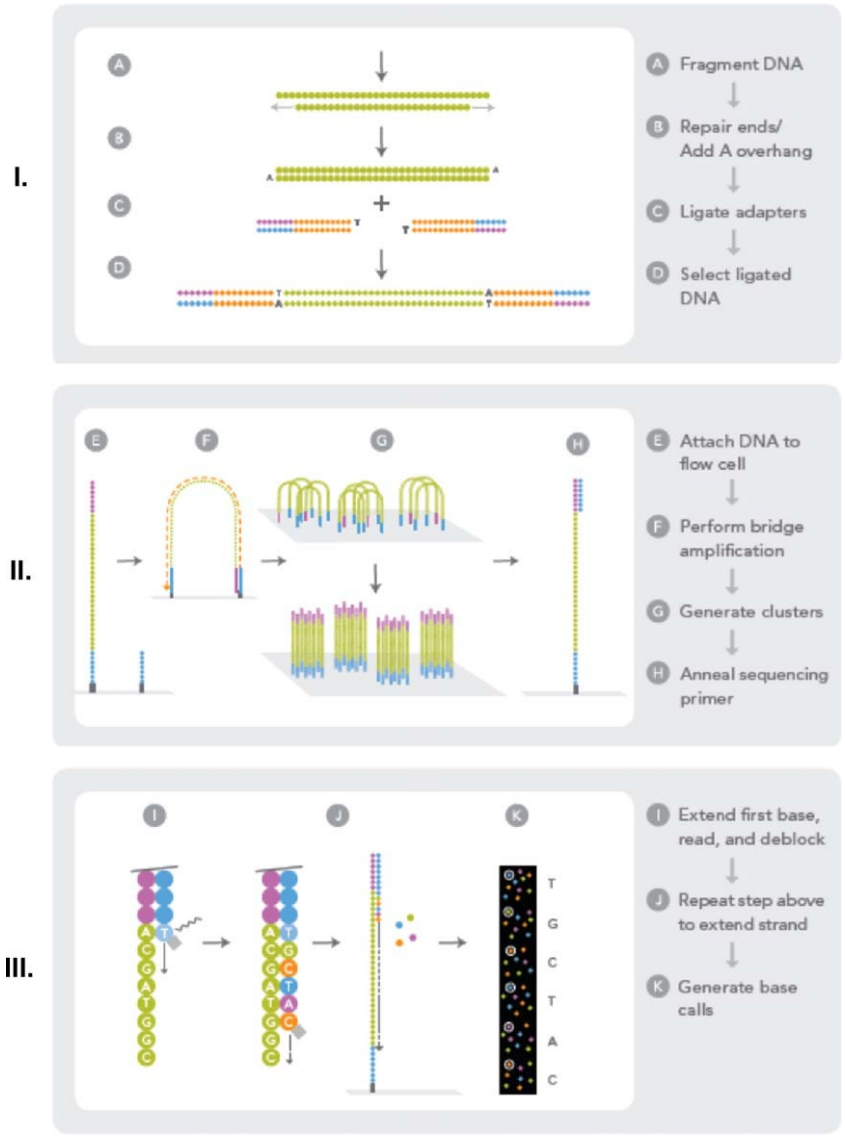

Figure 3. Outline of the Illumina Genome Analyzer workflow. (I) Similar fragmentation and adapter ligation steps take place, before applying the library onto the solid surface of a flow cell. (II) Attached DNA fragments form 'bridge' molecules, which are subsequently amplified via an isothermal amplification process, leading to a cluster of identical fragments that are subsequently denatured for sequencing primer annealing. (III) Amplified DNA fragments are subjected to sequencing-by-synthesis using 30 blocked labeled nucleotides [3].

Table 1 summarizes the advantages, mechanism, components, cost and application of three above discussed sequencers.

Table 1. (a) Advantage and mechanism of sequencers. (b) Components and cost of sequencers. (c) Application of sequencers [2].

(a)

Sequencer	454 GS FLX	HiSeq 2000	SOLiDv4	Sanger 3730xl
Sequencing mechanism	Pyrosequencing	Sequencing by synthesis	Ligation and two-base coding	Dideoxy chain termination
Read length	700 bp	50SE, 50PE, 101PE	50 + 35 bp or 50 + 50 bp	400~900 bp
Accuracy	99.9%*	98%, (100PE)	99.94% *raw data	99.999%
Reads	1 M	3 G	1200~1400 M	—
Output data/run	0.7 Gb	600 Gb	120 Gb	1.9~84 Kb
Time/run	24 Hours	3~10 Days	7 Days for SE 14 Days for PE	20 Mins~3 Hours
Advantage	Read length, fast	High throughput	Accuracy	High quality, long read length
Disadvantage	Error rate with polybase more than 6, high cost, low throughput	Short read assembly	Short read assembly	High cost low throughput

(b)

Sequencers	454 GS FLX	HiSeq 2000	SOLiDv4	3730xl
Instrument price	Instrument $500,000, $7000 per run	Instrument $690,000, $6000/(30x) human genome	Instrument $495,000, $15,000/100 Gb	Instrument $95,000, about $4 per 800 bp reaction
CPU	2* Intel Xeon X5675	2* Intel Xeon X5560	8* processor 2.0 GHz	Pentium IV 3.0 GHz
Memory	48 GB	48 GB	16 GB	1 GB
Hard disk	1.1 TB	3 TB	10 TB	280 GB
Automation in library preparation	Yes	Yes	Yes	No
Other required device	REM e system	cBot system	EZ beads system	No
Cost/million bases	$10	$0.07	$0.13	$2400

(c)

Sequencers	454 GS FLX	HiSeq 2000	SOLiDv4	3730xl
Resequencing		Yes	Yes	
De novo	Yes	Yes		Yes
Cancer	Yes	Yes	Yes	
Array	Yes	Yes	Yes	Yes
High GC sample	Yes	Yes	Yes	
Bacterial	Yes	Yes	Yes	
Large genome	Yes	Yes		
Mutation detection	Yes	Yes	Yes	Yes

As indicated in the above table, of the three NGS systems discussed, the Illumina HiSeq 2000 features the biggest output and lowest reagent cost, the SOLiD system has the highest accuracy, and the Roche 454 system has the longest read length [2].

Compact Personal Genome Machine (PGM) Sequencers

Ion (PGM) and MiSeq were launched by Ion Torrent and Illumina. They are both small in size and feature fast turnover rates but limited data throughput. They are targeted to clinical applications and small labs.

Ion PGM was released by Ion Torrent at the end of 2010. PGM uses semiconductor sequencing technology. When a nucleotide is incorporated into the DNA molecules by the polymerase, a proton is released [2]. By detecting the change in pH, PGM recognized whether the nucleotide is added or not. Each time the chip was flooded with one nucleotide after another, if it is not the correct nucleotide, no voltage will be found; if there are 2 nucleotides added, there is double voltage detected. PGM is the first commercial sequencing machine that does not require fluorescence and camera scanning, resulting in higher speed, lower cost, and smaller instrument size [2].

MiSeq, which uses sequencing by synthesis (SBS) technology was launched by Illumina. It integrates the functions of cluster generation, SBS, and data analysis in a single instrument and can go from sample to answer (analyzed data) within a single day (as few as 8 hours) [2]. The Nextera, TruSeq, and Illumina's reversible terminator-based sequencing by synthesis chemistry was used in this innovative engineering. The highest integrity data and broader range of application, including amplicon sequencing, clone checking, ChIP-Seq, and small genome sequencing, are the outstanding parts of MiSeq [2]. Table 2 compares the two PGM sequencers.

Table 2. Comparison between Ion Torrent and MiSeq [2].

	PGM	MiSeq
Output	10 MB–100 MB	120 MB–1.5 GB
Read length	~200 bp	Up to 2 × 150 bp
Sequencing time	2 hours for 1 × 200 bp	3 hours for 1 × 36 single read 27 hours for 2 × 150 bp pair end read
Sample preparation time	8 samples in parallel, less than 6 hrs	As fast as 2 hrs, with 15 minutes hand on time
Sequencing method	semiconductor technology with a simple sequencing chemistry	Sequencing by synthesis (SBS)
Potential for development	Various parameters (read length, cycle time, accuracy, etc.)	Limited factors, major concentrate in flowcell surface size, insert sizes, and how to pack cluster in tighter
Input amount	μg	Ng (Nextera)
Data analysis	Off instrument	On instrument

The Third Generation Sequencer

While the increasing usage and new modification in next generation sequencing, the third generation sequencing is coming out with new insight in the sequencing. Third-generation sequencing has two main characteristics. First, PCR is not needed before sequencing, which shortens DNA preparation time for sequencing. Second, the signal is captured in real time, which means that the signal, no matter whether it is fluorescent (Pacbio) or electric current (Nanopore), is monitored during the enzymatic reaction of adding nucleotide in the complementary strand [2].

Single-molecule real-time (SMRT) is the third-generation sequencing method developed by Pacific Bioscience, which made use of modified enzyme and direct observation of the enzymatic reaction in real time. SMRT cell consists of millions of zero-mode waveguides (ZMWs), embedded with only one set of enzymes and DNA template that can be detected during the whole process [2]. During the reaction, the enzyme will incorporate the nucleotide into the complementary strand and cleave off the fluorescent dye previously linked with the nucleotide. Then the camera inside the machine will capture signal in a movie format in real-time observation. This will give out not only the fluorescent signal but also the signal difference along time, which may be useful for the prediction of structural variance in the sequence, especially useful in epigenetic studies such as DNA methylation [2].

Nanopore sequencing is another method of the third generation sequencing. Nanopore is a tiny biopore with diameter in nanoscale, which can be found in protein channel embedded on lipid bilayer, which facilitates ion exchange. Because of the biological role of nanopore, any particle movement can disrupt the voltage across the channel [2]. The core concept of nanopore sequencing involves putting a thread of single-stranded DNA across α- haemolysin (αHL) pore. αHL, a 33 kD protein isolated from *Staphylococcus aureus*, undergoes self-assembly to form a heptameric transmembrane channel. It can tolerate extraordinary voltage up to 100 mV with current 100 pA. This unique property supports its role as building block of nanopore. In nanopore sequencing, an ionic flow is applied continuously. Current disruption is simply detected by standard electrophysiological technique [2]. Readout is relied on the size difference between all deoxyribonucleoside monophosphate (dNMP). Thus, for given dNMP, characteristic current modulation is shown for discrimination. Ionic current is resumed after trapped nucleotide entirely squeezing out [2]. Nanopore sequencing possesses a number of fruitful advantages over existing commercialized next-generation sequencing technologies. Firstly, it potentially reaches long read length >5 kbp with speed 1 bp/ns. Moreover, detection of bases is fluorescent tag-free. Thirdly, except the use of exonuclease for holding up ssDNA and nucleotide cleavage, involvement of enzyme is remarkably obviated in nanopore sequencing. This implies that nanopore sequencing is less sensitive to temperature throughout the sequencing reaction and reliable outcome can be maintained. Fourthly, instead of sequencing DNA during polymerization, single DNA strands are sequenced through nanopore by means of DNA strand depolymerization. Hence, hand-on time for sample preparation such as cloning and amplification steps can be shortened significantly [2].

Applications

The exponentially decreasing cost of next-generation sequencing data generation has put large-scale investigation of rare variation within reach, and there has been a resultant shift in the field of complex disease genetics over the past 5 years. GWAS data strongly suggest that the vast majority of the heritability of complex traits will not be due to a few common variants with low to moderate effects [5]. Rare variation with large effect sizes is likely contributing a significant proportion to the ‘‘missing heritability’’ of complex traits and disease. The common disease-common variant versus common disease-rare variant debate remains unresolved. There are still questions that remain as to whether the genetic contribution to common traits can be attributed to an infinite number of common alleles with

small effect, a large number of rare alleles with large effects, or some combination of genes and environment [5].

Variant Detection: The advent of NGS has enabled the inquiry of nearly every base in the genome, and thus techniques to reliably interpret and identify millions of variants are being developed. The advantage of sequencing in this regard is that most variants, common and rare, can be discovered with the appropriate sequencing read coverage, algorithmic methods to identify the variants, and a sufficient careful orthogonal validation to confirm true from false positives [5]. The exception to this discovery potential is due to the reliance on alignment to the Human Genome Reference sequence, which is the first step to analysis of NGS data, as this reference does not contain the entirety of novel genome content across all humans. Numerous variant calling algorithms have been developed for the detection and genotyping of germline and small indels in high-throughput sequencing data. Once detected, these variants can be analyzed in case-control studies using the same methods that have been developed for GWAS [5].

Rare Variation and Burden Testing: Unlike GWAS (which examines common mutations), sequencing facilitates the discovery of rare mutations that, combined with the continuing unexplained genetic contributions to complex phenotypes from GWAS, has sparked intense interest in measuring their association with complex phenotypes. In any single gene, there are a large number of rare variants due to recent human population growth, and there may be many non-associated variants in a gene. Furthermore, even in large cohorts, there may not be enough individuals with a given variant to achieve statistical significance [5].

Identifying De Novo Mutations: The rarest of variants are de novo mutations: those variants that arise first in an individual. They have tremendous relevance for disease biology, as they are more likely to have functional consequences in rare diseases. Characterizing these mutations also allows for the estimation of the baseline human mutation rate as well as its correlation to parental age. An entire class of computational tools has arisen that utilize both sequencing data and pedigree information to identify de novo mutations genome wide [5]. Most of these tools currently deal with trios (mother, father, and child) only and can identify de novo variants arising in the children. Because sequencing reads have a higher error rate than traditional genotyping, these tools incorporate information about coverage, the sequencing error rate, the expected de novo mutation rate, and family relationships. Although all of these tools identify potential de novo mutations, there remain significant feature differences between them, and no single tool has yet emerged as the frontrunner [5].

Studying Rare Mendelian Disorders: Rare monogenic disorders have provided unique opportunities to identify disease genes in humans. Traditionally, such disorders were studied by positional cloning or candidate gene approaches. Determining their molecular basis, however, was often hindered by small kindred sizes, genetic heterogeneity, and diagnostic classifications that may not reflect molecular pathogenesis. However, high-throughput sequencing of the full set of protein-coding genes (the exome) helps to overcome these obstacles by screening thousands of genes in a single experiment [5]. Although this limits the types of mutations that can be discovered, rare coding variants that are

predicted to have significant functional consequences can be discovered. In fact, it is estimated that, in ~60% of projects, exome sequencing will identify new Mendelian disease genes, and it is likely this approach also will contribute to complex disease genetics. Hence, the exome represents an enriched target space to identify rare variants with large effect sizes, as opposed to GWAS, wherein variants have low effect sizes [5].

Somatic Variant Detection: The comparison of an individual's cancer genome to the normal genome (derived from an unaffected tissue DNA) provides a comprehensive description of the somatic changes that have occurred in the transition from normal to cancerous cells. WGS approaches to somatic variant detection are more challenging due to the size of the data and the numerous types of variants that can be discovered by different algorithmic predictors, relative to exome sequencing [5]. However, structural variants, which are most difficult to predict accurately and with a reasonable false positive rate, occur frequently in cancer genomes and only can be discovered from WGS data. With an increasing focus on characterizing cancer heterogeneity, discussed below, the ability of somatic variant detection algorithms to predict low-frequency single-nucleotide variants (SNVs) in cancer cell populations is becoming critically important [5].

Rare Inherited Disorders: Although next-generation sequencing has impacted the human genetics field as a whole, few areas have benefited more than the study of rare genetic diseases. Some of the earliest applications of NGS to Mendelian disorders, as shown in Table 2, demonstrated that it was possible to identify disease-causing genes by sequencing the exomes of a few unrelated individuals or affected family members. Even the exome sequence of a single index case proved sufficient for genetic diagnosis for some disorders when information about the molecular underpinnings of the disease was known [5].

Table 2. OMIM Phenotypes for which the molecular basis is known, 2007 and 2013.

Inheritance Pattern	January 2007	July 2013
Autosomal	1,851	3,525
X Linked	169	277
Y Linked	2	4
Mitochondrial	26	28
Total	2,048	3,834

The impact of NGS technologies on rare genetic diseases is further evidenced by the growth of the Online Mendelian Inheritance in Man (OMIM) database, in which the number of inherited phenotypes for which the molecular basis is known has nearly doubled since 2007 (Table 2 above).

Family Studies of Complex Disease: There has been a return to family-based experimental designs for complex disease genetics recently, as it is expected that many members of the same family will carry a particular rare variant; hence, the number of individuals needed for rare variant discovery is much smaller than in cohorts of unrelated individuals. Using a combination of exome and WGS of affected individuals in consanguineous families, researchers can use homozygosity mapping to identify and characterize the variants contributing to genetically heterogeneous disorders [5].

De Novo Mutation Studies: Although genomic research in the past decade has largely emphasized inherited variation, NGS technologies also allow us to study, at base-pair resolution, the mutational processes that occur in humans from one generation to the next. Family based WGS studies have shown that each individual's genome harbors ~74 germline de novo mutations (DNMs). These mutations are potentially more deleterious because they have not been subject to natural selection and therefore are of considerable interest for sporadic diseases. Neurological and developmental disorders in particular highlight the impact of DNMs on disease risk [5].

Cancer Genomics Discovery: Over the past two years, the growth in cancer genomics discovery due to NGS is unprecedented, with multiple examples of large-scale WGS or WES based studies published in the literature for both adult and pediatric cancer types. The growth in our knowledge of the genes frequently mutated in cancer genomes is illustrated in Figure 4 [5].

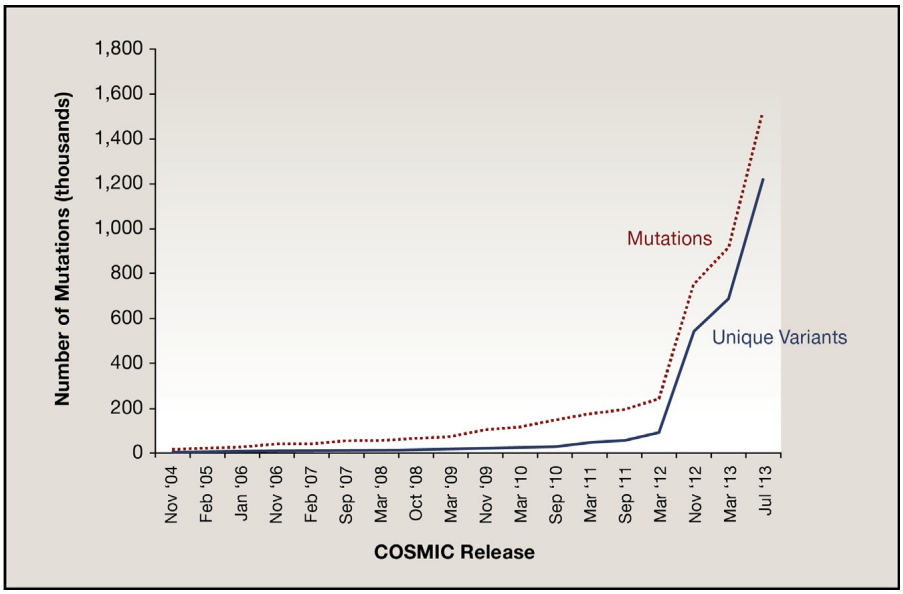

Figure 4. Growth in COSMIC Database Reports of Identified and Unique Mutations. Increases in the numbers of mutations and unique variants identified from DNA sequencing of cancer samples as cataloged in the COSMIC database, from November 2004 until the most recent release in July 2013. Note that the numbers of unique variants identified are increasing at a rate equal to the numbers of mutations discovered [5].

Cancer Genome Heterogeneity: For >100 years, the view of cancer cells through the pathologist's microscope has indicated that not all cancer cells in a tissue block are entirely similar. Several groups, using the digital nature of NGS data, now have proven this ''heterogeneity'' of cancer cells at the genomic level. Initially, genomic heterogeneity was demonstrated by copy number comparisons between primary and metastatic disease and by whole genome amplification and low-coverage sequencing of amplified genomic DNA from single breast cancer cells [5]. Within the past year, published studies using either WES or WGS have demonstrated the changes in genomic heterogeneity in cancers over the primary-to-relapse/metastatic transition or have characterized heterogeneity with primary tumor specimens. Specifically, these changes are determined by comparing the associated changes in the percentage of tumor cells carrying specific mutations detected by deep coverage NGS

data during disease progression. These studies evoke an evolutionary aspect to cancer's response to survival pressures, including therapy, and have fueled interest in better understanding the genomes of patients who are likely to recur in their disease [5].

Prediction of Targeted Therapy/Actionable Mutations: Since the earliest descriptions of specific mutations in *EGFR* predicting response to small-molecule inhibitors such as tyrosine kinase inhibitors, the association of somatic mutations to drug response has been of increasing interest. The use of NGS technologies in this regard has several advantages over the original methods (PCR and Sanger fluorescent sequencing) used to acquire these data [5]. Namely, the NGS-based inquiries required for discovering the gene-therapy association can be less hypothesis driven and examine all genes, the associated cost to generate the data for each patient sample is both less expensive and more rapidly obtained, and the ability to detect specific types of mutations such as insertions or deletions of one or several nucleotides is facilitated by NGS [5]. The first aspect is important because most small-molecule therapies target a range of mutated proteins, so multiple genes must be tested in each patient. The second aspect is important because these queries are now approaching clinical usage wherein identification of appropriate therapies must happen in a 2-3 week period to be applicable to patient care. Lastly, although small insertion/deletion mutations are rarer than single-nucleotide substitutions, their impact on the resulting protein may be more profound. Because Sanger sequencing typically fails to detect these variants, it is both likely that the frequency of these mutations is underestimated and certain that their response to therapy is less well understood as a result [5].

Circulating Tumor DNA Analysis: Many solid tumors shed cells and/or DNA into the blood stream at very low levels that are thought to fluctuate with increases or decreases in the disease burden of the patient. Hence, the ability to detect these changes with high sensitivity poses an interesting and potentially powerful disease-monitoring capability that likely would complement imaging modalities such as CT or MRI but at much lower cost and with lower associated [5]. In this regard, several groups have recently published manuscripts describing the selective capture of circulating tumor cells (CTCs) or the amplification and sequencing of circulating tumor DNA or RNA. This so-called ‘‘liquid biopsy’’ approach using plasma can detect the predominant somatic mutations for that tumor type, or if chromosomal translocations or structural variants already are known from prior characterization of the cancer genome, PCR primers can be designed to amplify the tumor- specific products for NGS and analysis [5].

Noninvasive Prenatal Testing: An NGS-based clinical assay that already has received widespread adoption is noninvasive prenatal testing for chromosomal abnormality diagnosis using samples such as maternal blood. In 1997, a group of scientist demonstrated that male sex could be determined from circulating fetal DNA in maternal plasma and serum samples and that the level of circulating fetal DNA increases with gestational age. However, achieving high sensitivity and specificity of fetal genotype was difficult, given the low levels of fetal DNA and the cost of high-depth sequencing [5]. With the advent of NGS, resolving the whole genome of a fetus from maternal blood sources became possible; however, major limitations likely will hinder clinical translation. For example, sequencing to sufficient depth to detect fetal DNA genotypes is still quite expensive. In addition, it is prohibitively

expensive and time consuming to routinely create and sequence maternal fosmid pools. As single-molecule sequencing technologies improve, it may be realistic to routinely resolve extended parental haplotypes to assist in fetal genotyping [5].

Conclusions

The significant cost reduction (shown in Figure 5 below) and high throughput of next generation DNA sequencing systems has resulted in generating sequence of huge numbers of different DNA strands simultaneously. These technologies are revolutionizing our understanding in molecular diagnostics, clinical interpretation and medical care, personalized medicine and nutrition as well as parental testing in near future.

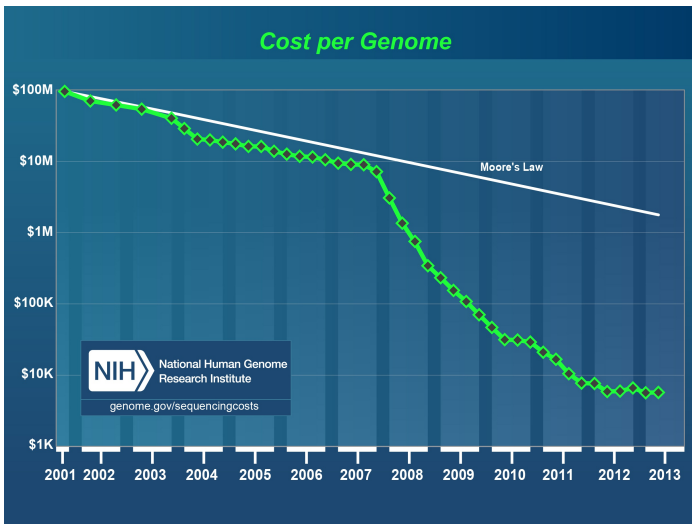

Figure 5. Cost per Genome: the cost of sequencing a human-sized genome [6].

Understanding the genetic variants (including the ability to accurately calling the functional variants, CNVs, chromosomal abnormalities) provides fascinating visions into the human disease for prevention approaches, diagnostic applications and therapeutic methods. Further sequencing cost reduction, improved read accuracy, and more importantly, computer-based analytics for correct data interpretation is required for incorporation of the technique into clinical diagnostic setting to improve life qualities.

References:

1. Bahareh Rabbani, Mustafa Tekin, and Nejat Mahdieh, "*The promise of whole-exome sequencing in medical genetics"*, Journal of Human Genetics, pp. 1-11 (2013).

2. Lin Liu, Yinhu Li, Siliang Li, Ni Hu, Yimin He, Ray Pong, Danni Lin, Lihua Lu, and Maggie Law, "*Comparison of Next-Generation Sequencing Systems"*, Journal of Biomedicine and Biotechnology, **2012**, pp. 1-11 (2012).

3. Wilhelm J. Ansorge, "*Next-generation DNA sequencing techniques"*, New Biotechnology, **25**, pp. 195-203 (2009).

4. Elaine R. Mardis, "*The impact of next-generation sequencing technology on genetics"*, Trends in Genetics, **24**, pp. 133-141 (2008).

5. Daniel C. Koboldt, Karyn Meltz Steinberg, David E. Larson, Richard K. Wilson, and Elaine R. Mardis, "*The Next-Generation Sequencing Revolution and Its Impact on Genomics"*, Cell, **155**, pp. 27-38 (2013).

6. National Human Genome Research Institute [Online] Available at http://www.genome.gov/SequencingCosts/ (accessed 27 November 2013)

7. Sian Ellard, George P. Patrinos, and William S. Oetting, "*Clinical Applications of Next-Generation Sequencing: The 2013 Human Genome Variation Society Scientific Meeting"*, HUMAN MUTATION, **34**, pp. 1583–1587 (2013).

8. Zhenqiang Su, Baitang Ning, Hong Fang, Huixiao Hong, Roger Perkins, Weida Tong and Leming Shi, "*Next-generation sequencing and its applications in molecular diagnostics"*, Expert Rev. Mol. Diagn., **11**, pp. 333-343 (2011).

9. Elaine R. Mardis, "*A decade's perspective on DNA sequencing technology"*, NATURE, **470**, pp. 198-203 (2011).

10. Peter M. Woollard, Nalini A.L. Mehta, Jessica J. Vamathevan, Stephanie Van Horn, Bhushan K. Bonde and David J. Dow, "*The application of next-generation sequencing technologies to drug discovery and development"*, Drug Discovery Today, **16**, pp. 512-519 (2011).

11. Elaine R Mardis, "*New strategies and emerging technologies for massively parallel sequencing: applications in medical research"*, Genome Medicine, **1**, pp. 1-4 (2009).

12. Wilhelm J. Ansorge, "*Next-generation DNA sequencing techniques"*, New Biotechnology, **25**, pp. 195-203 (2009).

13. Jorge S Reis-Filho, "*Next-generation sequencing"*, Breast Cancer Research, **11**, pp. 1-7 (2009).

14. Olena Morozova and Marco A. Marra, "*Applications of next-generation sequencing technologies in functional genomics"*, Genomics, **92**, pp. 255–264 (2008).

15. Jay Shendure & Hanlee Ji, "*Next-generation DNA sequencing"*, NATURE BIOTECHNOLOGY, **26**, pp. 1135-1145 (2008).

16. Elaine R. Mardis, "*Next-Generation DNA Sequencing Methods"*, Annu. Rev. Genomics Hum. Genet., **9**, pp. 387-402 (2008).