Platforms

DEBORAH JOSKO

LEARNING OBJECTIVES

1. Explain the Sanger method of DNA sequencing.
2. Discuss the difference between dNTPs and ddNTPs.
3. Describe the dye terminator and dye primer methods of sequencing.
4. Compare and contrast the Ion Torrent PGM™, MiSeq Illumina, and the SMRT Technology with regard to methodology, cost and turnaround time.

ABBREVIATIONS: ATP - adenosine triphosphate; dATP - deoxyadenosine triphosphate; dCTP - deoxycytidine triphosphate; ddNTP - dideoxynucleotide triphosphate ; dGTP - deoxyguanosine triphosphate; DNA - deoxyribonucleic acid; dNTP - deoxynucleotide triphosphate; DOE - Department of Energy; dTTP - deoxythymidine triphosphate; emPCR - emulsion Polymerase Chain Reaction; GS - genome sequencer; ISPs - Ion sphere particles; NGS – next generation sequencing; NIH - National Institutes of Health; OH - hydroxyl; PCR - Polymerase Chain Reaction; PGM - Personal Genome Machine; ^{32}P - Phosphorous 32 (isotope); SMRT - single molecule real time; ZMW - zero mode waveguides.

INDEX TERMS: Sequencing; DNA Sequencing, Sanger Method, Dideoxynucleotides; Ion Torrent PCM™, Illumina MiSeq; SMRT Technology.

Clin Lab Sci 2014;27(3):173

Deborah Josko, Ph.D. MLT(ASCP)M, SM, *Associate Professor and Program Director, Department of Clinical Laboratory Sciences, Medical Laboratory Science Program, Rutgers, The State University of New Jersey – School of Health Related Professions, Scotch Plains, NJ.*

Address for correspondence: Deborah Josko, Ph.D. MLT(ASCP)M, SM, *Associate Professor, Department of Clinical Laboratory Sciences, Medical Laboratory Science Program, Rutgers, The State University of New Jersey – School of Health Related Professions 1776 Raritan Road, Scotch Plains, NJ 07076. (908) 889-2422. daj100 @shrp.rutgers.edu.*

Introduction

Since the discovery of the DNA molecule by James Watson and Francis Crick in 1953, the field of molecular biology and genomics has escalated substantially especially in the last decade as a result of the completion of the Human Genome Project.[1,2] In 1990, the National Institutes of Health (NIH) and the Department of Energy (DOE) formulated and published a 5 year plan to begin sequencing the entire human genome.[1] This international undertaking which was estimated to last 15 years was completed a head of schedule in 2003 due to advances in deoxyribonucleic acid (DNA) sequencing techniques.[3,4]

DNA is a double helical structure that contains the sugar deoxyribose (a five carbon sugar which lacks an oxygen at the 2′ position hence the term “deoxy”); a phosphate group; and four deoxynucleotide triphosphates (dNTPs): deoxyadenosine triphosphate (dATP), deoxycytidine triphosphate (dCTP), deoxyguanosine triphosphate (dGTP) and deoxythymidine triphosphate (dTTP).[5,6] The sugar phosphate backbone when covalently linked to a nucleoside is referred to as a nucleotide.[6] Due to the chemical composition of the dNTPs, adenine always binds to thymine while cytosine always binds to guanine. DNA contains the genetic code necessary for an organism to reproduce and survive and can generate an exact copy of itself through the process known as DNA synthesis.

DNA synthesis requires the following components: DNA template; DNA primer; DNA polymerase which adds the nucleotides to the growing chain; four deoxynucleoside triphosphates: dATP, dTTP, dCTP and dGTP; and an energy source adenosine triphosphate (ATP) to catalyze the reaction.[5,7] During

the elongation phase of DNA synthesis, the free 3′ hydroxyl (OH) group from the growing strand attaches to the phosphate group of the next base forming a phosphodiester bond.[7] Nucleotides are added to the growing strand until the process is complete and termination occurs.

In 1975, utilizing this knowledge of DNA synthesis, Dr. Frederick Sanger developed the first DNA sequencing technique known today as the Sanger method. For his discovery and contribution to science and medicine, he was awarded the Nobel Prize in Chemistry in 1980.[8] In the Sanger method of sequencing, the following components are utilized: DNA template, DNA primer; DNA polymerase, dNTPs and a small quantity of dideoxynucleotide triphosphates (ddNTPs). The difference between dNTPs and ddNTPs is that the ddNTPs lack the 3′ OH group essential for chain elongation during DNA synthesis (ddNTPs lack two OH groups, one at the 2′ position (deoxynucleotide) and one at the 3′ position (dideoxynucleotide)).[9,10] Therefore, without the 3′ OH group necessary for chain elongation, chain termination occurs as a result when a ddNTP is incorporated into the growing DNA strand.[9,10]

The original Sanger method required that the reaction be performed in four separate tubes. For example, the first tube contains the DNA template, DNA primer, DNA polymerase, an ample supply of dATP and a small amount of ddATP. The second tube contains the DNA template, DNA primer, DNA polymerase, excess amounts of dTTP and a small amount of ddTTPs, and so on for dGTP and dCTP.[9,10] In the first tube, if a dATP is added during the synthesis cycle, the chain will continue to grow. However, if a ddATP is incorporated into the growing chain by DNA polymerase, chain termination occurs.[9] Due to the excess amounts of dNTPs in each tube and the small volume of ddNTPs, ddNTPs will be randomly incorporated into every nucleotide position.[9,10] One dNTP is labeled with a radioactive tag ^{32}P and used in the detection process by autoradiography. After multiple cycles, various size DNA fragments are generated and then separated by size using gel electrophoresis and read by autoradiography from the smallest to the largest fragment.[9,10] This technique however was very labor intensive and the gels were difficult to interpret. Refer to Figure 1 for an illustration of an autoradiography gel with hundreds of DNA bands.[11] Fortunately, the detection process was automated utilizing capillary electrophoresis which made interpretation easier. This technique was introduced in the early 1980s by Jorgensen and Lukacs.[12]

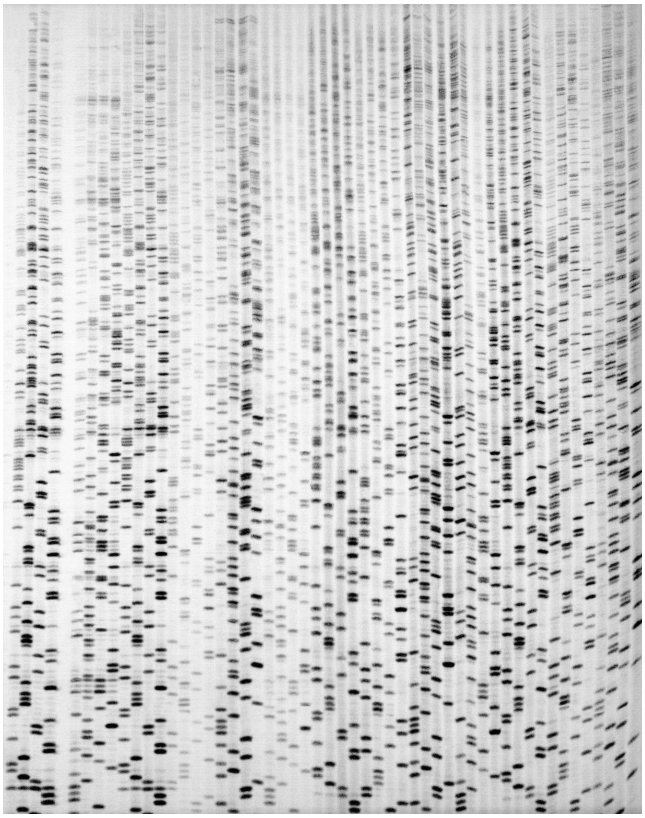

Figure 1. Autoradiography of Dideoxy Sequencing Gel. Image used with Permission by Cancer Research Technology, Wellcome Images.[11] Autoradiogram of a dideoxy sequencing gel using the Sanger Method. Four 35S-labelled sequence reactions are prepared for each sample; one for each base. There are ten reactions on this gel, each with approximately 250 bases, for a total of 2500 bases. The sequence is read from the bottom up of each lane. The gel is difficult to read, the process is laborious, and skill is required to interpret the resulting bands. Each gel and exposure takes approximately three days. This technique has been automated using capillary electrophoresis which aids in the interpretation of the sequence.[11,12]

One modification made to this technique is the dye terminator method.[13] This reaction which utilizes the same components: DNA template to be sequenced; DNA primer; DNA polymerase; excess amounts of dNTPs; and small volumes of ddNTPs can be carried out in one tube since each ddNTP is labeled with a

different colored fluorescent dye at the 3′ end.[13] After the cycle is complete, a small volume from the mix is loaded in one lane and placed in the automated sequencer. The sample undergoes electrophoresis and a computer generated image of the sequence is rendered.[13] Modern day sequencing techniques utilize lasers and automated instrumentation to aid in interpretation.

Another method that is a variation of the Sanger method is the dye primer method which utilizes the same components and occurs in four separate tubes as in the original Sanger method. However, the DNA primers are labeled with four different fluorescent tags instead of the ddNTPs. This method utilizes automated instruments in the detection process allowing for easier interpretation of bands.[14]

In summary, the various DNA sequencing techniques and platforms available today for both clinical and research use are either a modification or variation of the Sanger method of DNA sequencing or a completely different approach to DNA sequencing. The Sanger method is considered "first-generation" technology; however, the newer methods that have a higher output and are more cost effective are referred to as "next generation sequencing."[15]

Next Generation Sequencing Platforms:

Advances in genomic and genetic research are largely due to the development of NGS platforms. Three of the most common sequencers on the market today are the Ion Torrent PGM™ by Life Technologies,[16] Illumina MiSeq by Innovative Technologies,[17] and the SMRT Technology (single molecule, real time) by Pacific Biosciences[18] - all released in 2011.[19] Before that time, the Illumina HiSeq was the frontrunner in the market.[19] Other platforms available include the SOLID™ marketed by Life Technologies and AB Applied Biosystems[20] and the 454 GS Junior by Roche Diagnostics Company.[21] A brief description of the work flow, sample preparation, output, turnaround time and cost for a select few are discussed.

Ion Torrent PGM™

NGS platforms have the ability to sequence the entire genome or targeted areas/sequences. If a small region in the genome is of interest (from a few to several hundred regions), the Ion Torrent Personal Genome Machine(PGM)™ System by Life Technologies is the best platform to use.[22] Since this system is one of the most widely used systems on the market today, a detailed description of the principle is presented.

The system utilizes a semiconductor chip that is different from other sequencing principles. This chip allows for faster, more reliable output in a fraction of the time (approximately 3 hours) compared to other sequencers.[23] The use of a semiconductor chip allows chemical signals to be converted into digital information.[24] The workflow begins with the generation of a library of DNA.[25] DNA is extracted from the sample and either fragmented through sonication or an enzymatic shearing method. The size of the fragments depend on the how long the sample is sonicated or by the various restriction enzymes used.[26] Either procedure allows the DNA to be digested into fragments resulting in blunt ends.[26] For the Ion Torrent PGM™, DNA fragments should be in the range of 200 to 400 base pairs.[24,25,26] Once the DNA fragment library is created, the fragments undergo an initial amplification step using polymerase chain reaction (PCR). At this time bar coded adapters are added to the master mix (A and P1 adaptors) which will ligate to the DNA fragments.[24,25,26] The amplified products are then mixed with the Ion Sphere™ particles (ISPs) that contain specific primers attached to the entire particle. The P1 adapter on the DNA fragment will attach to the ISPs while the A1 adapter on the other end is used during the sequencing process.[24,25,26]

The DNA library fragments are further amplified onto the ISPs via emulsion PCR (emPCR). After emPCR is complete, the ISPs undergo an enrichment step to harvest the template ISPs.[24,25,26] At the end of this phase, the ultimate goal is to have one ISP and one strand of DNA that has been clonally amplified over the entire sphere.[26] Enriched ISPs are then loaded on the semiconductor chip. During the sequencing phase, different nucleotides are washed over the chip at a rate of one nucleotide every 15 seconds. [24,25,26] If a nucleotide was incorporated into the single stranded DNA, a hydrogen ion is released which changes the pH of solution in the well. An ion sensitive layer measures the change in pH and converts it to voltage. Thus, each well works as a small pH meter. [24,25] This is how the Ion Torrent sequences DNA by reading the chemical change generated directly on chip. If a nucleotide was not added, a hydrogen ion is not released, no pH

change is detected and no voltage is recorded. [24,25,26] If two of the same nucleotides are added to the chain, for example, two adenines were added, the voltage reading doubles. Once the run is complete, the data is processed and analyzed.[24,25,26]

Life Technologies markets three semiconductor chips that are utilized on the Ion Torrent PGM™: the 314™ chip contains 1.2 million wells; the 316™ chip contains 6 million wells; and the 318™ chip contains 12 million wells.[26,27] According to the manufacture's website, "with as little as 10 ng input DNA, the new Ion 314™, 316™, and 318™ Chips yield 99.99% consensus accuracy and up to a 50% increase in output over previous chips, in as little as 2 hours on the Ion PGM™ System."[28] The size of the DNA fragment and what is being sequenced dictates the size chip that should be used. Due to the bar coded adapters, individual patient samples can be run and detected simultaneously allowing for parallel sequencing to occur on one chip.

The actual sequencing step can take from two to three hours depending on the run and the size of the chip.[28] Preparing the DNA fragment library and performing both PCR reactions can require 4 hours to 1 day. Therefore, most users prefer to prepare their DNA library one day and perform the sequencing on a subsequent day.[26] Although pricing varies, the Ion Torrent PGM™ instrument costs around $80,000.[19] Another system that will be available soon by Life Technologies is the Ion Chef™ System which can be used with the Ion Torrent PGM™ and Ion Torrent Proton™ Systems.[29] This instrument is a fully automated, walk away system which has the capability to prepare template from DNA libraries as well as perform chip loading providing sequence ready chips to be sequenced. The entire set up takes less than 15 minutes.[29] This allows for lower costs, faster speed, and high throughput.[29] The applications for this instrument will be discussed in the Focus article entitled "Next Generation Sequencing - Applications."

Illumina MiSeq

The Illumina MiSeq platform is a fully automated system that incorporates DNA amplification, sequencing and data assessment in one instrument.[30] This platform, which utilizes the same chemistries as the HiSeq system, has been made smaller to fit better in research and clinical laboratories, is less expensive, more user friendly, accurate, but most importantly generates better quality data compared to older versions of the HiSeq.[31] This instrument is suitable when sequencing small genomes including bacterial genomes, performing targeted gene expression analysis, RNA sequencing, resequencing, etc.[30] The applications will be presented in the Focus article entitled "Next Generation Sequencing - Applications."

The workflow is fully automated and involves three steps: library preparation, cluster generation, and sequencing.[30,31,32] In the library preparation, DNA is fragmented, cut ends are repaired and adenylated, and adapters are added to both ends of the fragment.[32] These fragments are then size selected and purified.[32] In the cluster generation step, single molecules are isothermally amplified in a flow cell ultimately generating hundreds of millions of unique clusters.[32] There is no emulsion PCR required that was used in the Ion Torrent PGM™ system.[30] Reverse strands are cleaved and washed away and sequencing primers are hybridized to the DNA template and ready for sequencing.[30,32]

The sequencing step uses four fluorescently labeled reversibly terminated nucleotides.[30,32] All four bases compete with each other to bind to the template ensuring the highest accuracy in chain elongation. After each round, the clusters are excited by a laser emitting color that identifies the newly added base. The fluorescent label and blocking groups are then removed so the next base can be incorporated. Bases are read after each cycle to ensure accurate sequencing.[30,32]

This system is user friendly and offers content specific help for any run. The instrument instructs the user on which reagents to load for a particular run and will make sure all aspects of the instrument are working by making pre run checks.[31] However, the most exciting feature is the ability to interface with any smart phone, iPad, iPhone, etc. There is an application that can be downloaded called SeqMonitor.[31] The user can access performance runs, monitor estimated output, and quality of run, etc. The data can not only be accessed from a computer but can be placed in the iCloud (secure). This will create a community of NGS users that can store or share data and utilize applications developed in the iCloud. Information can be at ones fingertips no matter where they are located.[31]

Once the DNA is prepared (library and cluster generation steps) the turnaround time from DNA to results is approximately 8 hours (1.5 hours prep; 4 hours sequencing; 3 hours analysis).[33] Costs tend to vary; however, the Illumina MiSeq is priced at around $130,000.[19]

Pacific Biosciences SMRT Technology

Pacific Biosciences offers sequencing technology that differs from both the Ion Torrent PGM™ and the Illumina MiSeq. According to the manufacturer, "The PacBio RS II sequencing technology resolves single molecules in real time, allowing observation of structural and cell type variation not accessible with other technologies."[34] This technology utilizes a single molecule, real time (SMRT) cell that contains hundreds of thousands of zero mode waveguides (ZMW) very similar to wells on a microchip.[35] A DNA template and polymerase complex is bound to the bottom of the ZMW.[36,37] Phospholinked nucleotides, which have a different color florescent label, are attached to the terminal phosphate of each nucleotide instead of at the base. These nucleotides are then introduced into the ZMW chamber. As the DNA polymerase comes across the correct nucleotide, it takes a few milliseconds to incorporate it in the growing chain.[36,37] At this time the fluorescent label is exited emitting light which is captured by a sensitive detector and recorded. The label is then cleaved off and diffuses away leaving a natural strand of DNA.[36,37] This step is repeated until the process is complete. This technology utilizes the properties of DNA polymerase: high speed (10 or more bases per second), high fidelity, and long read length.[37] No amplification steps are required with run times as little as 30 minutes (30 to 180 minutes per SMRT cell).[38] The entire process including library preparation, sequencing and data analysis takes approximately 10 hours.[38] Costs can vary but the approximate fee of this instrument is $695.000.[19] For additional information and an in depth overview and comparison of all three sequencers described above refer to the article entitled, "A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers" by Quail et al. in 2012.[19]

Summary

The advent of DNA sequencing technologies and the various applications that can be performed will have a dramatic effect on medicine and healthcare in the near future. There are several DNA sequencing platforms available on the market for research and clinical use. Based on the medical laboratory scientist or researcher's needs and taking into consideration laboratory space and budget, one can chose which platform will be beneficial to their institution and their patient population. Although some of the instrument costs seem high, diagnosing a patient quickly and accurately will save hospitals money with fewer hospital stays and targeted treatment based on an individual's genetic make-up. By determining the type of disease an individual has, based on the mutations present or having the ability to prescribe the appropriate antimicrobials based on the knowledge of the organism's resistance patterns, the clinician will be better able to treat and diagnose a patient which ultimately will improve patient outcomes and prognosis.

ACKNOWLEDGEMENTS

The author would like to acknowledge the following individuals for their help and guidance in the development of this series: James Dermody, PhD; Puneet Dhawan, PhD; Robert Donnelly; PhD; and Helen Fernandes, PhD.

NOTE: The author is not endorsing any particular company or product and has no financial gain or otherwise interest in the products presented.

REFERENCES

1. The Human Genome Project Begins. Available from http://unlockinglifescode.org/timeline?tid=4. Accessed 2014 March 25.
2. All About The Human Genome Project (HGP). Available from http://www.genome.gov/10001772. Accessed 2014 March 25.
3. Human Genome Project completion announced. Available from http://unlockinglifescode.org/timeline?tid=4. Accessed 2014 March 25.
4. Human Genome Project. Available from http://web.ornl.gov/sci/techresources/Human_Genome/index.shtml. Accessed 2014 March 25.
5. DNA Structure & DNA Replication. Available from http://www.biology-online.org/1/5_DNA.htm. Accessed 2014 March 19.
6. Genetics Homepage: What is DNA. Available from http://ghr.nlm.nih.gov/handbook/basics/dna. Accessed 2014 March 19.
7. Ferrier DR. Biochemistry. 3rd ed. Lippincott Williams & Wilkins; 2005. Ch 29, p. 393.
8. Frederick Sanger – Biographical. Available from

http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1980/sanger-bio.html. Accessed 2014 March 25.
9. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol. 1975 May 25;94(3):441–448.
10. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA. 1977;74:5463–7.
11. Autoradiography of Dideoxy Sequencing Gel. Available from http://www.sanger.ac.uk/. Accessed 2014 March 29.
12. Jorgenson JW, Lukacs KD. Science. 1981;53:266–72.
13. Smith LM, Sanders JZ, Kaiser RJ, Hughes P., Dodd C., Connell C.R., et. al. Fluorescence detection in automated DNA sequence analysis. Nature. 1986;321:674–9.
14. Buckingham L, Flaws M. Molecular Diagnostics: Fundamentals, Methods, and Clinical Applications, F. A. Davis Co., 2007. Ch10, p 203.
15. Slatko BE , et al. "First generation" automated DNA sequencing technology. Curr. Protoc. Mol. Biol. 2011; S96:7.2.
16. Ion TorrentTM Next-Generation Sequencing. Available from http://www.lifetechnologies.com/us/en/home/life-science /sequencing.html?s _kwcid= AL!3652!10!33369080 44!26566630335&ef_id=UysVxwAA AAep8zDk:20140320162252:s. Accessed 2014 March 29.
17. Illumina MiSeq Desktop Sequencer . Available from http://www.illumina.com/systems/miseq.ilmn. Accessed 2014 March 29.
18. SMRT Technology. Available from http://www.pacificbiosciences.com/products/smrt-technology/. Accessed 2014 March 29.
19. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics. 2012;13:341. doi:10.1186/1471-2164-13-341.
20. SOLiD™ 4 System. Available from http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/solid-next-generation-sequencing/next-generation-systems/solid-4-system.html?CID=FL-091411_solid4. Accessed 2014 March 28.
21. The New GS Junior. Available from http://www.gsjunior.com/index.php. Accessed 2014 March 28.
22. Ion PGM™ System for Next-Generation Sequencing. Available from http://www.lifetechnologies.com/us/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-run-sequence/ion-pgm-system-for-next-generation-sequencing.html. Accessed 2014 March 28.
23. Ion Torrent PGM™ System Specifications. Available from http://www.lifetechnologies.com/us/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-run-sequence/ion-pgm-system-for-next-generation-sequencing/ion-pgm-system-specifications.html. Accessed 2014 March 28.
24. Ion Torrent™ Next-Generation Sequencing Technology. Available from http://www.lifetechnologies.com/us/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-technology.html. Accessed 2014 March 28.
25. Ion Torrent Next-Generation Sequencing Workflow. Available from http://www.lifetechnologies.com/us/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow.html. Accessed 2014 March 28.
26. Life Technologies Technical Support –Personal Communication. 2014 March 28.
27. Ion Torrent™ Next-Generation Sequencing Chips Product Table. Available from http://www.lifetechnologies.com/us/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-products-services/ion-torrent-next-generation-sequencing-chips-product-table.html. Accessed 2014 March 28.
28. The Chip is the Machine™. Available fromhttp://www.lifetechnologies.com/us/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-run-sequence/ion-pgm-ion-proton-system-chips.html. Accessed 2014 March 28.
29. Ion Torrent™ Next-Generation Sequencing Ion Chef™ System. Available from http://www.lifetechnologies.com/us/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/prepare-template/ion-torrent-next-generation-sequencing-ion-chef-system.html. Accessed 2019 March 28.
30. Illumina System. Available from http://www.illumina.com/systems/miseq/system.ilmn. Accessed 2014 March 28.
31. Illumina MiSeq Vision. Available from http://www.youtube.com/watch?v=t0akxx8Dwsk. Accessed 2014 March 29.
32. Illumina sequencing. Available from http://www.youtube.com/watch?v=l99aKKHcxC4. Accessed 2014 March 28.
33. Illumina Workflow. Available from http://www.illumina.com/systems/miseq/workflow.ilmn. Accessed 2014 March 28.
34. Pacific Biosciences Overview. Available from http://www.pacificbiosciences.com/applications/overview/index.html. Accessed 2014 March 28.
35. SMRT Cells. Available from http://www.pacificbiosciences.com/products/consumables/SMRT-cells/. Accessed 2014 March 28.
36. Overview of SMRT Technology. Available from http://www.youtube.com/watch?v=WMZmG00uhwU. Accessed 2014 March 28.
37. SMRT. Available from http://www.youtube.com/watch?v=GX6RSKh4J7E. Accessed 2014 March 28.
38. PACBIO RS II WORKFLOW. Available fromhttp://www.pacificbiosciences.com/products/pacificbio-rs-workflow-main/. Accessed 2014 March 28.