

Chapter 2

Overview of Sequencing Technology Platforms

Samuel Myllykangas, Jason Buenrostro, and Hanlee P. Ji

Abstract The high-throughput DNA sequencing technologies are based on immobilization of the DNA samples onto a solid support, cyclic sequencing reactions using automated fluidics devices, and detection of molecular events by imaging. Featured sequencing technologies include: GS FLX by 454 Life Technologies/Roche, Genome Analyzer by Solexa/Illumina, SOLiD by Applied Biosystems, CGA Platform by Complete Genomics, and PacBio RS by Pacific Biosciences. In addition, emerging technologies are discussed.

2.1 Introduction

High-throughput sequencing has begun to revolutionize science and healthcare by allowing users to acquire genome-wide data using massively parallel sequencing approaches. During its short existence, the high-throughput sequencing field has witnessed the rise of many technologies capable of massive genomic analysis. Despite the technological dynamism, there are general principles employed in the construction of the high-throughput sequencing instruments.

Commercial high-throughput sequencing platforms share three critical steps: DNA sample preparation, immobilization, and sequencing (Fig. 2.1). Generally, preparation of a DNA sample for sequencing involves the addition of defined sequences, known as “adapters,” to the ends of randomly fragmented DNA (Fig. 2.2). This DNA preparation with common or universal nucleic acid ends is commonly referred to as the “sequencing library.” The addition of adapters is required to anchor the DNA fragments of the sequencing library to a solid surface and define the site in

S. Myllykangas • J. Buenrostro • H.P. Ji (✉)

Division of Oncology, Department of Medicine, Stanford Genome Technology Center,
Stanford University School of Medicine, CCSR, 269 Campus Drive,
94305 Stanford, CA, USA

e-mail: smyllyka@stanford.edu; jdbuenrostro@gmail.com; genomics_ji@stanford.edu

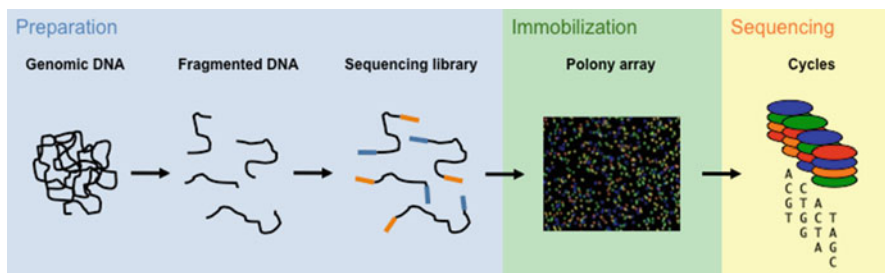


Fig. 2.1 High-throughput sequencing workflow. There are three main steps in high-throughput sequencing: preparation, immobilization, and sequencing. Preparation of the sample for high-throughput sequencing involves random fragmentation of the genomic DNA and addition of adapter sequences to the ends of the fragments. The prepared sequencing library fragments are then immobilized on a solid support to form detectable sequencing features. Finally, massively parallel cyclic sequencing reactions are performed to interrogate the nucleotide sequence

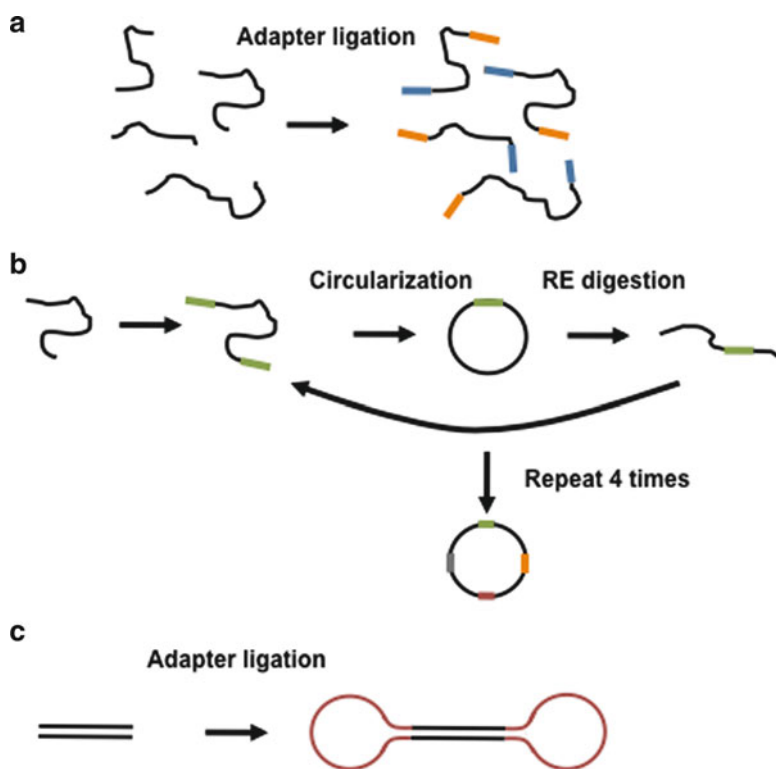


Fig. 2.2 Sequencing library preparation. There are three principal approaches for addition of adapter sequences and preparation of the sequencing library. **(a)** Linear adapters are applied in the GS FLX, Genome Analyzer, and SOLiD systems. Specific adaptor sequences are added to both ends of the genomic DNA fragments. **(b)** Circular adapters are applied in the CGA platform, where four distinct adaptor sequences are internalized into a circular template DNA. **(c)** Bubble adapters are used in the PacBio RS sequencing system. Hairpin forming bubble adapters are added to double-strand DNA fragments to generate a circular molecule

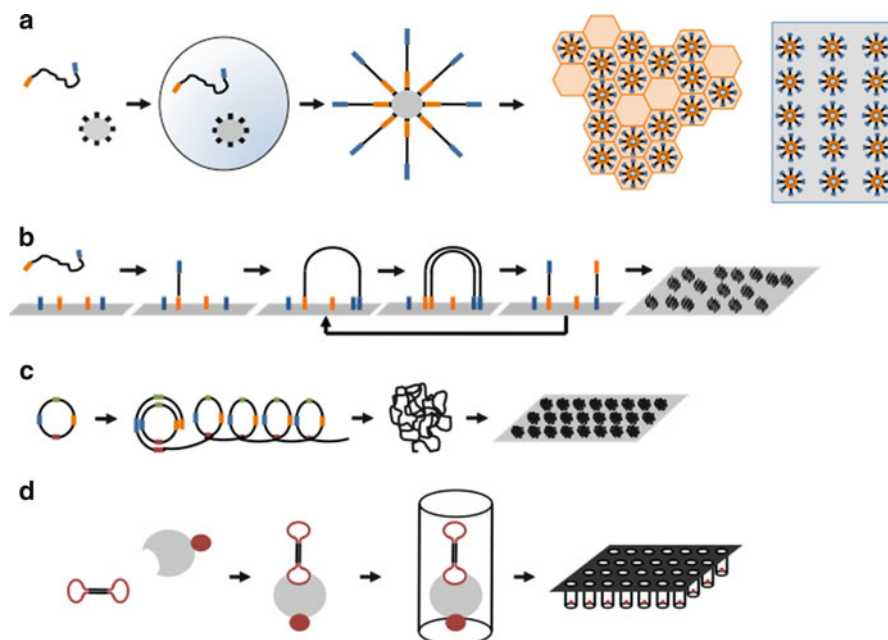


Fig. 2.3 Generation of sequencing features. High-throughput sequencing systems have taken different approaches in the generation of the detectable sequencing features. **(a)** Emulsion PCR is applied in the GS FLX and SOLiD systems. Single enrichment bead and sequencing library fragment are emulsified inside an aqueous reaction bubble. PCR is then applied to populate the surface of the bead by clonal copies of the template. Beads with immobilized clonal DNA collections are deposited onto a Picotiter plate (GS FLX) or on a glass slide (SOLiD). **(b)** Bridge-PCR is used to generate the in situ clusters of amplified sequencing library fragments on a solid support. Immobilized amplification primers are used in the process. **(c)** Rolling circle amplification is used to generate long stretches of DNA that fold into nanoballs that are arrayed in the CGA technology. **(d)** Biotinylated DNA polymerase binds to bubble adapted template in the PacBio RS system. Polymerase/template complex is immobilized on the bottom of a zero mode wave guide (ZMW)

which the sequencing reactions begin. These high-throughput sequencing systems, with the exception of PacBio RS, require amplification of the sequencing library DNA to form spatially distinct and detectable sequencing features (Fig. 2.3). Amplification can be performed in situ, in emulsion or in solution to generate clusters of clonal DNA copies. Sequencing is performed using either DNA polymerase synthesis for fluorescent nucleotides or the ligation of fluorescent oligonucleotides (Fig. 2.4).

The high-throughput sequencing platforms integrate a variety of fluidic and optic technologies to perform and monitor the molecular sequencing reactions. The fluidics systems that enable the parallelization of the sequencing reaction form the core of the high-throughput sequencing platform. Micro-liter scale fluidic devices support the DNA immobilization and sequencing using automated liquid dispensing mechanisms. These instruments enable the automated flow of reagents onto the immobilized

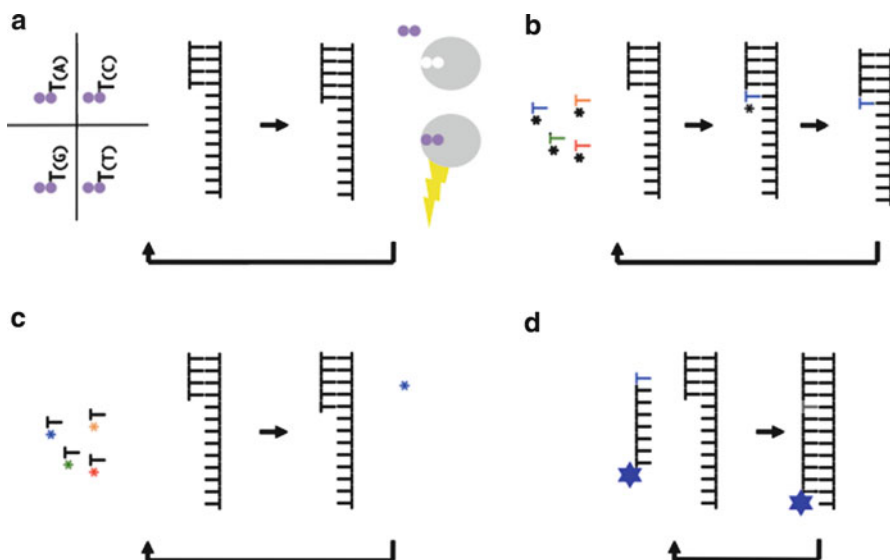


Fig. 2.4 Cyclic sequencing reactions. **(a)** Pyrosequencing is based on recording light bursts during nucleotide incorporation events. Each nucleotide is interrogated individually. Pyrosequencing is a technique used in GS FLX sequencing. **(b)** Reversible terminator nucleotides are used in the Genome Analyzer system. Each nucleotide has a specific fluorescent label and a termination moiety that prevents addition of other reporter nucleotides to the synthesized strand. All four nucleotides are analyzed in parallel and one position is sequenced at each cycle. **(c)** Nucleotides with cleavable fluorophores are used in the PacBio RS system. Each nucleotide has a specific fluorophore, which gets cleaved during the incorporation event. **(d)** Sequencing by ligation is applied in the SOLiD and CGA platforms. Although they have different approaches, the general principle is the same. Both systems apply fluorophore-labeled degenerate oligonucleotides that correspond to a specific base in the molecule

DNA samples for cyclic interrogation of the nucleotide sequence. Massive parallel sequencing systems apply high-throughput optical systems to capture information about the molecular events, which define the sequencing reaction and the sequence of the immobilized sequencing library. Each sequencing cycle consists of incorporating a detectable nucleic acid substrate to the immobilized template, washes, and imaging the molecular event. Incorporation–washing–imaging cycles are repeated to build the DNA sequence read. PacBio RS is based on monitoring DNA polymerization reactions in parallel by recording the light pulses emitted during each incorporation event in real time.

High-throughput DNA sequencing has been commercialized by a number of companies (Table 2.1). The GS FLX sequencing system (Margulies et al. 2005), originally developed by 454 Life Sciences and later acquired by Roche (Basel, Switzerland), was the first commercially available high-throughput sequencing platform. The first short read sequencing technology, Genome Analyzer, was developed by Solexa, which was later acquired by Illumina Inc. (San Diego, CA) (Bentley et al. 2008; Bentley 2006). The SOLiD sequencing system by Applied Biosystems

Table 2.1 High-throughput sequencing platforms

| Platform | Company | Sequencing library | Support | Feature generation | Sequencing reaction | Detection method |
|-----------------|--------------------------|--------------------|---------------------|------------------------------|----------------------|---|
| GS FLX | 454 Life Sciences, Roche | Linear adapters | Picotiter plate | Emulsion PCR | Synthesis | Pyrosequencing |
| Genome analyzer | Solexa, Illumina | Linear adapters | Flow cell | Bridge PCR | Synthesis | Fluorophore labeled reversible terminator nucleotides |
| SOLiD | Applied Biosystems | Linear adapters | Flow cell | Emulsion PCR | Ligation | Fluorophore labeled oligonucleotide probes |
| CGA platform | Complete genomics | Circular adapters | DNA nanoball arrays | Rolling circle amplification | Ligation | Fluorophore labeled oligonucleotide probes |
| PacBio RS | Pacific biosciences | Bubble adapters | Zero mode waveguide | Single molecule | Synthesis, real-time | Phospholinked fluorophore labeled nucleotides |

(Foster City, CA) applies fluorophore labeled oligonucleotide panel and ligation chemistry for sequencing (Smith et al. 2010; Valouev et al. 2008). Complete Genomics (Mountain View, CA) has developed a sequencing technology called CGA that is based on preparing a semiordered array of DNA nanoballs on a solid surface (Drmanac et al. 2010). Pacific Biosciences (Menlo Park, CA) has developed PacBio RS sequencing technology, which uses the polymerase enzyme, fluorescent nucleotides, and high-content imaging to detect single-molecule DNA synthesis events in real time (Eid et al. 2009).

2.2 Genome Sequencer GS FLX

The Roche GS FLX sequencing process consists of preparing an end-modified DNA fragment library, sample immobilization on streptavidin beads, and pyrosequencing.

2.2.1 *Preparation of the Sequencing Library*

Sample preparation of the GS FLX sequencing system begins with random fragmentation of DNA into 300–800 base-pair (bp) fragments (Margulies et al. 2005). After shearing, fragmented double-stranded DNA is repaired with an end-repair enzyme cocktail and adenine bases are added to the 3' ends of fragments. Common adapters, named “A” and “B,” are then nick-ligated to the fragments ends. Nicks present in the adapter-to-fragment junctions are filled in using a strand-displacing *Bst* DNA polymerase. Adapter “B” carries a biotin group, which facilitates the purification of homo-adapted fragments (A/A or B/B). The biotin labeled sequencing library is captured on streptavidin beads. Fragments containing the biotin labeled B adapter are bound to the streptavidin beads while homozygous, nonbiotinylated A/A adapters are washed away. The immobilized fragments are denatured after which both strands of the B/B adapted fragments remain immobilized by the streptavidin–biotin bond and single-strand template of the A/B fragments are freed and used in sequencing.

2.2.2 *Emulsion PCR and Immobilization to Picotiter Plate*

In GS FLX sequencing, the single-strand sequencing library fragment is immobilized onto a specific DNA capture bead (Fig. 2.3a). GS FLX sequencing relies on capturing one DNA fragment onto a single bead. One-to-one ratio of beads and fragments is achieved by limiting dilutions. The bead-bound library is then amplified using a specific form of PCR. In emulsion PCR, parallel amplification of bead captured library fragments takes place in a mixture of oil and water. Aqueous bubbles, immersed in oil, form microscopic reaction entities for each individual capture bead. Hundreds of thousands of amplified DNA fragments can be immobilized on the surface of each bead.

In the GS FLX sequencing platform, beads covered with amplified DNA can be immobilized on a solid support (Fig. 2.3a). The GS FLX sequencing platform uses a “Picotiter plate,” a solid phase support containing over a million picoliter volume wells (Margulies et al. 2005). The dimensions of the wells are such that only one bead is able to enter each position on the plate. Sequencing chemistry flows through the plate and insular sequencing reactions take place inside the wells. The Picotiter plate can be compartmentalized up to 16 separate reaction entities using different gaskets.

2.2.3 Pyrosequencing

The GS FLX sequencing reaction utilizes a process called pyrosequencing (Fig. 2.4a) to detect the base incorporation events during sequencing (Margulies et al. 2005). In pyrosequencing, Picotiter plates are flushed with nucleotides and the activity of DNA polymerase and the incorporation of a nucleotide lead to the release of a pyrophosphate. ATP sulfurylase and luciferase enzymes convert the pyrophosphate into a visible burst of light, which is detected by a CCD imaging system. Each nucleotide species (i.e., dATP, dCTP, dGTP, and dTTP) is washed over the Picotiter plate and interrogated separately for each sequencing cycle. The GS FLX technology relies on asynchronous extension chemistry, as there is no termination moiety that would prevent addition of multiple bases during one sequencing cycle. As a result, multiple nucleotides can be incorporated to the extending DNA strand and accurate sequencing through homopolymer stretches (i.e., AAA) represents a challenging technical issue for GS FLX. However, a number of improvements have been made to improve the sequencing performance of homopolymers (Smith et al. 2010).

2.3 Genome Analyzer

The Genome Analyzer system is based on immobilizing linear sequencing library fragments using solid support amplification. DNA sequencing is enabled using fluorescent reversible terminator nucleotides.

2.3.1 Sequencing Library Preparation

Sample preparation for the Illumina Inc. Genome Analyzer involves adding specific adapter sequences to the ends of DNA molecules (Fig. 2.2a) (Bentley et al. 2008; Bentley 2006). The production of a sequencing library initiates with fragmentation of the DNA sample, which defines the molecular entry points for the sequencing reads. Then, an enzyme cocktail repairs the staggered ends, after which, adenines (A) are added to the 3' ends of the DNA fragments. A-tailed DNA is applied as a template to ligate double strand, partially complementary adapters to the DNA fragments.

Adapted DNA library is size selected and amplified to improve the quality of sequence reads. Amplification introduces end-specific PCR primers that bring in the portion of the adapter required for sample processing on the Illumina Inc. system.

2.3.2 Solid Support Amplification

Illumina Inc. flow cells are planar, fluidic devices that can be flushed with sequencing reagents. The inner surface of the flow cell is functionalized with two oligonucleotides, which creates an ultra-dense primer field. The sequencing library is immobilized on the surface of a flow cell (Fig. 2.3b). The immobilized primers on the flow cell surface have sequences that correspond to the DNA adapters present in the sequencing library. DNA molecules in the sequencing library hybridize to the immobilized primers and function as templates in strand extension reactions that generate immobilized copies of the original molecules.

In the Illumina Inc. Genome Analyzer system, the preparation of the flow cell requires amplification of individual DNA molecules of a sequencing library and formation of spatially condensed, microscopically detectable clusters of molecular copies (Fig. 2.3b). The primer functionalized flow cell surface serves as a support for amplification of the immobilized sequencing library by a process also known as “Bridge-PCR.”

Generally, PCR is performed in solution and relies on repeated thermal cycles of denaturation, annealing, and extension to exponentially amplify DNA molecules. In the Illumina Inc. Genome Analyzer Bridge-PCR system, amplification is performed on a solid support using immobilized primers and in isothermal conditions using reagent flush cycles of denaturation, annealing, extension, and wash. Bridge-PCR initiates by hybridization of the immobilized sequencing library fragment and a primer to form a surface-supported molecular bridge structure. Arched molecule is a template for a DNA polymerase-based extension reaction. The resulting bridged double-strand DNA is freed using a denaturing reagent. Repeated reagent flush cycles generate groups of thousands of DNA molecules, also known as “clusters,” on each flow cell lane. DNA clusters are finalized for sequencing by unbinding the complementary DNA strand to retain a single molecular species in each cluster, in a reaction called “linearization,” followed by blocking the free 3' ends of the clusters and hybridizing a sequencing primer.

2.3.3 Sequencing Using Fluorophore Labeled Reversible Terminator Nucleotides

The prepared flow cell is connected to a high-throughput imaging system, which consists of microscopic imaging, excitation lasers, and fluorescence filters. Molecularly, Illumina Inc.'s sequencing-by-synthesis method employs four distinct fluorophores and reversibly terminated nucleotides (Fig. 2.4b). The sequencing reaction initiates by

DNA polymerase synthesis of a fluorescent reversible terminator nucleotide from the hybridized sequencing primer. The extended base contains a fluorophore specific to the extended base and a reversible terminator moiety, which inhibits the incorporation of additional nucleotides.

After each incorporation reaction, the immobilized nucleotide fluorophores, corresponding to each cluster, are imaged in parallel. X – Y position of imaged nucleotide fluorophore defines the first base of a sequence read. Before proceeding to next cycle, reversible-terminator moieties and fluorophores are detached using a cleavage reagent, enabling subsequent addition of nucleotides. The synchronous extension of the sequencing strand by one nucleotide per cycle ensures that homopolymer stretches (consecutive nucleotides of the same kind, i.e., AAAA) can be accurately sequenced. However, failure to incorporate a nucleotide during a sequencing cycle results in off-phasing effect – some molecules are lagging in extension and the generalized signal derived from the cluster deteriorates over cycles. Therefore, Illumina Inc. sequencing accuracy declines as the read length increases, which limits this technology to short sequence reads.

2.4 SOLiD

The Applied Biosystems SOLiD sequencer, featured in Valouev et al. (2008) and Smith et al. (2010), is based on the Polonator technology (Shendure et al. 2005), an open source sequencer that utilizes emulsion PCR to immobilize the DNA library onto a solid support and cyclic sequencing-by-ligation chemistry.

2.4.1 Sequencing Library Preparation and Immobilization

The in vitro sequencing library preparation for SOLiD involves fragmentation of the DNA sample to an appropriate size range (400–850 bp), end repair and ligation of “P1” and “P2” DNA adapters to the ends of the library fragments (Valouev et al. 2008). Emulsion PCR is applied to immobilize the sequencing library DNA onto “P1” coated paramagnetic beads. High-density, semi-ordered polony arrays are generated by functionalizing the 3' ends of the templates and immobilizing the modified beads to a glass slide. The glass slides can be segmented up to eight chambers to facilitate up scaling of the number of analyzed samples.

2.4.2 Sequencing by Ligation

The SOLiD sequencing chemistry is based on ligation (Fig. 2.4d). A sequencing primer is hybridized to the “P1” adapter in the immobilized beads. A pool of uniquely labeled oligonucleotides contains all possible variations of the complementary

bases for the template sequence. SOLiD technology applies partially degenerate, fluorescently labeled, DNA octamers with dinucleotide complement sequence recognition core. These detection oligonucleotides are hybridized to the template and perfectly annealing sequences are ligated to the primer. After imaging, unextended strands are capped and fluorophores are cleaved. A new cycle begins 5 bases upstream from the priming site. After the seven sequencing cycles first sequencing primer is peeled off and second primer, starting at $n-1$ site, is hybridized to the template. In all, 5 sequencing primers (n , $n-1$, $n-2$, $n-3$, and $n-4$) are utilized for the sequencing. As a result, the 35-base insert is sequenced twice to improve the sequencing accuracy.

Since the ligation-based method in the SOLiD system requires complex panel of labeled oligonucleotides and sequencing proceeds by off-set steps, the interpretation of the raw data requires a complicated algorithm (Valouev et al. 2008). However, the SOLiD system achieves a slightly better performance in terms of sequencing accuracy due to the redundant sequencing of each base twice by a dinucleotide detection core structure of the octamer sequencing oligonucleotides.

2.5 CGA Platform

The CGA Platform (Complete Genomics) represents the first high-throughput platform only available to the public as a service (Table 2.1). The CGA technology is based on preparation of circular DNA libraries (Fig. 2.2c) and rolling circle amplification (RCA) to generate DNA nanoballs that are arrayed on a solid support (Fig. 2.3c) (Drmanac et al. 2010).

2.5.1 Sequencing Library Preparation

DNA is randomly fragmented and 400–500 bp fragments are collected. The fragment ends are enzymatically end-repaired and dephosphorylated. Common adapters are ligated to the DNA fragments using nick translation. These adapter libraries are enriched and Uracils are incorporated in the products using PCR and uracil containing primers. Uracils are removed from the final product to create overhangs. The products are digested and methylated with *AcuI* and circularized using T4 DNA ligase in a presence of a splint oligonucleotide. The circularized products are purified using an exonuclease, which degrades residual linear DNA molecules. Linearization, adapter ligation, PCR amplification, restriction enzyme digestion, and circularization process are repeated until four unique adapters are incorporated into the circular sequencing library molecules. Prior to the final circularization step, a single-strand template is purified using strand separation by bead capture and exonuclease treatment. The final product contains two 13 base genomic DNA inserts and two 26 base genomic DNA inserts adjacent to the adapter sequences.

2.5.2 DNA Nanoball Array

To prepare the immobilized sequencing features for Complete Genomics sequencing, circular, single-strand DNA library is amplified using RCA and a highly processive and strand displacing Phi29 polymerase. RCA creates long DNA strands from the circular DNA library templates that contain short palindrome sequences. The palindrome sequences within the long linear products promote intramolecular coiling of the molecule and formation of the DNA nanoballs (DNBs). A nanoball is a long strand of repetitive fragments of amplified DNA, which forms a detectable, three-dimensional, condensed, and spherical sequencing object.

The hexamethyldisilazane (HDMS) covered surface of the CGA Platform's fluidic chamber is spotted by aminosilane using photolithography techniques. Three-hundred nm aminosilane spots cover over 95% of the CGA surface. While HDMS inhibits DNA binding, the positively charged aminosilane binds the negatively charged DNBs. Randomly organized but regionally ordered high-density array has 350 million immobilized DNBs within a distance of 1.29 μm between the centers of the spots.

2.5.3 Sequencing by Ligation Using Combinatorial Probe Anchors

Complete genomics' CGA Platform uses a novel strategy called combinatorial probe anchor ligation (cPAL) for sequencing. The process begins by hybridization between an anchor molecule and one of the unique adapters. Four degenerate 9-mer oligonucleotides are labeled with specific fluorophores that correspond to a specific nucleotide (A, C, G, or T) in the first position of the probe. Sequence determination occurs in a reaction where the correct matching probe is hybridized to a template and ligated to the anchor using T4 DNA ligase. After imaging of the ligated products, the ligated anchor-probe molecules are denatured. The process of hybridization, ligation, imaging, and denaturing is repeated five times using new sets of fluorescently labeled 9-mer probes that contain known bases at the $n+1$, $n+2$, $n+3$, and $n+4$ positions.

After five cycles, the fidelity of the ligation reaction decreases and sequencing continues by resetting the reaction using an anchor with degenerate region of 5 bases. Another five cycles of sequencing by ligation are performed using the fluorescently labeled, degenerate 9-mer probes. The cyclic sequencing of 10 bases can be repeated up to eight times, starting at each of the unique anchors, and resulting in 62–70 base long reads from each DNB.

Unlike other high-throughput sequencing platforms that involve additive detection chemistries, the cPAL technology is unchained as sequenced nucleotides are not physically linked. The anchor and probe constructs are removed after each sequencing cycle and the next cycle is initiated completely independent of the molecular

events of the previous cycle. A disadvantage of this system is that read lengths are limited by the sample preparation, even if, longer reads up to 120 bases can be achieved by adding more restriction enzyme sites.

2.6 PacBio RS

PacBio RS is a single-molecule real-time (SMRT) sequencing system developed by Pacific Biosciences (Eid et al. 2009).

2.6.1 *Preparation of the Sequencing Library*

SMRTbell is the default method for preparing sequencing libraries for PacBio RS in order to get high accuracy variant detection (Travers et al. 2010) (Fig. 2.3d). For genome sequencing, DNA is randomly fragmented and then end-repaired. Then, 3' adenine is added to the fragmented genomic DNA, which facilitates ligation of an adapter with a T overhang. Single DNA oligonucleotide, which forms an intramolecular hairpin structure, is used as the adapter. The SMRTbell DNA template is structurally a linear molecule but the bubble adapters create a topologically circular molecule.

2.6.2 *The SMRT Cell*

The SMRT cell houses a patterned array of zero-mode waveguides (ZMWs) (Korlach et al. 2008b; Levene et al. 2003). ZMWs are nanofabricated on a glass surface. The volume of the nanometer-sized aluminum layer wells is in zeptoliter scale. The SMRT cell is prepared for polymerase immobilization by coating the surface with streptavidin. The preparation of the sequencing reaction requires incubating a biotinylated Phi29 DNA polymerase with primed SMRTbell DNA templates. The coupled products are then immobilized to the SMRT cell using a biotin–streptavidin interaction.

2.6.3 *Processive DNA Sequencing by Synthesis*

When the sequencing reaction begins, the tethered polymerase incorporates nucleotides with individually phospholinked fluorophores, each fluorophore corresponding to a specific base, to the growing DNA chain (Korlach et al. 2008a). During the initiation of a base incorporation event, the fluorescent nucleotide is brought into

the polymerase's active site and into proximity of the ZMW glass surface. At the bottom of the ZMW, high-resolution camera records the fluorescence of the nucleotide being incorporated. During the incorporation reaction a phosphate-coupled fluorophore is released from the nucleotide and that dissociation diminishes the fluorescent signal. While the polymerase synthesizes a copy of the template strand, incorporation events of successive nucleotides are recorded in a movie-like format.

The tethered Phi29 polymerase is a highly processive strand-displacing enzyme capable of performing RCA. Using SMRTbell libraries with small insert sizes, it is possible to sequence the template using a scheme called circular consensus sequencing. The same insert is read on the sense and antisense strands multiple times and the redundancy is dependent on insert size. This highly redundant sequencing approach improves the accuracy of the base calls overcoming the high error rates associated with real-time sequencing and allowing accurate variant detection. For low accuracy and long read lengths, larger insert sizes can be used. The unique method of detecting nucleotide incorporation events in real time allows the development of novel applications, such as the detection of methylated cytosines based on differential polymerase kinetics (Flusberg et al. 2010).

2.7 Emerging Technologies

The phenomenal success of high-throughput DNA sequencing systems has fueled the development of novel instruments that are anticipated to be faster than the current high-throughput technologies and will lower the cost of genome sequencing. These future generations of DNA sequencing are based on technologies that enable more efficient detection of sequencing events. Instruments for detection of ion release during incorporation of label-free natural nucleotides and nanopore technologies are emerging. The pace of technological development in the field of genome sequencing is overwhelming and new technological breakthroughs are probable in the near future.

2.7.1 *Semiconductor Sequencing*

Life Technology and Ion Torrent are developing the Ion Personal Genome Machine, which represents an affordable and rapid bench top system designed for small projects. The IPG system harbors an array of semiconductor chips capable of sensing minor changes in pH and detecting nucleotide incorporation events by the release of a hydrogen ion from natural nucleotides. The Ion Torrent system does not require any special enzymes or labeled nucleotides and takes advantage of the advances made in the semiconductor technology and component miniaturization.

2.7.2 Nanopore Sequencing

Nanopore sequencing is based on a theory that recording the current modulation of nucleic acids passing through a pore could be used to discern the sequence of individual bases within the DNA chain. Nanopore sequencing is expected to offer solutions to limitations of short read sequencing technologies and enable sequencing of large DNA molecules in minutes without having to modify or prepare samples. Despite the technology's potential many technical hurdles remain.

Exonuclease DNA sequencing from Oxford Nanopores represents a possible solution to some of the technical hurdles found in nanopore sequencing. The system seeks to couple an exonuclease to a biological alpha hemolysin pore and plant that construct onto a lipid bilayer. When the exonuclease encounters a single-strand DNA molecule, it cleaves a base and passes it through the pore. Each base creates a unique signature of current modulation as it crosses through the lipid bilayer, which can be detected using sensitive electrical methods.

2.8 Conclusions

Although high-throughput sequencing is in its infancy, it has already begun to reshape the ways in which biology is portrayed. In principle, massive parallel sequencing systems are powerful technology rigs that integrate basic molecular biology, automated fluidics devices, high-throughput microscopic imaging, and information technologies. By default, to be able to use these systems requires comprehensive understanding of the complex underlying molecular biology and biochemistry. The ultra-high-throughput instruments are essentially high-tech machines and understanding the engineering principles gives the user the ability to command and troubleshoot the massive parallel sequencing systems. The complexity and size of the experimental results is rescaling the boundaries of biological inquiry. With the advent of these technologies, it is required that users acquire computational skills and develop systematic data analysis pipelines. High-throughput sequencing has presented an introduction to an exciting new era of multidisciplinary science.

References

- Bentley, D. R. 2006. Whole-genome re-sequencing. *Curr Opin Genet Dev* 16 (6):545–552. doi:S0959-437X(06)00208-5 [pii] 10.1016/j.gde.2006.10.009.
- Bentley, DR, S Balasubramanian, HP Swerdlow, GP Smith, J Milton, CG Brown, KP Hall et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456:53–59.
- Drmanac, R., A. B. Sparks, M. J. Callow, A. L. Halpern, N. L. Burns, B. G. Kermani, P. Carnevali et al. 2010. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 327 (5961):78–81. doi:1181498 [pii] 10.1126/science.1181498.

- Eid, J, A Fehr, J Gray, K Luong, J Lyle, G Otto, P Peluso et al. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* 323:133–138.
- Flusberg, B. A., D. R. Webster, J. H. Lee, K. J. Travers, E. C. Olivares, T. A. Clark, J. Korlach, and S. W. Turner. 2010. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods* 7 (6):461–465. doi:nmeth.1459 [pii] 10.1038/nmeth.1459.
- Korlach, J, A Bibillo, J Wegener, P Peluso, TT Pham, I Park, S Clark, GA Otto, and SW Turner. 2008. Long, processive enzymatic DNA synthesis using 100% dye-labeled terminal phosphate-linked nucleotides. *Nucleosides Nucleotides Nucleic Acids* 27:1072–1083.
- Korlach, J., P. J. Marks, R. L. Cicero, J. J. Gray, D. L. Murphy, D. B. Roitman, T. T. Pham, G. A. Otto, M. Foquet, and S. W. Turner. 2008. Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures. *Proc Natl Acad Sci USA* 105 (4):1176–1181.
- Levene, M. J., J. Korlach, S. W. Turner, M. Foquet, H. G. Craighead, and W. W. Webb. 2003. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* 299 (5607):682–686.
- Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bembien, J. Berka et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437 (7057):376–380. doi:nature03959 [pii] 10.1038/nature03959.
- Shendure, J., G. J. Porreca, N. B. Reppas, X. Lin, J. P. McCutcheon, A. M. Rosenbaum, M. D. Wang, K. Zhang, R. D. Mitra, and G. M. Church. 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 309 (5741):1728–1732. doi:1117389 [pii] 10.1126/science.1117389.
- Smith, A. M., L. E. Heisler, R. P. St Onge, E. Farias-Hesson, I. M. Wallace, J. Bodeau, A. N. Harris et al. 2010. Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples. *Nucleic Acids Res* 38 (13):e142.
- Travers, K. J., C. S. Chin, D. R. Rank, J. S. Eid, and S. W. Turner. 2010. A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Res* 38 (15):e159. doi:gkq543 [pii] 10.1093/nar/gkq543.
- Valouev, A., J. Ichikawa, T. Tonthat, J. Stuart, S. Ranade, H. Peckham, K. Zeng et al. 2008. A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res* 18 (7):1051–1063. doi:gr.076463.108 [pii] 10.1101/gr.076463.108.

Bioinformatics for High Throughput Sequencing

Rodríguez-Ezpeleta, N.; Hackenberg, M.; Aransay, A.M.
(Eds.)

2012, XI, 255 p. 29 illus., 25 illus. in color., Hardcover

ISBN: 978-1-4614-0781-2