Published in final edited form as:
Recent Pat Biomed Eng. 2008 ; 2008(1): 60–67.

Recent Patents and Advances in the Next-Generation Sequencing Technologies

Biaoyang Lin[1,2,*], **Jun Wang**[2], and **Yin Cheng**[2]

[1]The Institute for Systems Biology, Seattle, WA, 98021 USA

[2]Zhejiang-California International Nanosystems Institute, Hangzhou, 310029, China

Abstract

We are now witnessing a new genomic revolution due to the arrival and continued advancements in the next-generation high-throughput sequencing technologies, which encompass sequencing by synthesis including fluorescent *in situ* sequencing (FISSEQ) and pyrosequencing, sequencing by ligation including using polony amplification and supported oligonucleotide detection (SOLiD), sequencing by hybridization in combination with sequencing-by-ligation and nanopore technology, nanopore sequencing and other novel sequencing technologies using nano-transistor array, scanning tunneling microscopy and nanowire molecule sensors etc. We review here major technologies and recent patents for achieving high-throughput, ultra-fast, extremely cheap, and highly accurate sequencing. We will see enormous impacts of these next-generation sequencing methods for solving complex biological problems and for ushering in the practice of personalized medicine.

Keywords

Sequencing by synthesis; sequencing by ligation; sequencing by hybridization; single molecule DNA sequencing; nanopore; cleavable dye; reversible terminator; polony sequencing; pyrosequencing; supported oligonucleotide detection (SOLiD)

INTRODUCTION

Sequencing technologies have recently under dramatic development. In the 1970's Sanger *et al*. developed the technology of sequencing by chain termination and gel separation [1,2]. In the method, a mixed population of nucleic acid fragments representing terminations at each base was generated using 'terminator'-the 2′,3′-dideoxy and arabinonucleoside analogues of the normal deoxynucleoside triphosphates. They are run on an electrophoretic gel and the sequence can be 'read' from the order of fragments in the gel. A similar sequencing method that relies on chemical degradation of nucleic acid fragments at each base was also developed by Maxam and Gilbert [3]. The gel-based sequencing technology has undergone dramatic improvement in throughput from the parallelization, automation, and refinement of sequencing methods and chemistry, evidenced by the wide availability and usages of advanced parallel capillary sequencing machines such as the ABI 3730 / 3730xl Genetic Analyzer, Beckman CEQ 8000 Genetic Analyzer and MegaBACE 4000 DNA Sequencers. Recent advances in microfabrication have resulted in further throughput improvements of

[*]Address correspondence to this author at the Institute for Systems Biology, Seattle, WA, 98021 USA; Tel: (206) 732 1297; Fax: (206) 732-1299; blin@systembiology.org .

Sanger sequencing by multiplexing and miniaturization. These advances have been reviewed by Metzker [4].

Despite many improvements in the past decades, the gel-based Sanger sequencing technology has limitations in cost and throughput. To achieve a much higher sequencing throughput requires technological revolution; numerous commercial companies and scientific labs have come up with many different ways of achieving ultra high throughput sequencing with reasonable cost. The technologies, named together as the next-generation sequencing technologies, include: sequencing by synthesis; sequencing by ligation; sequencing by hybridization; single molecule DNA sequencing; nanopore sequencing and polony sequencing. Next generation sequencing technologies offer rapid, low-cost sequencing that have revolutionary impacts on genetic applications including: metagenomics, comparative genomics, high-throughput polymorphism detection, analysis of small RNAs, mutation screening, transcriptome profiling, methylation profiling, and chromatin remodeling. Many different technology platforms and sequencing principles co-exist. The major measurements for the success of the next generation technology are: sequence (read) length, sequence quality, throughput and cost. This review is not intended to cover all inventions in this area, we will focus on novel patents that can improve one or more of the four success measurements previously mentioned.

SEQUENCING BY SYNTHESIS

Sequencing by synthesis using fluorophore-labeled, reversible-terminator nucleotides is the most common platform of sequencing by synthesis. It is sometimes named “fluorescent *in situ* sequencing” (FISSEQ). It usually involves these following steps: attaching the DNA to be sequenced in a solid surface, then adding polymerase and labeled nucleotides with cleavable chemical group to cap an-OH group at a 3′-position of the deoxyribose so that incorporation of the nucleotides terminates the reaction. The sequence can be read from the labels used for nucleotides. After washing and un-caping the 3′-OH group at a 3′-position of the deoxyribose, the sequencing cycle continues. In this procedure, each base was detected while the DNA chain is growing, thus eliminating the need for gel-based size-fractionation, and providing much higher throughput.

The Pyrosequencing technology is another SBS technology developed by Ronaghi *et al*. at Stanford University [5,6]. In brief, it is based on the detection of pyrophosphate (PPi) released during DNA synthesis when inorganic PPi is released after nucleotide incorporation by DNA polymerase. The released PPi is then converted to ATP by ATP sulfurylase. A luciferase reporter enzyme uses the ATP to generate light, which is then detected by a charge coupled device (CCD) camera. The light signal is proportional to the number of nucleotides incorporated (e.g. A, TT, CCC etc) and because the G, A, T, and C nucleotides are added step-wise in a sequencing cycle, the DNA sequences are easily derived. The discovery of apyrase has given pyrosequencing momentum for its implementation as a high throughput sequencing strategy. Apyrase continuously degrades unin-corporated dNTPs and excess ATP, eliminating the need for a washing step between nucleotides in a pyrosequencing cycle [6]. Pyrosequencing has evolved into an ultra-high throughput sequencing technology with the combination of several technologies such as template carrying microbeads deposited in microfabricated picoliter-sized reaction wells connecting to optical fibers [7].

In pyrosequencing, luciferase was used convert luciferin to oxyluciferin using the ATP, thus emitting a photon for detection. However, luciferase is not absolutely specific for ATP, it can also use dATP, and to a lesser extent other nucleoside triphosphates, resulting in high background and non-specific reactions (sequence errors). A method to remove the cause of

background light causing problems in pyrosequencing was recently invented by Kanbara at Hitachi Ltd. and Kikkoman Corp., Japan (JP2007097471A2) [8]. In this method, the step to synthesize ATP in the standard pyrosequencing was changed to synthesizing ATP by reacting pyrophosphoric acid with pyruvate phosphate dikinase with 30-800 uM AMP.

454 Life Sciences/Roche Diagnostics has the Genome Sequencer 20 System and the Genome Sequencer FLX System, two high-throughput commercial sequencing platforms based on the pyrosequencing technology. In this platform, DNA helices are fractionated into 300-500 bp fragments and linkers are added to their 3′ and 5′ ends. Single stranded DNA helices are isolated and captured on beads. The beads with DNAs are then emulsified in a 'water-in-oil' mixture with amplification reagents to create microreactors for emPCR (Emulsion PCR). Finally, beads with amplified DNAs are loaded onto a PicoTiterPlate (allowing only one bead per well) for sequencing. 200,000 DNA sequences can usually be obtained in a single pyrosequencing run, but the major advantage of the technology is its long sequence read length (>100 bases), which enables extremely reliable mapping to the transcriptome. Several recent inventions by the 454 Life Sciences Corporation were employed in commercial platform including US patent No. 7211390; 7244559; 7264929 [9-11]. These inventions proposed sequencing nucleic acids attached to solid substrates. The substrate can be a fiber optic surface that is cavitated, e.g., by etching the core of a fiber optic. The invention also proposed an apparatus that include a reagent delivery chamber (the perfusion chamber), a conduit, an imaging system such as a fiber optic system, and a data collection system.

To obtain enough signals for the detection of nucleotide additions (e.g. by fluorescence), it is useful to pre-amplify the DNA to be sequenced. DNA can be amplified *in situ* after they bind to a solid support. Adessi *et al*. at Solexa Ltd (UK) invented a method of solid-phase nucleotide acid amplification, immobilised nucleic acids array fabrication, and sequencing [12]. In their method, a large number of distinct nucleic acid sequences can be arrayed and amplified simultaneously at a high density. This invention is implemented as the "Clonal Single Molecule Array technology". Using "bridge amplification" in a flow cell to generate DNA clonal clusters, it greatly simplifies the template-amplification process and increases the sequencing reaction signal (A detailed technology description is available at http://www. illumina.com/).

Solexa (recently merged with Illumina Inc.) has a high throughput sequencing platform, the Genome Analyzer system, which is based on a SBS technology [13,14]. In its current implementation in the Genome Analyzer system, DNA 'clone' are amplified directly in the planar surface of the flow cells. Alternatively, DNA can be bound to microparticles, and the microparticles loaded into the flow cells as Bridgham *et al*., described in their recent patent [15]. However, the density (therefore throughput) of this technology is lower than the "Clonal Single Molecule Array" technology that is currently implemented in the Soelxa/ Illumina's Genome Analyzer system.

Intelligent Bio-Systems Inc. is also developing a next-generation technology. Their method is based on a four-color DNA sequencing-by-synthesis method using cleavable fluorescent nucleotide reversible terminators [16,17] that was developed at Columbia University. In their SBS, they have developed and used chemically-cleavable, fluorescent-nucleotide reversible terminators structured as such: 3′-O-allyl-dNTPs-allyl-fluorophore.

Hardin *et al*. filed several patents regarding engineering of a polymerase and/or dNTPs with atomic and/or molecular tags for high throughput sequencing [18-20]. The engineered nucleotide triphosphate probes contain a mole-cular and/or atomic tag on a gamma and/or beta phosphate group and/or a base moiety. In this technology, the interaction between a

fluorescently-tagged polymerase and a fluorescently modified nucleotide is exploited for real-time sequencing by synthesis. During the sequencing-by-synthesis reaction, when a nucleotide is incorporated, energy transfers from the polymerase to the nucleotide by Fluorescent Resonance Energy Transfer (FRET) resulting in the emission of a base-specific incorporation signature that can be detec-ted real time. Visigen Biotechnologies Inc. is comer-cializing this technology.

Affymetrix Inc. is also entering the field of next-generation sequencing with their recent patent filing “Sequencing of surface immobilized polymers utilizing microfluorescence detection” [21]. The invention covers means for parallel-sequence analysis of a large number of biological polymers such as DNA including the apparatus and reagents and methods for sequencing using fluorescent labels for templates immobilized at defined positions on a solid substrate.

NOVEL NUCLEOTIDE ANALOGUES OR MODIFIED NUCLEOTIDES FOR SBS

In sequencing by synthesis (SBS), it would be advantageous to be able to add one nucleotide at a time using reversible blocker or terminator. In a SBS reaction, continued sequencing requires 3′-OH protecting groups be removed for the next round of sequencing reaction. An ideal blocker or reversible terminator should not interfere with incorporation of the nucleotides, i.e. be effective substrates for DNA polymerases; and should be de-blocked efficiently. Different blocking and de-blocking chemistries have recently being developed. Kwiatkowski at Helicos Biosciences Corp. invented a hydrocarbyldithiomethyl-modified compound that is useful for protecting and/or blocking hydroxyl groups in nucleotides and will be useful in SBS based sequencing [22]. Parce *et al*. at Caliper Life Sciences developed nucleotide analogues comprising 3′-blocking groups for reversible chain-termination for sequencing by synthesis [23]. Their blocking groups include phosphate groups and carbamate groups. Sequencing can be performed using fluorescent nucleotides incorporation and detection, and then followed by de-blocking and additional cycles of sequencing. Optionally photobleaching the fluorescent label after detecting an added fluorescent nucleotide can be used, or intercalating dyes can be used to detect addition of a non-labeled nucleotide during sequencing. Ju and Itagaki *et al*. [17,24] developed nucleotide analogues that have unique labels attached to them through a cleavable linker. A cleavable chemical group $-CH_2OCH_3$ or $-CH_2CH{=}CH_2$ is used to cap the −OH group at the 3′-position of the deoxyribose. Barnes *et al*. at Soelxa Limited also developed nucleosides and nucleotides that are linked to detectable labels via a cleavable linker group [25]. They came up with many different linkers such as disulfide linkers, acid labile linkers, Sieber linkers, indole linkers and t-butyl Sieber linkers. These linkers can be electrophilically cleavable linkers (cleaved by protons), nucleophilically cleavable linkers, or photocleavable linkers etc.

Hoser filed a patent application [26] for a method of sequencing using nucleotide analogues that are capable of binding, but not incorporated into, to the active site of the polymerase and to complementary bases in the DNA being sequenced. For example, native nucleotides can be mixed with α-β methyl DNTP that can compete with the native nucleotides but which are non-incorporable. Each nucleotide (G, A, T, C) analogue is labelled with different intercalator dyes so that each fluoresces at a different wavelength when binding to the polymerase-primer-template complex. As the nucleotide analogue cannot be incorporated, it eventually leaves the complex and a native base binds and then is incorporated. Therefore the enzyme can move along the template and sequencing can continue.

Other recent inventions of novel nucleotide analogues and modified nucleotides include chemically-cleavable 3′-oallyl-DNTP-allylfluorophore fluorescent nucleotide analogues

(WO07053719A2), a reversibly-terminating nucleotide modified at the 2′ position (WO07075967A2), four color 3′-o-allyl modified photocleavable fluorescent nucleotides (WO07053702A2), photocleavable nucleotides [27] and 3′-O-modified nucleotides [28].

Other than developing better and more efficient nucleotide analogues as reversible terminators, the polymerase itself can also be engineered to better and more efficiently incorporate modified nucleotide analogues [29,30]. For example, Smith *et al*. developed a modified polymerase enzymes that exhibit improved incorporation efficiency for nucleotide analogues that are modified at the 3′ position of the sugar moiety group and are larger in size than the naturally occurring 3′ hydroxyl group [29].

SEQUENCING HOMOPOLYMER

One problem in next generation sequencing is the sequencing of homopolymer regions, i.e. a sequence of the same nucleotide runs. When a stretch of the same base is encountered in pyrosequencing, the polymerase continues adding nucleotides until the end of the stretch. The increase in light emission is related to the number of nucleotides in the stretch (e.g. from A to AA), but as the length of the stretch increases, the accuracy of the estimate decreases (e.g for TTT vs TTTT). Fuller at GE Healthcare Bio-Sciences Corp. invented a method that improves the pyrosequencing technology by overcoming these limitations [31]. In his invention, Fuller developed a method to form a closed complex of nucleic acid template, nucleotide and polymerase without divalent metal ions. This closed complex traps the nucleotide complementary to the next template nucleotide to be sequenced and technically ‘freezes’ the polymerase, and sequencing is paused. Detection of the trapped nucleotide allows determination of the next nucleotide in the sequence. In this way, the formation of the closed-complex system allows sequencing of a single base at a time, even when a homopolymer region is encountered.

Recenly, Wu *et al*. reported the development of a method to sequence homopolymer regions in pyrosequencing using nucleotide reversible terminators [28]. They used an allyl or a 2-nitrobenzyl group as the reversible moiety to cap the 3′-OH of the four nucleotides and synthesized 3′-O-allyl-dNTPs and 3′-O-(2-nitrobenzyl)-dNTPs as reversible terminators for pyrosequencing. During pyrosequencing, only one nucleotide is incorporated at a time even in homopolymeric regions. The capping moiety on the 3′-OH of the DNA extension product is efficiently removed after PPi detection by either a chemical method or photolysis.

In sequencing homopolymer regions using incorporation of fluorescently-labeled nucleotides in the sequencing reactions, conventional optics are not able to reliably distinguish one from two identical bases, and sometimes two from three base incorporations, which resulted in the failure to sequence homopolymer regions. Lapidus *et al*. at Helicos Bioscience Inc. invented a method for controlling at least one parameter of a nucleotide extension reaction in order to regulate the rate at which nucleotides are added to a primer, thus increasing the resolution and reliability of single molecule sequencing [32]. The invention also improves the imaging systems in accurately resolving a sequence at the single-molecule level. This invention allows the determination of the number of nucleotides in a homopolymer sequence.

Church *et al*., [33] invented a method named “WOBBLE SEQUENCING” and associated regents that are useful for sequencing homopolymeric regions of DNA. The method also prevents the accumulation of mistakes and inefficiencies in the sequencing reaction.

PATENTS FOR PAIR-END SEQUENCING

It will be advantageous and often necessary to generate sequence data from both ends (nick-named pair-end, dual-end or double ended sequencing) of a template DNA fragment to confirm or help shotgun sequence assemblies. Pair-end sequencing will also be useful for characterization of genomic rearrangement and insertion and deletion, such as in cancer genome characterization.

Chen *et al*. at the 454 Life Science Corp. invented a method named 'Double ended sequencing' [34]. In their method, sequencing can be initiated from multiple primers that are hybridized using a single primer hybridization step. Two or more sequencing primers are hybridized to the template DNA but all the sequencing primers, except one, are then blocked. Sequencing can be initiated from the unblocked primer. After finishing the sequencing from the 1st unblocked primer, one of the blocked primers is unblocked and sequencing is initiated from the newly unblocked primer. This process is repeated until all the sequencing primers are de-blocked and sequenced.

Smith *et al*. invented methods for producing a paired tag from a nucleic-acid sequence and methods for generating a paired-end library of DNA fragments for sequencing using asymmetrical adapters [35,36]. In one embodiment, the asymmetrical adapters are comprised of an asymmetrical oligonucleotide adapter with a single-stranded 3′ overhang and, a second asymmetrical double-stranded oligonucleotide adapter with a single-stranded 5′ overhang and more than one blocking group (double-stranded) 3′ inside of the 5′ overhang to form so-called "tail adapters". In other embodiments, the asymmetrical adapters can be created to form so-called " Y adapters" or "bubble adapters" depending on the different configurations. They can be used to generate and amplify paired ends from the ends of DNA ranging in size from 2-200 kbs.

Berka *et al*. also developed a method of obtaining a DNA construct comprising two ends from a large piece of DNA [37]. Sheared DNAs were ligated to "capture element oligonucleotides". In one method, a "hairpin adaptor" can be used as the capture element. The " hairpin adaptor" comprises four segments of nucleotide sequences: the hairpin region, restriction endonuclease recognition site, a biotinylated region, and a type IIS restriction endonuclease recognition site.

OTHER IMPROVEMENTS IN SBS

Methods of sequencing nucleotides without amplification have also been invented. Stemple and Armes developed a method [38] for sequencing without the need for amplification. In their invention, polymerases are immobilized onto a solid support. Sequencing and detection are carried out by adding to a reaction chamber a solution that contains a nucleic-acid sample, oligonucleotide primers and all four labeled nucleoside triphosphate terminators.

The accuracy of sequencing in the SBS is also under consistent improvement. Misincorporation can happens during nucleotide synthesis, especially when nucleotide analogues or modified nucleotides are used in SBS. Harris at Helicos Bioscience Corp. invented a method of increasing accuracy of nucleic acid sequencing [39].

POLONY SEQUENCING

George Church's laboratory at Harvard University has developed a high throughput technology based on polony amplification and FISSEQ [40]. Polony amplification is a method to amplify DNAs *in situ* on a thin polyacrylamide film [41]. The DNA movement is limited in the polyacrylamide gel, so the amplified DNAs are localized in the gel and form

the so-called “polonies”, polymerase colonics. Up to 5 million polonies (i.e. 5 million PCRs) can form on a single glass microscope slide. This sequencing technology has been tested in a bacterial genome and the sequence read length was found to be about 13-bases per colony [42]. There are current attempts to improve this read length to 17-18 bases. Recently, Church *et al*. filed a patent for polonyfluorescent-in situ-sequencing beads [43]. They provided a method for preparing and arraying large numbers of microspheres bearing amplified DNA. In one embodiment, PCR reactions in water-in-oil emulsions are used. The beads can then be enriched, arrayed and immobilized onto a glass surface to which the acrylamide gel was coated. Theses beads can be used in parallel nucleic-acid analysis such as DNA sequencing and RNA expression profiling. An epifluorescence microscope can be used to image the individual beads during successive sequencing reactions in which fluorescent nucleotides are used. Kim *et al*. developed a technology named PMAGE (for “polony multiplex analysis of gene expression”), which combines polony amplification and a sequence-by-ligation method, to sequence 14-base tags [44]. Up to 5 million polonies can be sequenced in parallel.

SINGLE MOLECULE DNA SEQUENCING

Most current sequencing technologies are based on sequencing many identical copies of DNA molecules (often amplified). There are potential problems associated with sequencing amplified multiple copies of identical DNAs such as: How to achieve synchronous priming of each copy of the multiple DNAs by the sequencing primers? How to solve the problems with phasing that resulted from a loss in the synchronization of different copies of DNAs during successive sequencing reaction steps? If some of the molecules fail to incorporate a nucleotide in a sequencing step, subsequent sequencing reactions and results obtained for these molecules will no longer be in synchronization with results obtained for the other molecules. This problem can increase through successive sequencing steps, resulting in ambiguous sequencing results and a high number of sequencing errors.

One way to solve these problems is to sequence a single DNA molecule. Helicos Biosciences Corp. developed true single molecule sequencing technology based on sequencing by synthesis method [32]. DNA can be attached to solid support to form single molecule arrays, and the single DNA molecule be sequenced directly. Buzby at Helicos Biosciences Corp. also invented a method of stabilizing a nucleic acid duplex on a surface for single molecular sequencing [45]. In another patent application filed by Sun at Applera Corp. [46], fluorescent intercalators are employed as a donor in fluorescence resonance energy transfer (FRET) for use in single molecule sequencing reactions.

Woudenberg and Taing invented [47] systems and methods that are capable of single molecule sequencing. They invented an apparatus that comprises a sample holder configured to hold a single-molecule. A light source is used to illuminate a confined single molecule for detection. A second light source is used to produce light pulses for controlling the timing of incorporation events so that one nucleotide base is incorporated at each single molecule DNA per each light pulse. A labeled and modified nucleotide, which comprises a nucleotide, a fluorescent label and a photocleavable linker between the nucleotide and the fluorescent label, are used for sequencing. The photocleavable linker is selected as it can be cleaved by light after the nucleotide is incorporated to release the linker and the fluorescent label, thus allowing the next incorporation event to occur.

Hardin *et al*. filed an application for single molecular sequencing using engineered polymerase and labeled dNTPs that can be detected when they are incorporated [48], and achieve directly readout the exact base sequence of an RNA or DNA molecule. They also proposed an alternative way of sequencing, using depolymerizing agents such as

exonucleases for sequencing a polymer such as DNA by depolymerization instead of polymerization. The final goal is to improve fluorescent molecule chemistry, computer modeling, base-calling algorithms, and genetic engineering of biomolecules, especially for real-time or near real-time sequencing.

NANOPORE SEQUENCING

Nanopores are nanometer-scale pores and are touted as one of the most promising technologies in achieving true real time, ultra-fast, true single molecule DNA sequencing [49]. Nanopore have been used for the detection, counting and characterization of single molecules by observing the changes in ionic current in a nanopore when molecules traverse through a nanopore [50,51]. Recent advancements in nanofabrication have made it possible to fabricate pores at the nanometer scale, i.e. the scale of nucleotides. The fabrication of nanopores such as the alpha-hemolysin pore and synthetic nanopore was reviewed recently [52] and will not be covered here.

Attempts to use nanopores for DNA sequencing usually involves the measurement of the so called 'blockade current', i.e the changes in ionic current across the nanopore while a DNA molecule is pulled through a nanopore using electric field force [53]. However, because signal difference between each of the different nucleotides is not big enough when a DNA molecule passes through a nanopore, obtaining individual nucleotide sequences has proven difficult.

Recent progress has made a big step toward ultrafast sequencing using nanopore technologies. Lagerqvis *et al*. [54] proposed a novel idea to measure the electric current perpendicular to the DNA backbone. They envisioned embedding electrodes in the walls of nanopores; however, this is difficult to reduce to practice. They calculated that in theory it would be feasible to perform ultra-fast DNA sequencing based on the distributions of transverse electrical currents when a DNA molecule passes through a nanopore. Zhao *et al*. reported that a single nucleotide polymorphism can be detected by a change in the threshold voltage of a nanopore [55].

The most interesting invention using nanopore for DNA sequencing came from Ling and Bready's invention of a "hybridization assisted nanopore sequencing" technique [56]. In one embodiment of the method, DNA to be sequenced is first hybridized to a probe library (e.g. 4096 possible six-mers). The hybridized probe and DNA is then passed through the nanopore, the fluctuations in current measured across the nanopore are recorded. These fluctuations in current as a function of time are then used to determine the attachment positions of the probes along the DNA. These probe position information is calculated using a computer algorithm to derive the sequence of the DNA as the probe sequences are known. Ling continued to develop methods of using solid state nanopores for DNA sequencing and methods to make electrically-addressable nanopore Array [57]. Nabsys Inc. is commercializing this technology.

Joyce and Dowell at Agilent Technologies Inc. [58] also invented devices, systems, and methods for nanopore analysis of polymers such as DNA. A set of resonant tunneling electrodes are placed adjacent to a partitioned nanopore, and changes in tunneling current as monomers of a polymer (e.g. each base of a DNA) sequentially travel through a partitioned nanopore can be recorded. Barth and Myerholtz at Agilent Technologies Inc. [59] invented a nanopore device for DNA sequencing that comprises a fluid barrier with a nanopore with its top inner diameter smaller than its bottom inner diameter, and containing perimeter electrodes that are separated by insulator element and biopolymeric channel. Wu at the Hewlett Packard company invented a molecular analysis device that comprises a molecule sensor and a nanopore for DNA sequencing [60]. The molecule sensor can be a single

electron transistor or a nanowire with two termini for detecting signals. The molecule sensor can measure the electronic effect such as the electrical current change responsive to a molecule passing through a nanopore.

SEQUENCING BY HYBRIDIZATION

Sequencing by hybridization (SBH) is an old concept which dates back to the late 80's [61,62]. It is a method of reconstructing a DNA sequence based on its k-mer content. All possible k-nucleotide oligomers (k-mers) are hybridized to identify overlapping k-mers in an unknown DNA sample. These overlapping k-mers are subsequently aligned by algorithms to produce the DNA sequence. Drmanac *et al*. reviewed the technology of sequencing by hybridization [63]. The use of traditional SBH alone in *de novo* DNA sequencing has not yet come to full fruition due to the following problems: there is a loss of unique reconstructability of target DNA with > 200 bases in length, it becomes computationaly expensive and complex to analyze sequencing data of longer DNA pieces, and errors in microarray-hybridization image analysis persist.

However, using SBH with pre-defined probe sets derived from a known sequence has been used to resequence a specific region of genomic DNA or cDNA for the identification of small deletions, insertions, polymorphisms and mutations [64]. Affymetrix and Perlegen have microfabricated arrays of immobilized oligonucleotide probes for such applications. Although traditional SBH has not proven its utility in large-scale DNA sequencing, sequencing by hybridization in combination with use of DNA ligase was developed to perform large-scale and high-throughput DNA sequencing (see the section on sequencing by ligation). In addition, there are a few recent improvements that may improve the applicability of using traditional SBH for *de novo* high-throughput DNA sequencing. Sosnowski *et al*. [65] invented a method for electronically stabilizing the hybridization of nucleic acids bound to a microelectronic device using zwitterionic buffers. Jacobson invented a method of nucleotide sequencing via repetitive single-molecule hybridization [66]. Yakhini at Agilent Inc. [67] invented a method and system for sequencing nucleic-acid molecules using sequencing by hybridization and determination of "decoration patterns" using nanopore-based methods. The "decoration pattern" is the change in current recorded over time as a DNA fragment passes through a nanopore aperture.

SEQUENCING BY LIGATION

The first high-throughput sequencing by ligation was probably realized by Lynx Therapeutics Inc. (later merged with Solexa Inc. and now part of Illumina Inc.) in their early technology of Massively Parallel Signature Sequencing [68]. They patented a method of massively parallel signature sequencing by ligation of encoded adaptors [69]. Church's laboratory at Harvard Medical School also advanced the sequencing by ligation method for ultra fast and high throughput sequencing [42,44]. They converted an epifluorescence microscope for rapid DNA sequencing. DNA molecules were amplified in parallel onto microbeads by emulsion polymerase chain reaction. Millions of beads were immobilized in a polyacrylamide gel and sequenced using sequencing by ligation method. Applied Biosystems Inc. acquired Agencourt Personal Genomics who developed the SOLiD (Supported Oligo Ligation Detection) System for high-throughput DNA sequencing. The most recent version the technology is based on a similar sequence-by-ligation method in combination with emulsion PCR developed at Harvard. In brief, DNA fragments are amplified by emulsion PCR as we described earlier and captured on beads. The beads are then loaded on a glass surface to form a random array for sequencing. Sequencing primers are added together with four oligo probes with different fluorescent labels (colors). Each oligo probe is eight-bases long and the middle 2 bases (4^{th} and 5^{th} position) are encoded.

After hybridization, ligation and detection, sequences at the 4^{th} and 5^{th} position are determined. Repeating this process leads to the determination of dinucleotide sequences every five bases (e.g., at position 4, 5,…9, 10,…). After 5 cycles, the sequencing reaction is reset by denaturing the DNAs. Another sequencing primer that is offset by one base (n-1) is then added and the same sequence-by-ligation processes are repeated. The dinucleotide sequences at position 3, 4, …8, 9,… are determined. Repeated sequencing cycles with sequencing primers that offset 2, 3, 4 bases (n-2, n-3, n-4) eventually lead to complete sequencing of the DNA (for technical details, please see the company's web site).

OTHER NOVEL SEQUENCING TECHNOLOGIES

Other novel technologies have been pursued for next-generation sequencing. We will not do a comprehensive coverage here as most of them are still in the early stages of development. Nano-transistor array based devices, such as Carbon NanoTube Field Effect Transistor (CNTFET), have been proposed for ultra-fast DNA sequencing. Huang and Tsai have claimed development of a method that uses nano-transistors for ultra-rapid DNA sequencing [70]. In this method, DNA molecules are stretched and driven over nanotubes by micro-fabricated electrodes. DNA molecules come into contact with the carbon nanotube surface base after base, and the difference in current flow from each base is generated due to charge transfer between the DNA molecule and the nanotubes. These differences are recorded, and DNAs are thus sequenced. This technology is similar to the aforementioned nanopore sequencing; however, it may overcome some problems associated with nanopore sequencing such as that of efficiently stretching the curly DNA molecules through the nanopores and that of preventing interaction of DNA with the nanopore-channel proteins (if protein nanopores are used).

Scanning tunneling microscopy (STM) has also been proposed for DNA sequencing. STM uses a piezo-electric-controlled sharp probe that performs a raster scan of a specimen to form images of its surface [71]. STM has been used to image the physical properties of single DNA molecules [72,73]. Recently, Zorn claimed development of a method for generating coherent electron tunneling imaging and spectroscopy by integrating scanning tunneling microscope with an actuator-driven flexible gap. This device is a coherent electron quantum interferometer for scanning-probe microscopy when combined with a nanomanipulator such as nanotweezers using Raman optical and mass spectroscopic means and can be used to sequence DNAs [74].

Li *et al*. at the Hewlett Packard company recently filed a patent application for a molecular-analysis device which is comprised of a nanowire-molecule sensor that detects the interactions of the nitrogenous material disposed on the nanowires and molecules such as DNAs. A molecule guide is configured for guiding a molecule near the molecule sensor for an interaction to occur and be detected [75].

CURRENT & FUTURE DEVELOPMENTS

We will continue to see improvements in next-generation sequencing technologies. The quest for 'sequencing a human genome under $1000' and contest for the X prize '$10M to the First Team to Sequence 100 Human Genomes in 10 Days' (http://genomics.xprize.org/) have added to the fanfare to develop and improve the next generation of sequencing technologies. For the more mature next-generation sequencing technologies, improvements will need to focus on improving throughput and sequence length, for example, achieving longer sequence reads for the Solexa/Illumina's (currently up to 35 nucleotides) and ABI/Agencourt's (up to 50) technology platforms and achieving higher throughput for the Roche/454 Life Sciences's platform. Other critical areas of needed improvement are simplifying sample preparation, increasing sequencing accuracy and eliminating DNA manipulation

steps such as amplifications. Also, technology that is able to perform whole-genome sequencing from a single cell will also be important. Genomic sequencing from a single microbial cell can be achieved using DNA amplified by the Multiple-Displacement-Amplification (MDA) reaction [76]. However, we have yet to see that this is possible for more complex genomes such as that of the human (some of the genomic sequences were lost after amplification). A sequencing technology obviating the need for template amplification will be critical to achieve single-cell genome sequencing. Continued technology improvements in single-molecule sequencing, as pioneered by Helicos Inc. using single-molecule imaging and Nabsys Inc. using “hybridization assisted nanopore sequencing”, will take the next-generation sequencing to a new level in throughput. The application and development of next-generation sequencing based on other novel technologies such as nano-transistor arrays, scanning probe microscopy and nanowire molecule sensors is still at the proof-of-principle stage, and further technology improvements are necessary.

We are at the verge of a new genomic revolution with recent advances in next-generation sequencing technologies. We will see enormous impacts of these next-generation sequencing methods in dealing with complex biological problems, for example, the identification of all sequence changes in drug-resistant HIV and drug resistance tuberculosis bacteria (TB), in the delineation of sequence changes for individual cells during cancer initiation and progression, and in global transcription-factor mapping (CHIP to sequencing) [77]. In addition, next-generation sequencing technology will hasten the arrival of personalized medicine. It will make routine sequencing of a person’s genome a reality, as evidenced by the recent release of the genome sequence of an individual human [78] and the initiation of the Personal Human Genome Project. It will have a tremendous impact on the health care system as well as the way medicine is practiced in the future.

REFERENCES

[1]. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA. 1977; 74:5463–5467. [PubMed: 271968]

[2]. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol. 1975; 94:441–448. [PubMed: 1100841]

[3]. Maxam AM, Gilbert WA. New method for sequencing DNA. Proc Natl Acad Sci USA. 1977; 74:560–564. [PubMed: 265521]

[4]. Metzker ML. Emerging technologies in DNA sequencing. Genome Res. 2005; 15:1767–1776. [PubMed: 16339375]

[5]. Ronaghi M, Karamohamed S, Pettersson B, et al. Real-time DNA sequencing using detection of pyrophosphate release. Anal Biochem. 1996; 242:84–89. [PubMed: 8923969]

[6]. Ronaghi M, Uhlen M, Nyren PA. Sequencing method based on real-time pyrophosphate. Science. 1998; 281:363, 365. [PubMed: 9705713]

[7]. Margulies M, Egholm M, Altman WE, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005; 437:376–380. [PubMed: 16056220]

[8]. Kanbara, H.; Shu, K.; Kajiyama, T.; Suzuki, S. JP097471. 2007.

[9]. Rothberg, JM.; Bader, JS.; Dewell, SB.; McDade, K.; Simpson, JW.; Berka, J.; Colangelo, CM. US20077211390. 2007.

[10]. Rothberg, JM.; Bader, JS.; Dewell, SB.; McDade, K.; Simpson, JW.; Berka, J.; Colangelo, CM. US20077244559. 2007.

[11]. Rothberg, JM.; Bader, JS.; Dewell, SB.; McDade, K.; Simpson, JW.; Berka, J.; Colangelo, CM. US20077264929. 2007.

[12]. Adessi, C.; Kawashima, E.; Mayer, P.; Mermod, J-J.; Turcatti, G. US20067115400. 2006.

[13]. Bentley DR. Whole-genome re-sequencing. Curr Opin Genet Dev. 2006; 16:545–552. [PubMed: 17055251]

[14]. Balasubramanian, S.; Klenerman, D.; Barnes, C.; Osborne, MA. US20077232656. 2007.

[15]. Bridgham, J.; Corcoran, K.; Golda, G.; Pallas, MC.; Brenner, S. US20077282370. 2007.

[16]. Ju J, Kim DH, BiL, et al. Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. Proc Natl Acad Sci USA. 2006; 103:19635–19640. [PubMed: 17170132]

[17]. Ju, J.; Li, Z.; Edwards, JR.; Itagaki, Y. EP1790736. 2007.

[18]. Hardin, SH.; Gao, X.; Briggs, J.; Willson, R.; Tu, S. US20070172869. 2007.

[19]. Hardin, SH.; Gao, X.; Briggs, J.; Willson, R.; Tu, S-C. US20070172860. 2007.

[20]. Hardin, SH.; Gao, X.; Briggs, J.; Willson, R.; Tu, S-C. US20070172819. 2007.

[21]. Dower, W.; Fodor, S. US20070105131. 2007.

[22]. Kwiatkowski, M. US20077279563. 2007.

[23]. Parce, JW.; Nikiforov, TT.; Mehta, TB.; Kopf-Sill, AR.; Chow, AW.; Knapp, MR. US20067105300. 2006.

[24]. Itagaki, YJJ.; Englewood, CLZ.; Edwards, JR. EP1337541. 2007.

[25]. Barnes, C.; Balasubramanian, S.; Liu, X.; Swerdlow, H.; Milton, J. US20067057026. 2006.

[26]. Hoser, MJ. US20070148645. 2007.

[27]. Olejnik, J.; Krzymanska-Olejnik, E.; Rothschild, K. US20060252923. 2006.

[28]. Wu J, Zhang S, Meng Q, et al. 3′-O-Modified nucleotides as reversible terminators for pyrosequencing. Proc Natl Acad Sci USA. 2007; 104:16462–16467. [PubMed: 17923668]

[29]. Smith, GP.; Bailey, DMD.; Sanches, RM.; Swerdlow, H.; Earnshaw, DJ. US20060240439. 2006.

[30]. Barr, O.; Tobias, W.; Smith, GP.; Balasubramanian, S.; Rigatti, R.; Sanches, RM. US20060281109. 2006.

[31]. Fuller, CW. US20077264934. 2007.

[32]. Lapidus, SN.; Buzby, PR.; Harris, T. US20077169560. 2007.

[33]. Church, GM.; Shendure, J.; Porreca, GJ. US20070207482. 2007.

[34]. Chen, YJ.; Leamon, JH.; Lohman, KL.; Ronan, MT.; Rothberg, JM.; Srinivasan, M.; Weiner, MP. US20077244567. 2007.

[35]. Smith, D.; Malek, J.; McKernan, KJ. US20060024681. 2006.

[36]. Smith, D.; Malek, JA. US20070172839. 2007.

[37]. Berka, J.; Chen, Z.; Egholm, MC.; G.B.; Hutchison, SK.; Leamon, JH.; Sarkis, GJ.; Simons, JF. US20060292611. 2006.

[38]. Stemple, DL.; Armes, NA. US20077270951. 2007.

[39]. Harris, TD. US20077282337. 2007.

[40]. Mitra RD, Shendure J, Olejnik J, et al. Fluorescent *in situ* sequencing on polymerase colonies. Anal Biochem. 2003; 320:55–65. [PubMed: 12895469]

[41]. Mitra RD, Church GM. *In situ* localized amplification and contact replication of many individual DNA molecules. Nucleic Acids Res. 1999; 27:34–37.

[42]. Shendure J, Porreca GJ, Reppas NB, et al. Accurate multiplex polony sequencing of an evolved bacterial genome. Science. 2005; 309:1728–1732. [PubMed: 16081699]

[43]. Church, GM.; Shendure, J.; Porreca, GJ.; Zhu, J. US20070087362. 2007.

[44]. Kim JB, Porreca GJ, Song L, et al. Polony multiplex analysis of gene expression (PMAGE) in mouse hypertrophic cardiomyopathy. Science. 2007; 316:1481–1484. [PubMed: 17556586]

[45]. Buzby, PR. US20077220549. 2007.

[46]. Sun, H. US20070202521A1. 2007.

[47]. Woudenberg, TM.; Taing, M. US20070154921A1. 2007.

[48]. Hardin, S.; Gao, X.; Briggs, J.; Willson, R.; Tu, S-C. US20070184475. 2007.

[49]. Rhee M, Burns MA. Nanopore sequencing technology: research trends and applications. Trends Biotechnol. 2006; 24:580–586. [PubMed: 17055093]

[50]. Meller A, Branton D. Single molecule measurements of DNA transport through a nanopore. Electrophoresis. 2002; 23:2583–2591. [PubMed: 12210161]

[51]. Meller A, Nivon L, Brandin E, et al. Rapid nanopore discrimination between single polynucleotide molecules. Proc Natl Acad Sci USA. 2000; 97:1079–1084. [PubMed: 10655487]

[52]. Rhee M, Burns MA. Nanopore sequencing technology: nanopore preparations. Trends Biotechnol. 2007; 25:174–181. [PubMed: 17320228]

[53]. Heng JB, Aksimentiev A, Ho C, et al. Stretching DNA using the electric field in a synthetic nanopore. Nano Lett. 2005; 5:1883–1888. [PubMed: 16218703]

[54]. Lagerqvist J, Zwolak M, Di Ventra M. Fast DNA sequencing via transverse electronic transport. Nano Lett. 2006; 6:779–782. [PubMed: 16608283]

[55]. Zhao Q, Sigalov G, Dimitrov V, et al. Detecting SNPs using a synthetic nanopore. Nano Lett. 2007; 7:1680–1685. [PubMed: 17500578]

[56]. Ling, XS.; Bready, B. US20070190542. 2007.

[57]. Ling, XS. US20070042366. 2007.

[58]. Joyce, TH.; Dowell, JT. US20070048745. 2007.

[59]. Barth, PW.; Myerholtz, CA. US20060231419. 2006.

[60]. Wu, W. US20070178507. 2007.

[61]. Drmanac R, Labat I, Brukner I, et al. Sequencing of megabase plus DNA by hybridization: theory of the method. Genomics. 1989; 4:114–128. [PubMed: 2737674]

[62]. Khrapko KR, Lysov YP, Khorlyn AA, et al. An oligonucleotide hybridization approach to DNA sequencing. FEBS Lett. 1989; 256:118–122. [PubMed: 2680594]

[63]. Drmanac R, Drmanac S, Chui G, et al. Sequencing by hybridization (SBH): advantages, achievements, and opportunities. Adv Biochem Eng Biotechnol. 2002; 77:75–101. [PubMed: 12227738]

[64]. Schirinzi A, Drmanac S, Dallapiccola B, et al. Combinatorial sequencing-by-hybridization: analysis of the NF1 gene. Genet Test. 2006; 10:8–17. [PubMed: 16544997]

[65]. Sosnowski, RG.; Butler, WF.; Tu, E.; Nerenberg, MI.; Heller, MJ.; Edman, CF. US20070178516. 2007.

[66]. Jacobson, JM. US20077276338. 2007.

[67]. Yakhini, Z. US20060287833. 2006.

[68]. Brenner S, Williams SR, Vermaas EH, et al. *in vitro* Cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs. Proc Natl Acad Sci USA. 2000; 97:1665–1670. [PubMed: 10677516]

[69]. Albrecht, G.; Brenner, S.; DuBridge, RB.; Lloyd, DH.; Pallas, MC. US20006013445. 2000.

[70]. Huang, J-T.; Tsai, C-H. US20060246497. 2006.

[71]. Gerd, B.; Heinrich, R. US4343993. 1982.

[72]. Tanaka H, Kawai T. Imaging and manipulation of DNA molecules using scanning tunneling microscopy. Tanpakushitsu Kakusan Koso. 2000; 45:170–175. [PubMed: 10667073]

[73]. Herrick J, Bensimon A. Imaging of single DNA molecule: applications to high-resolution genomic studies. Chromosome Res. 1999; 7:409–423. [PubMed: 10560964]

[74]. Zorn, MD. US20070194225. 2007.

[75]. Li, Z.; Wang, S-Y.; Wu, W.; Islam, MS. US20060275779. 2006.

[76]. Lasken RS. Single-cell genomic sequencing using Multiple Displacement Amplification. Curr Opin Microbiol. 2007

[77]. Johnson DS, Mortazavi A, Myers RM, et al. Genome-wide mapping of *in vivo* protein-DNA interactions. Science. 2007; 316:1497–1502. [PubMed: 17540862]

[78]. Levy S, Sutton G, Ng PC, et al. The Diploid Genome Sequence of an Individual Human. PLoS Biol. 2007; 5:254.