

Next Generation DNA Sequencing (II): Techniques, Applications

Wilhelm J Ansorge*

Ecole Polytechnique Federal Lausanne, EPFL, Lausanne, Switzerland

*Corresponding author: Wilhelm J Ansorge, Ecole Polytechnique Federal Lausanne, EPFL, Route Cantonale, 1015 Lausanne, Switzerland, Tel: 41795900517; E-mail: wilhelm.ansorge@epfl.ch

Rec date: Jan 2, 2016; Acc date: Jan 30, 2016; Pub date: Feb 2, 2016

Copyright: © 2016 Ansorge WJ. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Development and applications of DNA sequencing techniques in the last years have exceeded by far the expectations. Historical development of the field from the start up to the initial Next Generation systems, technical principles of the platforms existing 6 years ago and many applications, have been described in the previous reviews (e.g. 1,2). In this updated review is discussed the status of the technology in the second half of the year 2015, both of the novel commercially available platforms, and of the systems in development. Mentioned are techniques with potential for future DNA sequencing techniques. Discussed are some of the not solved challenges for the platforms, the price to performance ratio, complexity of sample preparation, and detection of structural variations in genome. The aims for the technology will be to lower the cost of the equipment and biochemicals involved, increasing simultaneously the reproducibility, reliability and simplicity of the techniques and protocols in operation, of importance for diagnostics and applications in the clinics. During the last few years have been rapidly developing technologies and methods that permit analysis of the genome and transcriptome of a single cell. The first observations suggest that both genomic and transcriptomic heterogeneities within an organism are more common than expected, during normal development and disease. Recent public support for precision medicine, for novel therapeutic approaches and efforts for improvements in the healthcare, will be a motivation for further innovations and developments in the field.

Keywords: DNA sequencing; Transcriptome; Single cell; Nanopore

Introduction

From the time of the previous reviews 6 years ago [1,2], DNA sequencing techniques and its applications in the diagnostics field have made significant progress. The innovations are reflected particularly in higher throughput and lower cost per base, e.g. on Illumina and Ion Torrent systems, or the single-molecule real time sequencing approaches of Pacific Bioscience and Oxford Nanopore. The human genome sequence has complemented the understanding of biology, human diversity, and disease. The transition from the first human genome sequences to the personal genomes and genomic medicine has been made possible only because of the advances in DNA sequencing technologies over the past 12 years. NGS utilizes massively parallel sequencing to generate up to several gigabases of sequence information per day, making possible new applications that were in the past difficult to perform. The technique has spread to almost all fields concerning DNA, to such various areas as food control, screening of the purity of vaccines and other medicaments (where it can detect not only traces of the active viral components but also the presence of non-active viral particles), evolution of organisms, analysis of fetus DNA from the mother's blood, forensic field, and others.

Historical development of the DNA sequencing field from the start up to the initial next generation systems, as well as the technical principles of some of the platforms existing already 6 years ago, and many applications, have been described in the previous reviews (e.g. 1,2). In this updated review is discussed the status of the technology in the second half of the year 2015, both of the commercially available platforms, and of the systems in the development (announced to be in an advanced phase close to an introduction on the market). Discussed

are some of the not solved challenges for the platforms, the price to performance ratio, simplicity and reliability of operation, of importance for diagnostics and applications in the clinics. Mentioned are also some techniques reported to be in the initial phase of tests for DNA sequencing.

During the last few years have been rapidly developing technologies and methods that permit analysis of the genome and transcriptome of a single cell. The first observations suggest that both genomic and transcriptomic heterogeneities within an organism are more common than expected, during normal development and disease.

The diagnostics field has already profited from new generations of DNA sequencing technology, sample preparation, single-cell isolation techniques, DNA amplification schemes, and computer data analysis with improving software tools. The applications resulted in better understanding of causes for diseases, in elucidation of reasons why in some persons a drug efficiently treats a disease, while the treatment is not successful in others with the same disease. Recent public support for Precision Medicine, for novel therapeutic approaches and efforts for improvements in the healthcare, will be a motivation for further innovations and developments in the field.

The aims for the technology will be to lower the cost of the equipment and biochemicals involved, increasing simultaneously the reproducibility, reliability and simplicity of the techniques and protocols in operation. The goal to be achieved formulated in [1,2] is still valid, in spite of the great progress in DNA sequencing techniques since then. For genomic sequencing, and for analysis of the ever-more important structural genetic variations in genomes (e.g. copy-number variation CNVs, for chromosomal translocations, inversions, large deletions, insertions, and duplications) it would be a great advantage if sequence read length on the original single DNA molecule could be

increased to tens of thousands of bases and more, in a relatively short time. Ideally, the goal would be the sequence determination of a whole chromosome, from a single original DNA molecule, from a single cell.

Commercially Available Platforms

Since the last reviews [1,2], several new companies entered the Next Generation DNA Sequencing (NGS) field (e.g. Oxford Nanopore, Genia, GnuBio, others), with new devices described below.

Some companies limited the distribution of their systems, or are no longer active in the market. The first was Helicos and although pioneer in the field, its revolutionary Single Molecule Detection technique used in the device was not yet mature enough and its production stopped. The technique was recently re-designed, and a prototype of the new device called GenoCare developed and presented in October 2015 by the Direct Genomics company (see section 3.3 below). The SOLiD system platform from Life Sciences did not find decisive favor with users, possibly because of the somewhat complex biochemical sample preparation, high cost and relatively short reading length around 100 bases. The Roche company stopped the production and distribution of the pyrosequencing bioluminescence detection Gene Sequencer, because of its higher not competitive sequencing cost. It was developed originally by the 454 company, as the first system from the Next Generation devices on the market, and was for several years very successfully applied in many projects. After the attempt to take over Illumina, the Roche company has now initiated collaborations with sequencing platforms of Pacific Biosystems and Genia companies, discussed below.

Illumina platform

The company (www.illumina.com) pushed very successfully its technology developed originally by the Solexa company [1,2], and continuously introduced novel devices both for very high throughput, as well as smaller and more modestly priced versions. At the same time, they improved the reading length from the originally relatively short 50 bases to about 300 bases, improved several times the biochemistry of the sample preparation, and reduced the sequencing cost down to \$0.001 per 1000 bases, which would be not expected few years ago.

The sequencing devices offer starts with the MiSeq series, with speed and simplicity suitable for small genomes, amplicons and targeted gene panel sequencing, and it is also foreseen for applications in clinical research. The system has 1 flow cell, output ranges from 0.3 to 15 Gb, number of reads per flow cell 25 million, the run time 5-55 hours, Maximum Read Length 2x300 bp.

The HiSeq series in the middle range is used for sequencing of genomes, exomes and transcriptomes. As example, the HiSeq 2500 device in the rapid mode has 1 or 2 flow cells per run, the output range 10-300 Gb, run time 7-60 hours, results in 300 million reads per flow cell, and maximum read length 2x250 bp.

The HiSeq series with very high throughput is used in the production and population-scale for whole genomes sequencing. As example, the HiSeq 4000 device has 1 or 2 flow cells per run, the output range 125-1500 Gb, run time 1-3.5 days, 2.5 milliards (in the US 2.5 billions) reads per flow cell, and maximum read length 2x150 bp.

The data flow resulting from the large sequencing throughput, and its analysis, requires great efforts on the bioinformatics part. One of the

causes may also be the still relatively short reads obtainable with the technique.

There are now many Illumina systems purchased by the laboratories worldwide. It is possible that the global sequencing capacity is sufficiently large for present needs, so the next wave of acquisitions may not grow as fast as the previous one, taking into account also the competing systems being introduced in operation, as described below.

Schema of operation, the numerous applications in analysis of gene expression and genomes, relevant publications, collaborations and operation protocols involving the Illumina systems, are listed at the web site of the company (www.illumina.com). Opinion of users can be found on the Genome web site (www.genomeweb.com).

Ion torrent-thermofisher scientific

Ion Torrent Technology has been a continuation of the original pyrosequencing technique, as explained e.g. in [1,2]. In its original design, the 454 company detected the pyrophosphate group released during the incorporation of a nucleotide in the complementary strand, during the DNA polymerization process. The present Ion Torrent system detects the H⁺ ions, another product released during the DNA polymerization. The basic principle, the detection of H⁺ ions, has been licensed from the DNA Electronics company in London. The technology combines semiconductor sequencing technology with biochemistry, enabling the direct translation of chemical information into digital sequence data. This eliminates the need for expensive optics, lasers, and complex sequencing chemistries with fluorescently labeled nucleotides. The result is a sequencing system that is more affordable and cost-effective, faster, scalable and relatively simple to operate.

The Ion PGM System (depending on the Chip used and the application goals) can run up to 5.5 million reads, output reaching to 2 giga bases, reading length up to 400 bases, run time between 2-7 hours.

The Ion Proton System has been designed for high throughput sequencing of exomes, transcriptomes and genomes. With the Ion PI Chip, the throughput is up to 10 Gb, number of reads up to 80 million, read length 200 base, and run time 2-4 hours. It has applications in sequencing of: Genes, Human genomes, De novo sequencing, ChIP technique, whole transcriptomes, exomes, methylation analysis, Gene expression by sequencing, small genomes, and small RNAs.

The NGS devices Ion S5 and Ion S5 XL are being introduced for a simple and rapid workflow for panels, microbes, exomes, transcriptomes, and are applicable both for high and low weekly throughput. The output with Ion 540 Chip is up to 15 Gb (about half with reads 400 bases), number of reads up to 80 million, run time 2.5-5 hours (for reads of 200 or 400 bases). The claim is only 10 ng of low-quality DNA or RNA are needed to generate mutational or gene expression profiles. Target selection is fast and simple using Ion AmpliSeq panels, requiring also only as little as 10 ng of input material. Ion Torrent systems have simplified NGS data analysis with end-to-end bioinformatics solutions, including the Ion Reporter Software and server. Operating software and Plugins are available for analysis of genome sequencing, targeted sequencing, variant analysis and annotation, microbial sequencing, transfer to 3rd party bioinformatics packages, and more.

Because of its affordable price and simplicity of operation, the Ion Torrent systems are widely used worldwide in many projects. Schema of operation, list of publications and collaborations using this system,

are obtainable from the company web page (www.thermofisher.com). Opinion of users can be found on the genome website (www.genomeweb.com).

Pacific biosystems

The Pacific Biosystems has been developing the single-molecule real-time (SMRT) technology for many years. As the technology continuously improved over the years it was successfully attracting an increasing number of users. PacBio Sequencing System may provide very high depth of genetic information through exceptionally long sequencing reads (average >10,000 bp), and the high consensus accuracy.

The basic principle of the technology, explained in (www.pacbio.com), is the detection of the fluorescently labeled nucleotide as it is being incorporated in the complementary strand of the single molecule by the DNA polymerase. In the moment of the incorporation is detected the fluorescent label specific for the base, while the polymerase simultaneously cuts off the label from the nucleotide. The process is repeated for the next labelled nucleotide, and from the order of the 4 different labels detected successively is determined the base sequence. The process takes place in Zero-mode waveguides (ZMW) chambers. Within the ZMWs, single molecule of polymerase is immobilized and fluorescently labeled nucleotide added, so that DNA sequencing can be observed optically and recorded in real time.

In the first PacBio RS II system the SMRT cell contains 150,000 ZMWs, with a throughput of 1 gb per SMRT cell. It has high-performance optics, automated liquid handling, SMRT cells and reagents, instrument control software, run-time from 30 minutes to 6 hours per SMRT cell, run-size from 1 to 16 SMRT cells per run. With this established technology, there are over 850 scientific publications in a variety of research areas, such as human biomedical research, plant and animal genomics, and microbiology. The PacBio RS II has been applied to whole genome sequencing of small genomes, targeted sequencing, complex population analysis, RNA sequencing of targeted transcripts, and microbial genomics.

In autumn 2015 Pacific Biosciences launched a new Sequel single-molecule sequencing system that offers higher throughput and costs, at about half the price of the original RS II device. PacBio developed the Sequel as part of its collaboration with Roche to develop a clinical-grade sequencing system for diagnostic purposes. The device for the Roche sequencing instrument is expected to launch in 2016 for clinical research, and later as an in vitro diagnostics instrument. The Sequel system features redesigned SMRT cells, which each contains 1 million zero-mode waveguides, as compared to the 150,000 ZMWs in the RS II device. The increased number of ZMWs per SMRT cell enables an approximately seven-fold improvement in throughput, so the total will range between 5 gb and 10 gb and the system has an average read lengths between 8 kb and 12 kb. Like the RS II, the Sequel can run between 1 and 16 SMRT cells per run. Sample preparation will be the same on both instruments, taking around six hours to prepare a library. The chemistry is the same as in the RS II, therefore the consensus accuracy will be the same. The sequel is equipped with improved optics, with computer system that processes more data. It supports all applications that were used in the RS II device, including transcript and base modification analysis.

Although the consumables cost per SMRT cell will increase on the Sequel (to around \$700 from about \$250 on the RS II), the scientists

will obtain a seven-fold improvement in overall throughput. Assuming 7 gb of data per SMRT cell, PacBio estimates that sequencing a human genome at 10 coverage-which should be deep enough to do structural variant analysis-would cost around \$3,000. A 30x human genome would cost around \$10,000, and a 50x human genome would cost \$15,000. (For comparison, in the previous RS II system the SMRT cell contains 150,000 ZMWs, with the throughput of 1 gb per SMRT cell, a 30x human genome would cost around \$22,500, while a 50x genome would cost \$37,500).

The price to sequence a human genome on the Sequel will be still higher, compared to sequencing on Illumina, but the cost for computer and analysis time may be somewhat lower due to longer reads and resulting easier assembly. With its long reads, it will be used in projects for rapid and cost-effective generation of whole genome de novo assemblies, particularly of plant genomes, or in those projects analyzing hard-to-sequence regions of the human genome. The diagnostic instrument (in operation in 2016, with Roche collaboration) is planned for the clinical market.

Schema of operation, the numerous applications in analysis of genes and genomes, relevant publications, collaborations and operation protocols involving the PacBio system, are listed at the web site of the company (www.pacbio.com). Opinion of users and latest news can be found on the Genome Web site (www.genomeweb.com).

Complete genomics and BGI

Complete Genomics, founded in 2006, plans to become one of the leaders in whole human genome sequencing. Its technology [3] is based upon shearing the Human DNA into pieces about 400 bases long, then using rolling circle amplification to produce so called Nanoballs, which are then spotted upon a patterned substrate. The patterned substrate with nanoballs is subsequently exposed to a repeated proprietary hybridization technique by ligation with fluorescent labeled probes. From the fluorescence signals resulting from the repeated hybridization on the nanoballs the base sequence is determined. The company has developed for its technique the sequencing instruments, chemistry, and software. Complete genomics has sequenced more than 20,000 whole human genomes over the past five years. In a study published in Nature [4], Complete Genomics demonstrated its long fragment read technology, which enables whole-genome sequencing and haplotyping from 10 to 20 cells with an error rate of one in 10 million bases, which it hopes would enable truly clinical grade genomes.

In March 2013 Complete Genomics was acquired by BGI-Shenzhen, the world's largest genomics services company, with headquarter in Shenzhen, China. It provides comprehensive sequencing and bioinformatics services for commercial science, medical, agricultural and environmental applications, and collaborates globally in many projects with leading research institutions. Prior to the acquisition by BGI, the Complete Genomics Company was focused on selling whole genome sequencing services to over 150 research customers. It focuses now on building a new generation of high-throughput sequencing technology for clinical and consumer applications. In 2015 Complete Genomics produced and introduced the Revelocity system, the end-to-end genomics system for large-scale genomes. The Revelocity system will use a 300-base insert and mate pair sequencing of 28 bases from each end of the insert. The older technology with the four-adaptor construction had mate pair reads of 35 bases [3]. The remainder of the sequencing technology is the same. DNA nanoballs are formed from

the circular pieces of DNA and deposited onto patterned silicon arrays and sequenced by ligation.

Revelocity enables 96 percent of the genome to be covered with an error rate of about 10^{-6} . The system will initially be applied to exome and whole-genome sequencing, but the company plans to develop other applications, like RNA-seq.

In October 2015, BGI has launched a new desktop sequencing platform BGISEQ-500, based on the technology of Complete Genomics. It uses a variation on the "DNA nanoballs" sequencing method, with new engineering features. While Complete Genomics' instruments are applied to sequencing whole human genomes and exomes, the new instrument is more flexible and is planned for RNA sequencing, including in single cells, panels for non-invasive prenatal sequencing, genotyping panels, and other applications. It is a benchtop machine with high-throughput, about 200 gigabases per run. The BGISEQ-500 will also be capable of rapid runs producing as little as 8 gigabases of data, an advantage for clinical assays.

BGI and its customers have generated over 250 publications in top journals. Schema of operation, the numerous applications in analysis of genes and genomes, relevant publications, collaborations and operation protocols involving the Complete Genomics and BGI systems, are listed at the web site of the companies (www.completegenomics.com). Opinion of users and latest news can be found on the Genome Web site (www.genomeweb.com).

Oxford nanopore platform

The principle of the operation of nanopore sequence technique is the analysis of DNA strand directly as the molecule is drawn through a tiny pore suspended in a membrane. Changes in electrical current, or tunneling currents, are used to read off the chain of bases. Great hopes and attention are devoted to the emerging Nanopore technology, because it would allow to sequence Single DNA Molecule in real-time, requiring no amplification step, combined with less expensive hardware and simplicity of operation.

The promise is great, and the progress is finally slowly taking place. When Oxford Nanopore presented the MinIon device in 2012, it was met with slightly pessimistic expectations. In the meantime, many labs worldwide are part of the MinIon analysis and reference consortium (MARC), a group formed by a number of participants in Oxford nanopore's MinIon early-access program (MAP) that plans to conduct a series of projects around the technology. They all wish the technology to be a success, and are assessing how reproducible the device's performance is across laboratories, and developing standard protocols and reference data. The consortium made the raw and aligned nanopore reads of the first test results available recently (corresponding to the state of the MinIon technology in April 2015, www.genomeweb.com, 15.10.2015). The European Nucleotide Archive and the European Bioinformatics Institute coordinated the data distribution and analysis. The goal of the study was to assess the yield, accuracy, and reproducibility of the MinIon by performing replicate experiments at several sites, and to find out which technical factors determine high performance. For the study, the laboratories each generated MinIon sequence data from the same substrain of *E. coli* using the same protocols.

Overall, using R7.3 flow cells and SQK-MAP005 chemistry, a typical experiment resulted in about 20,000 two-dimensional reads, with a median length of 6.5 kilobases, generating about 115 megabases of data. When they used an 8-kilobase shearing protocol, almost 5

percent of 2D reads were at least 10,000 bases long, some of them more than 50,000 bases. The total error for individual 2D base calls was 12 percent, consisting of 3 percent miscalls, further 4 percent were insertions, and 5 percent deletions. A single run yielded enough 2D bases to cover the *E. coli* genome 25-fold.

The performance of the MinIon device itself was consistent in the tests, and no experiment failed due to problems with the device. The researchers also did not observe any GC-bias, although they noted this may be hard to detect with an *E. coli* genome. One reason the quantity and quality of data varied between runs is that many steps in the standard protocol are sensitive to the quality of materials and reagents being used. The results of the MARC report suggest how to improve the performance of the MinIon, including clearer protocol steps, methods for longer library molecules, and improved operation run software scripts.

Overall, MinIon nanopore sequencing has several advantages over short-read sequencing technologies, e.g. its long reads enable sequencing through repetitive regions, its speed generating results in minutes rather than hours, its low capital cost of \$1,000 per device, and its small size. Its main limitations are the currently high mismatch and indel error rate and the limited yield per run, ranging from megabases to gigabases. But both will continue to improve, through better bioinformatic tools and new hardware and chemistry.

Oxford Nanopore recently released the new MinIon Mk1 device and reagents, and the company is constructing systems with larger throughput. But the priority is to concentrate on reproducibility and standardization of the performance.

Detection of structural variants with Oxford Nanopore MinIon device: Nanopore sequencing can detect structural variants in a mixture of PCR amplicons. In a proof-of-principle the scientists led by Jim Eshleman and Winston Timp from Johns Hopkins University, showed (GenomeWeb.com, 5.10.2015) that they could detect a number of well-characterized structural variants-including large deletions, inversions, and translocations-that affect two tumor suppressor genes in pancreatic cancer cell lines, using amplicon sequence data from Oxford nanopore's MinIon. It shows the ability of nanopore sequencing to correctly and reliably detect SV with only hundreds reads, instead of millions of reads.

In the detection of structural variants, Oxford Nanopore is still behind the other technologies that offer long-range genome information, such as Pacific Biosciences, BioNano Genomics, Illumina synthetic long reads, and OpGen optical mapping. But the MinIon has more potential for diagnostic and screening use. Oxford nanopore platform cost and its operation are very low compared to other systems in use. This will push sequencing to spread from research laboratories to the clinics, where nanopore sequencing could become the tool for the low-level detection of cancer-associated SVs and their early detection. Clinical applications will require improved throughput (compared to amplicon sequencing), needed both for genome-wide detection of SVs and for targeted sequencing of regions likely to contain an SV. Increase in throughput will be accomplished with advances in the current MinIon platform, and with the larger instruments that Oxford Nanopore is developing.

As a response to the alleged infringement of patents (by Illumina), relating to the use of Mycobacterium smegmatis porin A (MspA), the company plans in April 2016 to release for the MinIon (and the upcoming Large throughput system PromethIon) the new R9 pore, which is based on the CsgG pore from *E. coli*.

There are now an increasing number of publications using the technique, with applications in analysis of clinical samples and small genomes. Schema of the principle of operation, relevant publications, collaborations and operation protocols involving the Oxford Nanopore system, are listed at the web site of the company (www.oxfordnanopore.com). Opinion of users and latest news can be found on the Genome Web site (www.genomeweb.com).

Other projects with protein nanopores, solid state and graphene nanopores

The nanopore systems in operation (Oxford Nanopore, Genia), use the protein nanopore channels in a membrane. It has been demonstrated that using special protein channels can give resolution of single DNA bases [5-7]. There are efforts underway to develop the nanopores with solid state techniques used in semiconductor industry, which could have some advantages in the mass manufacturing process of chips with large number of nanopores, with the ambition to increase the reproducibility of the nanopore performance. At present the thickness of the solid state membranes is too large, 10-20 nm, to allow the resolution of a single DNA base, since with this thickness there may be 30-50 bases simultaneously inside the nanopore. Further difficulty in the present techniques is the large speed with which the DNA molecule passes through the solid nanopore under the influence of an applied electrical field, making it impossible for the electronic detection system to achieve the single base resolution.

To slow down and control the translocation speed of the single strand DNA through the nanopore, one of the possible way is the use of DNA polymerases, as proposed e.g. in 2010 at EPFL in Lausanne (Ansorge W., note about patent proposal for DNA polymerase as a motor for controlled transport of DNA strand through a nanopore, in an EPFL internal e-mail communication on 4. August 2010 to prof. L. Forro and prof. A. Radenovic, and mailed to prof. Hagan Bayley on 17. August 2010), or in a publication in 2012 [6].

In an effort to reduce the nanopore thickness, several projects were recently described using graphene material for the nanopore [8-11]. The thickness of the graphene nanopore is so small, that just about one third of a nucleotide is inside the pore during the translocation of the DNA molecule through the pore, presenting a theoretical possibility to scan the DNA molecule with a single base resolution. However, as reported in the reports on graphene pores above, there are so far unsolved obstacles. Among them is the large speed of DNA translocation, needing to design much faster electronics signal detection, or alternatively developing some techniques allowing to lower the speed, e.g. by some controlled interactions of the passing DNA strand and the walls of the pore. Reported were also wetting problems of the graphene pore, possibly caused by the hydrophobic surface, requiring perhaps some surface passivation process for the pore.

There were other thin materials tested for the nanopore membranes, e.g. hafnium oxide and boron nitride [11], or molybdenum disulfide [12,13].

Besides the measurement of the changes in the ionic current as the DNA passes through the pore, there are efforts to measure the electric signals perpendicular to the DNA molecule (called tunneling currents DNA sequencing) as it traverses the pore or channel. The expectations are that each base will show sufficiently different tunneling current, in dependence on its electronic structure, and the base sequence will be deduced from the differences in the tunneling currents.

Genia-Roche

As the various platforms strive toward the \$1,000 genome, Genia (www.genia.com) is looking beyond, to the \$100 genome. The technology has the potential to reduce the price of sequencing and increase speed, accuracy, and sensitivity by moving away from complex sample preparation and optical detection. The base of Genia's technology is the biological nanopore, a protein pore embedded in a lipid bilayer membrane. The planar electronic sensor technology enables highly efficient accuracy of current readings, which is a common limitation in nanopore sequencing efforts. Genia's NanoTag sequencing approach, developed in collaboration with Columbia and Harvard University, [14], uses a DNA replication enzyme to sequence a template strand with single base precision, as base-specific engineered tags cleaved by the enzyme are captured by the nanopore. As the cleaved tags travel through the pore, they attenuate the current flow across the membrane in a sequence-dependent manner (Figure 1). Thus it is the tag, not the nucleotide, passing through the pore.

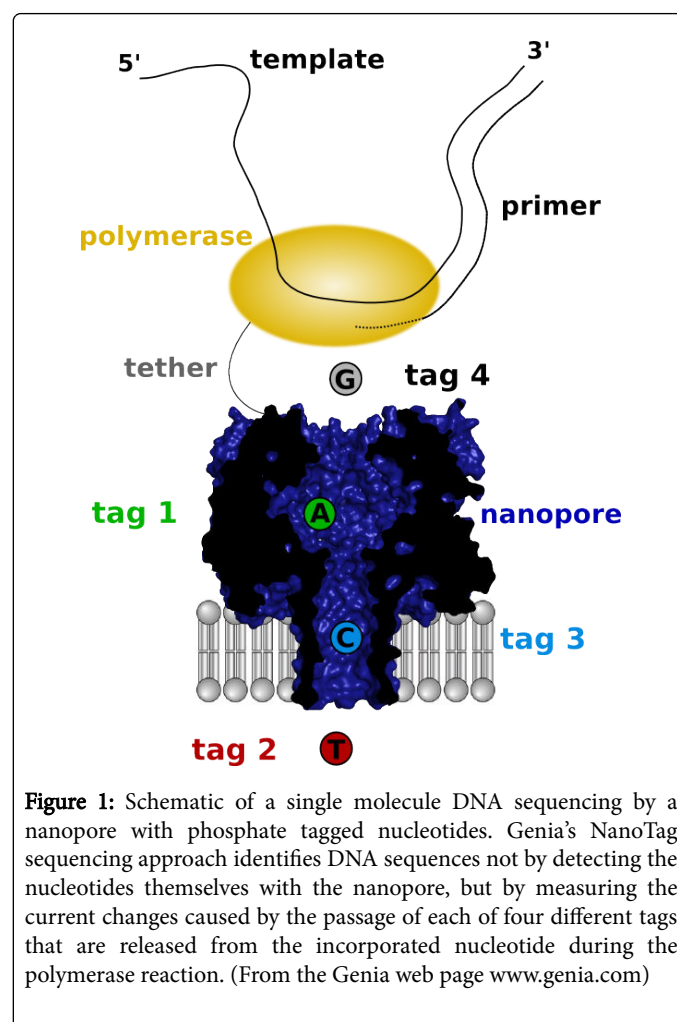


Figure 1: Schematic of a single molecule DNA sequencing by a nanopore with phosphate tagged nucleotides. Genia's NanoTag sequencing approach identifies DNA sequences not by detecting the nucleotides themselves with the nanopore, but by measuring the current changes caused by the passage of each of four different tags that are released from the incorporated nucleotide during the polymerase reaction. (From the Genia web page www.genia.com)

Electrochemically, each of the four tag-types interact with the nanopore recognition site differently, partially blocking the ion current by a characteristic amount which results in a tag-specific electronic signature (Figure 2). DNA sequences are computed from the residual currents measured on the nanopore/DNA complex during the passage of the cleaved tags through the pore. The high sensitivity of electronic detection circuitry underneath each sensor enables to obtain the

sequence of single DNA molecules with the Genia platform. The nanotags on the four nucleotides can be in the future optimized with the aim to achieve the highest resolution of the single bases [14].

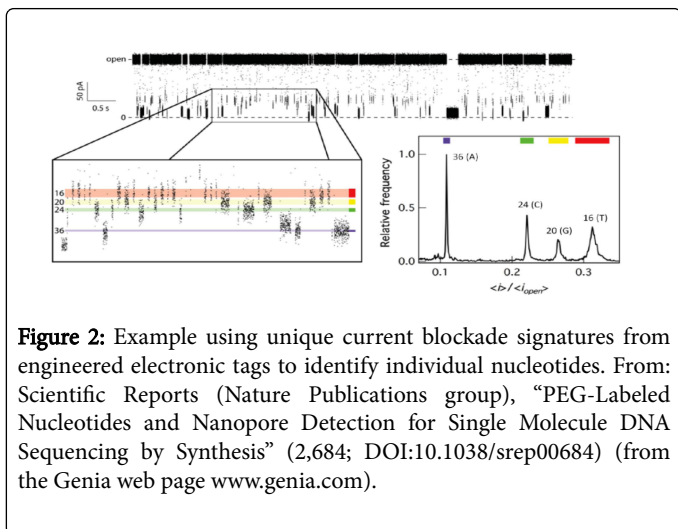


Figure 2: Example using unique current blockade signatures from engineered electronic tags to identify individual nucleotides. From: Scientific Reports (Nature Publications group), “PEG-Labeled Nucleotides and Nanopore Detection for Single Molecule DNA Sequencing by Synthesis” (2,684; DOI:10.1038/srep00684) (from the Genia web page www.genia.com).

Many of today’s sequencing platforms rely on specialized, expensive optical sensors. The Genia technology senses changes in electrical currents, detected by standard semiconductor devices, without optical systems. Involving semiconductor technology with the DNA sequencing will reduce the cost of Genia’s sequencer operation, making the \$100 genome a possibility. The company was acquired recently by Roche.

BioNano genomics

The BioNano Genomics company is not producing a classical DNA sequencing devices, although it uses rather similar nanochannel technology and fluorescence detection. But by providing a detailed physical Genome Map, it helps to finish sequencing, and to remove sequencing errors caused by repetitive regions. Genome Maps are a complement to sequencing and increase the quality of the results. NGS technologies are essential for nucleotide-level information, but some are limited in long range resolution. The reason is that the fragmented reads are often too short compared to repetitive regions, and longer base reading length would be needed to make a complete map and overcome repetitive regions within and between genes. To detect variation and architecture on a larger scale, scientists need a technology to directly map the genome with a resolution matching the variation in question, e.g. kilobases, not base pairs. Adding a genome map to sequencing data enables to obtain a better view of the whole genome, showing its features in context and functional relationships, across kilobases to megabases. Sequencing can be completed to a higher standard, in less time, with the help of a physical map to facilitate de novo sequence assembly and scaffolding.

The high-resolution Irys System from BioNano Genomics produces a detailed whole-genome map with highly precise anchor-points against which to align data from NGS. Genome maps orient contigs and size gaps by bridging across repeats and other complex elements that break NGS assemblies. The system improves the long-range contiguity across the genome, and helped in many cases to find errors in sequence assembly. Genome Maps represent a very useful validation method for genome contigs and scaffolds in regions assembled from fragment reads.

Structural variations in chromosomal DNA account for a large amount of the variability between human genomes, influencing also phenotypic variations. In the clinical research, structural variants are increasingly associated with several diseases, such as Crohn’s disease, autism, schizophrenia, morbid obesity and cancer. To study structural variation, researchers have traditionally been limited to techniques where the genomic DNA is first sheared into very small fragments, which limits the correct assembly of complex genomes. During assembly, repeats and structural relationships between regions are lost by the fragmentation. In the Irys System is the excessive fragmentation avoided, by relaxing and straightening large DNA molecules in nanochannels, so they can be imaged intact. Analysis of large whole molecules and generation of de novo genome maps allows viewing directly the elements in the genome, and to analyse structural variations. Other technologies that offer long-range genome information are Pacific Biosciences, Illumina synthetic long reads, OpGen optical mapping, and recently Oxford nanopore.

In the Irys system, the DNA to be analyzed is labeled at specific sequence motifs for imaging and identification in IrysChips. The labeling method in the system uses a nicking endonuclease to create a single-strand cut in the long DNA molecules, at a specific recognition site, wherever it occurs in the genome. Fluorescently labeled nucleotides are then incorporated by repairing the single-strand nicks with a polymerase. These labeling steps result in a uniquely identifiable sequence-specific pattern of labels to be used for de novo map assembly or for anchoring sequencing contigs. Other labeling techniques may be used in the system. The labeled DNA sample is pipetted onto the IrysChip in the flow cells, movement of DNA in the flowcell is controlled electrophoretically. The DNA stretches in solution, to confine chromosomal-length DNA inside the Nanochannels. The current is transiently turned off and the molecules in solution are stationary and stretched uniformly. At this time imaging is performed, capturing high-resolution, single-molecule images of the labeled DNA, that contains sequence motifs along hundreds of kilobases, (or even over a mega base), in a single contiguous molecule. Once imaged, the molecules are flushed and the process is repeated, allowing imaging of several gigabases of DNA per hour.

This technique reveals meaningful biological information in spite of the presence of repetitive elements, which cannot be resolved by some of the other technologies. It is possible to detect also epigenetic modifications across the genome. In contrast, in techniques where the DNA molecules are sheared to shorter fragments, the biological information may be disrupted or completely lost, e.g. the order and arrangement of functional regions.

Raw image data of labeled long DNA molecules are converted to digital representations of the motif-specific label pattern. These data are then assembled de novo using data analysis software to recreate a whole genome consensus map of the original genome. This genome map enables a variety of analyses, such as sequence finishing and structural variation detection.

Schema of principle of operation, relevant publications, collaborations and operation protocols involving the BioNano Genomics system, are listed at the web site of the company (www.bionanogenomics.com). Opinion of users and latest news can be found on the Genome Web site (www.genomeweb.com).

Techniques in Development

GnuBIO, BioRad

The GnuBIO platform is a new desktop sequencer which incorporates all steps for DNA sequencing into a single system. Utilizing emulsion microfluidics, GnuBIO has developed a scalable DNA sequencing reaction, which encompasses all of the steps necessary for DNA sequencing inside of pico liter-sized aqueous drops. Most other sequencing technologies require separate workflows for target selection, DNA amplification, DNA sequencing and analysis; the GnuBIO technology integrates the complete workflow into a single instrument. The platform provides a single step process which produces genomic result within hours. It is designed to meet the workflow of both research and clinical laboratories, in applications from single genes to whole genomes. GnuBio and its DNA sequencing platform has been acquired by BioRad Company in 2014, (schema of principle operation, and more information see www.gnubio.com).

Qiagen platform

Qiagen announced several years ago a plan to develop and launch a DNA sequencing platform mainly for clinical applications, to add to its wide range of products for preparation and purification of nucleic acids. For this purpose they bought the DNA sequencing technology from the SBH Company (www.qiagen.com). The complete chain for DNA analysis, from consumables for DNA preparation, over sequencing to data and functional analysis, would present a powerful combination. Although some initial test results were shown, the company has not yet launched the DNA sequencing platform, although all products for the other steps in the analysis chain, including the software for data and function analysis, are already distributed by Qiagen, including its expertise in single cell analysis. Progress, applications and projects can be followed on the web site www.qiagen.com.

Lasergen and Agilent

Lasergen company (www.laserger.com) has developed several years ago a novel "Lightning Reversible Terminators sequencing chemistry", which they expect to have the potential to be faster, more accurate, and less expensive than existing sequencing technologies. The Agilent company (www.agilent.com) has announced in March 2016, that it will collaborate with Lasergen, and will use its instrument engineering expertise to help Lasergen build a sequencing instrument; Agilent will also build several associated techniques for completing the workflow, including liquid handling, target enrichment, quality control, and informatics analysis.

Direct genomics single-molecule sequencing

Direct Genomics, a spin-off of the BGI institute in China, licensed its Single-Molecule Sequencing technique from Caltech, using a sequencing method first developed by Helicos Biosciences, Cambridge, Massachusetts. The GenoCare system will operate with chemistry and reagents from Helicos, but the technique and device were optimized and redesigned. The GenoCare is expected to have a throughput around 10 gigabases per run, run with very high consensus accuracy (approaching 100% by 5x coverage), read length 35 bases initially, and 140 bases with four color chemistry and improvements in hardware. The key improvement compared with the original Helicos technology is the use of total internal reflection fluorescence microscopy (TIRF).

The laser is reflected internally and illuminates only very thin region at the flow cell surface with captured DNA, resulting in a significantly improved signal-to-noise ratio. Other improvement is the use of latest generation of CCD camera, with higher sensitivity. The single molecule technique does not require sample preparation module, the DNA capture takes place directly on the flow cell surface. The GenoCare has been designed for clinical applications in China, concentrating on sequencing of gene cancer panels and noninvasive prenatal screening NIPT.

More information on the progress will be available at BGI and Direct Genomics web pages.

Potential Future Techniques

DNA sequencing by tunneling currents

Besides the measurement of the changes in the ionic current, as the DNA passes through the pore, there are efforts to measure electric signals perpendicular to the DNA molecule (in a technique called Tunneling Currents DNA Sequencing) as it traverses the pore or channel. The expectations are that each base will show sufficiently different tunneling current and the base sequence will be deduced from the differences in the tunneling currents.

Depending on its electronic structure each base may affect the tunneling current differently, allowing differentiation between the bases during the translocation through the pore, or channel. The technique could have the potential to sequence orders of magnitude faster than the ionic current methods. Sequencing of several DNA oligomers and micro-RNA by the tunneling currents technique has been reported [15,16].

Mechanical identification and sequencing of single DNA molecule

Recently was presented the proof of concept for a single-molecule platform that allows DNA identification and sequencing [17], based on the mechanical properties of DNA, as explained in the publication. As an entirely novel interesting approach, the technique is not based on the detection of the fluorescent nucleotides, but on DNA hairpin length. By mechanically pulling (in a magnetic tweezers system) on magnetic beads tethered by a DNA hairpin to the surface, the molecule can be un-zipped, and the two DNA strands mechanically separated. In this open state it can hybridize with complementary oligonucleotides, which transiently block the hairpin re-zipping when the pulling force is reduced. By measuring from the surface to the bead of a blocked hairpin, one can determine the position of the hybrid along the molecule, with nearly single-base precision. The approach can be used to identify a DNA fragment of known sequence in a mix of various fragments, and to sequence an unknown DNA fragment by hybridization or ligation. The proof of concept of the technique has been demonstrated, but it may not be easy to develop it into a simple system for routine DNA sequencing applications, competing with other techniques.

Sequencing with mass spectrometry

Matrix-assisted laser desorption ionization time-of-flight mass spectrometry, or abbreviated (MALDI-TOF MS), has been investigated specifically as an alternative method to gel electrophoresis for visualizing DNA fragments [18,19]. With this method, DNA fragments

generated by chain-termination sequencing reactions are compared by mass rather than by size, and the DNA sequence is deduced. The mass of each nucleotide is different from the others and this difference is detectable by mass spectrometry. Single-nucleotide mutations in a fragment can be more easily detected with MS than by gel electrophoresis alone [18,20].

Early chain-termination and TOF MS methods demonstrated read lengths of up to 100 base pairs. This appeared so far as the limit for the measurable average read size. MS-based DNA sequencing may not be suitable for large de novo sequencing projects, just like chain-termination sequencing alone, but it has been used e.g. in bacterial and forensic studies. A recent study did use the short sequence reads and mass spectroscopy to compare single-nucleotide polymorphisms in pathogenic streptococcus strains [21].

DNA sequencing by electron microscopy

This approach directly visualizes the sequence of DNA molecules using electron microscopy. The detection is not straight forward, since the atoms composing the DNA generally are of low mass, thus are not easily visualized by electron microscopic techniques. To be clearly visualized under an electron microscope, the DNA must be labeled with heavy atoms. In addition, specialized imaging techniques and aberration corrected optics are beneficial for obtaining the resolution required to image the labeled DNA molecule. In theory, transmission electron microscopy DNA sequencing could provide extremely long read lengths, but the issue of electron beam damage to the DNA has not been entirely resolved, and the technology has not yet been commercially developed. But the first identification of DNA base pairs within intact DNA molecules has been shown, proving that modified bases containing atoms of increased atomic number were enzymatically incorporated. Direct visualization and identification of individually labeled bases within a synthetic 3,272 base-pair DNA molecule and a 7,249 base-pair viral genome has been demonstrated [22,23].

There are at present two companies working on the DNA sequencing technique by Electron Microscopy, with different approaches. One company, ZS Genetics, published an initial proof-of-concept study in 2012 [22], demonstrating that it was possible to label (with a single mercury atom) and identify one of four bases of DNA with an electron microscope. The company claims, it is now able to uniquely label and distinguish between all four bases. The goal for DNA analysis with an EM is that it will enable the long reads and high accuracy. Without the need for fragmentation, the DNA is first labeled, then the EM takes an image of it, while analyzing and distinguishing between the different labels to read out the sequence. If confirmed in the future constructed device, such a strategy could enable reads of around 50,000 base pairs.

ZS genetics uses the transmission electron microscopy (TEM), where the electrons have about a thousand times more energy. In TEM, the high-energy electrons are not scattered enough by the light elements that make up DNA, so the molecule cannot be visualized directly. The heavy metal labeling approach has several problems. One is that the labeling reaction might not go to completion, thus some DNA bases may be missed and not detected. Further problem is that at present it is possible to label only one type of base at a time, because the labels are difficult to distinguish and interfere with each other when they get too close. Thus the sequence is reconstructed from four separate images. In addition, high-energy electrons often damage the

DNA, causing the labels to move around, so their position is no longer precisely located on the base.

Another start-up company Electron Optica and its collaborators have experimented and demonstrated in principle, that sequencing of non-labelled DNA by electron microscopy should be feasible. Technical approaches of the two companies differ in several aspects. Electron Optica will use LEEM, where the electrons have low energies up to several hundred electron volts. Because of the lower beam energy in the LEEM device, damage to the DNA molecule will be smaller, and it could reduce sequencing errors. Also, the low-energy electrons are slower, and get scattered by the light elements of DNA (mostly carbon, oxygen, and nitrogen) enabling to obtain sufficient contrast to visualize DNA directly, without the need to label the DNA with heavy metals, as was demonstrated in [23].

They show that the four base types provide sufficient contrast in low-energy electron microscopy (LEEM) for DNA sequencing. Electron Optica uses monochromatic aberration-corrected dual-beam low energy electron microscopy, where two beams illuminate the sample with low-energy electrons, and the reflected electrons compose an image. The LEEM technique has not been used widely in life sciences, because of the somewhat lower resolution, as compared with the other types of electron microscopes. Another reason for the less frequent use of the technique is that most bio-samples are electrical insulators. As such they charge up under the electron beam in the microscope, the charge acquired by the sample affects the trajectory and lowers the resolution of low-energy electrons. To get around this problem, the company had to develop a charge-control technique.

The Electron Optica is designing a new aberration corrector that can bring the resolution to about half a nanometer, needed for a demonstration that all four bases can be distinguished.

On four different kinds of substrates, and involving the use of advanced spectroscopy techniques such as X-ray photo-electron spectroscopy (XPS), and Auger electron spectroscopy, the company achieved contrast needed to distinguish the individual bases. The difference in the contrast among the bases is caused mainly by the differences in the nitrogen content of the bases [23].

The price of a finished sequencing microscope is estimated to be \$0.5 million to \$1million. The platform would compete with other technologies on read length (the goal is to obtain base readings up to 50 megabases), as well as cost per base, accuracy (error rate of 10^{-6}), and throughput (one genome per day). Besides DNA sequencing, it should also be useful for RNA sequencing and detection of methylation on bases.

More details and progress of the techniques can be found on the web site of the companies, (www.zsgenetics.com, and www.electronoptica.com).

RNAP sequencing technique

This method [1,2,24,25] is based on use of RNA polymerase (RNAP), which is attached to a polystyrene bead. One end of the DNA to be sequenced is attached to another bead, and both beads are placed in laser optical traps. RNAP motion during transcription brings the beads closer together, and as their relative distance changes, the change is recorded at a single nucleotide resolution. The sequence is deduced based on the four readouts with lowered concentrations of each of the four nucleotide types, similarly to the Sanger method. Comparison is made between regions and sequence information deduced by

comparing the known sequence regions to the unknown sequence regions. Due to the special equipment needed, this technique may be useful in special applications, and may not be easily developed into a wide technique competing with other technologies available.

FRET-based and Raman spectroscopy approach for DNA sequencing

A sequencing-by-synthesis approach that applies fluorescence resonance energy transfer (FRET) has been proposed by VisiGen Biotechnologies (www.visigenbio.com) already more than 10 years ago. In this technique a polymerase containing a donor fluorophore is used in combination with four different acceptor fluorophores for the respective bases. Whenever the polymerase incorporates a nucleotide, a FRET signal is generated through the proximity of the donor and acceptor fluorophores. Sequence information is obtained consequently based on the specific labels for each base. Upon release of the pyrophosphate, which contains the fluorophore, the FRET signal is quenched until the incorporation of the next base. The technique has been very promising, and taken over by Life Technologies in the past, without any news recently concerning its eventual further development.

The use of surface-enhanced raman scattering (SERS), and tip-enhanced raman scattering (TERS) techniques for DNA Sequencing are described in a publication [26], as well as in a technical review [27]. A test on a single RNA strand is reported as a promising step towards the development of a novel label-free single-molecule sequencing technique.

Perspectives for Future Applications and Diagnostics Techniques

As the techniques and technologies are developing at a fast pace, the DNA sequencing applications are spreading to all areas which are even remotely in connection with DNA analysis. As examples may serve the control of vaccines and medicament purity in pharma industry, where detected are not only active viral components, but also not active dormant viruses. Control of claimed meat quality in food processing industry, forensic analysis, fetal analysis from its DNA circulating in mother's blood, evolution biology of organisms, are just few several further fields where the technique found application. Here will be mentioned two emerging fields for which the latest DNA sequencing technology is of fundamental importance.

Precision medicine personal genomics

In recent years, precision medicine has gained increasing public attention and support [28]. As a result of technological advances that followed the Human Genome Project, emerged the expectation that the medical field can now include the study of DNA, detecting genetically conditioned illnesses and predicting personalized responses to potential therapies. This new form of medical treatment would allow concentrating on patient's specific needs in prevention, diagnosis, and treatment. Precision medicine has already revolutionized diagnosis and therapy in a number of cancers. Recognized as a 2013 breakthrough of the year by science magazine, a novel type of cancer immunotherapy has been developed. Genetically modified T-cells taken from blood and re-engineered to detect a protein on cancer cells, are used to fight the specific cancer cells. Compared with chemotherapy, the molecular profile of each individual patient will provide physicians with sufficiently specific details about the genetic

condition to enable personalized treatment, minimizing adverse side effects and unnecessary exposure to less efficient treatments. The precision medicine will be also of help in devising preventive approaches.

At present a person's genome can be sequenced at a cost of about \$1, and some companies already see the price being lowered to around \$100. This development indicates that DNA techniques providing individual molecular signatures may replace some traditional tools for diagnosis.

Single cell DNA sequencing, chips for diagnostics

The isolation and analysis of single cells is a field of increasing interest. The goal of obtaining the sequence of a single DNA molecule from a single cell, ideally the entire length of a chromosome, without amplification and without destruction of the original DNA strand, as discussed in [1,2], has moved closer to reality. The rapid progress in the single cell techniques was described in detail recently in the reviews [29,30]. The technique enables the study of rare cells, and will find applications in microbiology, neurobiology, development, immunology, cancer research, as well as in clinical environment.

Commercial devices for the cell isolation are available (www.fluidigm.com), but most of the analysis still requires amplification steps. After the single cell isolation, recent advances in whole-genome and whole-transcriptome amplification have enabled the sequencing of the low amounts of DNA and RNA present there. The technique made possible studies of genomic and transcriptomic heterogeneity, as they develop in normal as well as disease cells. It will also help to understand the evolution of genomic, epigenomic, and transcriptomic variations as they occur during the life time.

Conclusions

The diagnostics field has already profited from the new generation of DNA sequencing technology, sample preparation, single-cell isolation techniques, DNA amplification schemes, and computer data analysis with improving software tools. The recent novel therapeutic approaches and public support for efforts in improvements in the healthcare, will no doubt give steady push to develop and innovate further the field. The aims for the technology will be to lower the cost of the equipment and biochemicals involved, increasing simultaneously the reproducibility, reliability and simplicity of the techniques and protocols in operation.

The goal to be achieved, as formulated in [1,2] is still valid, in spite of the great progress in DNA sequencing techniques since then. For genomic sequencing, and for analysis of the ever-more important structural genetic variations in genomes (e.g. copy-number variation CNVs, for chromosomal translocations, inversions, large deletions, insertions, and duplications) it would be a great advantage if sequence read length on the original single DNA molecule could be increased to tens of thousands of bases and more, in a relatively short time. Ideally, the goal would be the sequence determination of a whole chromosome from a single original DNA molecule, from a single cell.

Since the previous reviews, a progress in this field has been very encouraging, particularly in throughput and cost per base on Illumina systems, single-molecule real time approaches on Pacific Bioscience and Oxford Nanopore. Many new developments and improvements are still needed. The promise for the benefits in health care and public interest in the Precision Medicine project, to understand novel

therapies, will be the stimuli for continuous efforts in developments of technology, diagnostics and therapies.

References

1. Ansorge WJ (2009) Next Generation DNA Sequencing Techniques and Applications. Review in New Journal of Biotechnology N Biotechnol. 25: 195-203.
2. Ansorge WJ (2010) Novel Next-Generation DNA Sequencing Techniques for Ultra High Throughput Applications in Bio-Medicine, Molecular Diagnostics (2nd edn), ScienceDirect.
3. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, et al. (2010) "Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays". *Science* 327: 78-81.
4. Brock AP, Bahram GK, Andrew BS, Oleg A, Peter H, et al. (2012) Accurate whole-genome sequencing and haplotyping from 10 to 20 human cells. *Nature* 487: 190-195.
5. Manrao EA, Derrington IM, Laszlo AH, Langford KW, Hopper MK, et al. (2012) Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nature Biotechnol* 30: 349-353.
6. Cherf GM, Lieberman KR, Rashid H, Lam CE, Karplus K, et al. (2012) Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precision. *Nature Biotechnol* 30: 344-348.
7. Laszlo AH, Derrington IM, Ross BC, Brinkerhoff H, Adey A, et al. (2014) Decoding long nanopore sequencing reads of natural DNA. *Nature Biotechnol* 32: 829-833.
8. Merchant CA, Healy K, Wanunu M, Ray V, Peterman N, et al. (2010) *Nano Lett* 10: 2915-2921.
9. Schneider GF (2010) *Nano Lett*. 10: 3163-3167.
10. Garaj S, Hubbard W, Reina A, Kong J, Branton D, et al. (2010). *Nature* 467: 190-193.
11. Drndic M (2014) Sequencing with Graphene pores. *Nature Nanotechnol* 9: 743.
12. Liu K, Feng J, Kis A, Radenovic A (2014) *ACS Nano* 8: 2504-2511.
13. Feng J, Liu K, Bulushev R, Khlybov S, Dumcenco D, et al. (2015) Identification of single nucleotides in MoS₂ nanopores. *Nature Nanotechnology* 10: 1070-1076.
14. Kumar S, Tao C, Chien M, Hellner B, Balijepalli A, et al. (2012) PEG-Labeled Nucleotides and Nanopore Detection for Single Molecule DNA Sequencing by Synthesis. *Scientific Reports* 2, Nature Publications group.
15. Ohshiro T, Matsubara K, Tsutsui M, Furuhashi M, Taniguchi M, et al. (2012) Single-molecule electrical random resequencing of DNA and RNA. *Sci Rep* 2: 501.
16. Ventra MD (2013) "Fast DNA sequencing by electrical means inches closer". *Nanotechnology* 24: 342501.
17. Ding F, Manosas M, Spiering MM, Benkovic SJ, Bensimon D, et al. (2012) Single DNA molecule mechanical identification and sequencing. *Nature Methods* 9: 367-372.
18. Monforte JA, Becker CH (1997) "High-throughput DNA analysis by time-of-flight mass spectrometry". *Nature Medicine* 3: 360-362.
19. Hall TA, Budowle B, Jiang Y, Blyn L, Eshoo M, et al. (2005) "Base composition analysis of human mitochondrial DNA using electrospray ionization mass spectrometry: A novel tool for the identification and differentiation of humans". *Analytical Biochemistry* 344: 53-69.
20. Edwards JR, Ruparel H, Ju J (2005) "Mass-spectrometry DNA sequencing". *Mutation Research* 573: 3-12.
21. Beres SB, Carroll RK, Shea PR, Sitkiewicz I, Martinez-Gutierrez JC, et al. (2010) "Molecular complexity of successive bacterial epidemics deconvoluted by comparative pathogenomics". *Proceedings of the National Academy of Sciences* 107: 4371-4376.
22. Bell DC, Thomas WK, Murtagh KM, Dionne CA, Graham AC, et al. (2012) "DNA Base Identification by Electron Microscopy". *Microscopy and microanalysis, Journal of Microscopy Society of America, Microbeam Analysis Society, Microscopical Society of Canada* 18: 1-5.
23. Mankos M, Shadman K, N'diaye AT, Schmid AK, Persson HH, et al. (2012) Progress toward an aberration-corrected low energy electron microscope for DNA sequencing and surface analysis. *J Vac Sci Technol B Nanotechnol Microelectron* 30.
24. Pareek CS, Smoczynski R, Tretyn A (2011) "Sequencing technologies and genome sequencing". *Journal of Applied Genetics* 52: 413-435.
25. Fujimori S, Hirai N, Ohashi H, Masuoka K, Nishikimi A, et al. (2012) Next-generation sequencing coupled with a cell-free display technology for high-throughput production of reliable interactome data. *Scientific reports* 2: 691.
26. Bailo E, Deckert V (2008) Tip-enhanced Raman spectroscopy of single RNA strands: towards a novel direct-sequencing method. *Angew Chem Int Ed* 47: 1658-1661.
27. Treffer R, Deckert V, (2010) Recent advances in single-molecule sequencing. *Current Opinion in Biotechnology* 21: 4-11.
28. Collins FS, Varmus H (2015) A New Initiative on Precision Medicine. *N Engl J Med* 372: 793-795.
29. Wang Y, Navin NE (2015) Advances and Applications of Single Cell Sequencing Technologies. *Molecular Cell* 58: 598-609.
30. Macaulay IC, Voet T (2014) Single Cell Genomics: Advances and Future Perspectives. *PLoS Genet* 10: e1004126.

This article was originally published in a special issue, entitled: "**Sequencing Technologies**", Edited by Jianping Wang