New Biotechnology • Volume 30, Number 4 • May 2013

ELSEVIER

DNA sequencing – spanning the generations

Steven McGinn[1] and Ivo Glynne Gut[2,]

[1] CEA, DSV, IG-Centre National de Genotypage, 2 rue Gaston Cremieux, 91057 Evry, France
[2] Centro Nacional de Analisis Genomico, PCB, C/Baldiri Reixac 4, 08028 Barcelona, Spain

Nucleic acid sequencing is the mainstay of biological research. There are several generations of DNA sequencing technologies that can be well characterized through their nature and the kind of output they provide. Dideoxy terminator sequencing developed by Sanger dominated for 30 years and was the workhorse used for the Human Genome Project. In 2005 the first 2nd generation sequencer was presented with an output orders of magnitude higher than Sanger sequencing and dramatically decreased cost. We are now at the dawn of 3rd generation with nanopore systems that are being developed for DNA sequencing. Meanwhile the field is also broadening into applications that complement 1st, 2nd and 3rd generation sequencing systems to get high resolution genetic information. The REvolutionary Approaches and Devices for Nucleic Acid analysis (READNA) consortium funded by the European Commission under FP7 has made great contributions to the development of new nucleic acid analysis methodology.

Introduction

A landmark paper published by Frederick Sanger in 1977 [1] presented a DNA sequencing technique that would become the gold standard for the next 30 years. The original dideoxy terminator DNA sequencing method was implemented with automated gel electrophoresis and fluorescent terminator chemistry, initially on slab gel and later capillary gel-based systems [2]. Most of the technological improvements of this method took place during the Human Genome Project (HGP) which was initiated in 1990 and over the next 13 years, in an international effort sequenced the first complete human genome at a cost approaching three Billion dollars [3]. Simultaneously a competing private venture also sequenced a human genome. In 2001 the initial results of this gargantuan effort were published [4], and in 2004 the finished sequence was published [5]. The Human Genome Project proved that whole genome sequencing (WGS) could be achieved, but at a cost prohibitive to be done on a routine basis. An effect of the HGP has been a technological revolution and the arrival of so called next generation sequencing (NGS). The aim of these technologies was to reduce the time, effort and cost of WGS to a level where the human genome could be sequenced on a routine basis for research and clinical applications. Today we can distinguish at least four generations with distinct features. The transition from first generation to second generation DNA sequencing was a disruptive one, with output increasing more than 5 orders of magnitude and cost dropping by more than 5 orders of magnitude. With current 2nd generation sequencing systems WGS of a human genome at 30× coverage can be achieved in 10 days at a cost of less than $10,000. Technological development continues to drive the cost of human WGS below $1000. The additional incentive from the X Prize foundation that has pledged to award $10 million to the first team that accurately sequences the whole genome of 100 subjects within 30 days for $1000 or less per genome has accentuated the intensity of technological developments within the field (http://genomics.xprize.org/). The central focus of 2nd generation sequencing is on developing systems that are very cheap and very fast with the main objective to generate large volumes of sequence. Because of the nature of the chemistry a genome sequence is in billions of short fragments between 70 and 400 bases long, which is a drawback and is dealt computationally. Data quality and

Corresponding author: Gut, I.G. (igut@pcb.ub.es)

1871-6784/$ - see front matter http://dx.doi.org/10.1016/j.nbt.2012.11.012

quantity are major issues as systems now produce 60 GBases of sequence per day on a single DNA sequencer. As a result computation for data analysis is what puts most strain on 2nd generation sequencing. Increases in data output will only accentuate this problem and new computational ways of dealing with this quantity of data are being developed. It can be anticipated that techniques that will be available in the future will provide sequence with markedly different characteristics than the current 2nd generation. It is also likely that additional, refined tools for certain sequencing applications will be required to complement the data that are produced using 2nd generation sequencers.

Even though 2nd generation sequencing is not an old technology, 3rd generation systems are being announced that hail again to be disruptive. The European Union realized that many issues in the DNA sequencing field need to be addressed and as a result a European project entitled READNA (REvolutionary Approaches and Devices for Nucleic Acid analysis) was funded with 12 million Euros to develop new nucleic acid analysis technologies. The review will present 3rd and 4th generation technologies that are being developed within the context of this innovative project.

Characteristics of the generations of DNA sequencing

There are at least 4 generations of methods for DNA sequencing that can be classified by key distinguishing features (Table 1). Other classifications are used, however, we focus on particular characteristics for this classification that make categorization clear.

1st generation sequencing

For 1st generation sequencing, fragment ladders of the sequence are generated either by enzymatically extending a primer hybridized to a pool of template molecules and introducing specific T, C, G, or A terminations along the template [1]. Alternatively, base-specific cleavages were introduced in a method devised by Maxam and Gilbert [6]. Fragment ladders are separated by gel-electrophoresis. Products are either revealed by radioactive labeling and exposure to a film or by the introduction of fluorescent dyes into the termination reaction and fluorescent imaging. This is now achieved during a sequencing run in real time.

2nd generation sequencing

A paradigm shift underlies 2nd generation sequencing [7]. Individual clonal DNA template are generated and all sequenced in parallel using microfluidics. Sequencing reactions are carried out with cycles of base additions and imaging. To determine the sequence consecutive images are overlaid and bases are called at the positions of all template molecules. Base additions can be done using DNA polymerization or ligation and the visualization is with chemiluminescence (Pyrosequencing chemistry used in the Roche FLX instrument), fluorescence (in the Illumina and LifeTechnologies 5500) or with pH changes (in the IonTorrent instrument). The key characteristics of 2nd generation sequencing are the use of many clonal templates in parallel and a sequence determination process using enzymatic replication.

2.5th generation sequencing

The SMRT sequencer from Pacific Biosciences is sometimes referred to as a 3rd generation sequencer, however, it satisfies the same characteristics as 2nd generation sequencing instruments in that individual clonal molecules are sequenced using an enzymatic template replication system [8]. However, it breaks with the random array concept of ''traditional'' 2nd generation systems. The polymerase enzymatic system is positioned at the bottom of the wells of zero mode wave guides [9]. Template DNA molecules are captured and base incorporation is monitored by cleavage of fluorescent dye-linked pyrophosphate in the volume limited observation window.

3rd generation sequencing

Third generation is defined by direct reading of individual molecules. No replication enzymatic system is used to identify the sequence. Nanopore sequencers (as outlined below) are examples of this.

4th generation sequencing

Fourth generation is a method that is currently still very experimental. It defines carrying out a sequencing experiment in context, such as the DNA of individual cells in a histological section, thus in their context. Applications of this method would be interrogation of defined DNA sequences such as ones that are likely to have undergone somatic mutation, rather than the generation of complete sequences from each cell.

Devices of 2nd and 2.5th generation DNA sequencing [10]

2nd generation DNA sequencing devices that did take off

The first instrument to reach the market was from 454 Corporation [11]. This system used emulsion PCR with oligonucleotides attached to beads for clonal amplification. The sequencing reaction is carried out using the Pyrosequencing chemistry originally developed by Ronaghi et al. [12–14]. For this method bases are added one at a time. The release of pyrophosphate is converted into luminescence. The amount of light emitted corresponds to the number of incorporated bases of that type (T, C, G, or A). 454 Corporation was later bought by Roche and the system evolved into the Roche GS FLX and a miniaturized version called the Roche GS FLX junior. The second system to reach the market was the 1G system from Solexa [15]. This system used bridge PCR for clonal amplification with surface grafted oligonucleotides and single base primer extension (also referred to as SBE) for sequencing. Solexa merged with Illumina and the system has evolved to the Illumina HiSeq2500, with which a human genome can be sequenced to 30× coverage in just over a day. Recently Illumina released the MiSeq benchtop instrument which offers shorter run times of lower volumes of data. The third system, the SOLiD was released by Applied Biosystems in 2007 [16]. In this system emulsion PCR is combined with oligonucleotide ligation for sequence determination. After a merger with Invitrogen the system developed into the Life Technologies 5500. Life Technologies recently acquired IonTorrent that developed a 2nd generation sequencing system with an innovative detection system (see below).

2nd generation DNA sequencing devices that never took off

There are several 2nd generation DNA sequencing systems that did not take off. Polony sequencing was implemented in an ''open source'' device called the Polonator [17]. The underlying concepts are similar to the SOLiD system oligonucleotide ligation used for the sequencing reaction [18]. The Heliscope from Helicos was a device that did not apply clonal amplification, but was only capturing the

TABLE 1

Overview of sequencing methods and characteristics

	Generation	Input	Template enrichment	Nature of sequencing template	Sequencing chemistry	Detection	Particularity	Reference
Sanger sequencing	1st	DNA	PCR	Pool	Primer extension, dideoxy termination	Radioactivity, fluorescence	Used for Human Genome Project	Sanger [1]
Maxam-Gilbert	1st	DNA	Sufficient template required	Pool	Chemical base-specific cleavage	Radioactivity		Maxam Gilbert [6]
454-Roche	2nd	DNA	Emulsion PCR	Amplified Clone	Primer extension one base at a time	Real-time chemiluminescence	Pyrosequencing	Margulies [11]
Polonator	2nd	DNA	Emulsion PCR	Amplified Clone	Oligonucleotide ligation	Fluorescence imaging		Shendure [17]
Solexa-Illumina	2nd	DNA	Bridge PCR	Amplified Clone	Single-base primer extension	Fluorescence imaging		Bentley [15]
Helicos	2nd	DNA	None	Individual DNA molecule	Single-base primer extension	Fluorescence imaging	Single molecule	Harris [19]
Applied Biotechnologies-Life Technologies	2nd	DNA	Emulsion PCR	Amplified Clone	Oligonucleotide ligation	Fluorescence imaging	Two-base encoding	McKernan [39]
Complete Genomics	2nd	DNA	Rolling circle replication	Amplified Clone	Oligonucleotide ligation	Fluorescence imaging	Proprietary system	Drmanac [20]
IonTorrent	2nd	DNA	Emulsion PCR	Amplified Clone	Primer extension one base at a time	pH measurement in semiconductor device	Non-optical	Rothberg [18]
Pacific Biosciences	2.5th	DNA	None	Individual DNA molecule	Primer extension	Real-time fluorescence imaging	Zero-mode wave guides	Eid [9]
Oxford Nanopore (Exonuclease)	3rd	DNA	None	Individual DNA molecule	Exonuclease digestion	Nanopore	Exonuclease nanopore	http://www.nanoporetech.com
Oxford Nanopore (Strand)	3rd	DNA	None	Individual DNA molecule	Direct measurement	Nanopore		http://www.nanoporetech.com
In situ	4th	Tissue section	*In situ* rolling circle	Rolling circle amplicon	Primer extension	Fluorescence imaging		Larsson [36]

individual template DNA molecules on the surface of a glass slide [19]. Sequencing was achieved by single base primer extension. To work around the erroneous base incorporations that result in sequencing errors if isolated molecules are analyzed, the entire sequencing cycle is repeated, thus generating two times the sequence for each template DNA molecule. Differences between the two runs are sequencing errors that are then not used for downstream processing.

All of these systems use optical detection of the sequencing process with high-resolution imaging system, followed by image processing to determine all the individual sequences of a run.

Innovative detection methods

Recently 2nd generation DNA sequencing systems have appeared that use alternative detection systems. IonTorrent uses enzymatic base incorporation one at a time similar to the Roche FLX system. Successful base incorporations are determined by pH changes due to the release of H+ during base incorporation. A microsystem with many pH sensors in parallel is used to detect the bases. Complete Genomics uses a random amplification scheme for sample preparation. Clonal templates are captured with a probe-anchor system and sequences are determined using oligonucleotide ligation [20]. Complete Genomics does not sell its systems but uses them exclusively to carry out service sequencing. Pacific Biosciences developed a technology that goes beyond 2nd generation sequencing and what we labeled as generation 2.5. In this technology a DNA polymerase is grafted to the bottom of a microwell. Detection is done from below taking advantage of a quantum physical principle. The diameter of the microwell does not allow light to penetrate deep into the well (zero-mode wave guides) and so restricts the irradiation volume [9]. The entry of the light into the well is deep enough to excite the sequencing system but not the other (fluorescent) components that are in solution. A template DNA molecule is captured by the DNA polymerase attached to the bottom of the well. Incorporation of nucleotides is monitored by following fluorescence that is linked to the pyrophosphate that is cleaved from the nucleotide in the polymerization reaction. Incorporation is followed in real time. This technology allows sequencing very long stretches of DNA. Read lengths of several kb have been reported. A few years ago Life Technologies announced that it was developing a similar technology except that it did not make use of the zero-mode wave guides, but rather uses nano-crystals and a FRET system. However, this very interesting system has not reached the market yet.

Data analysis

Data analysis challenges associated with 2nd generation sequencing

Data analysis of 2nd generation sequencing results has three major components: (1) Base Calling, (2) Alignment and (3) Variant Calling. For most systems the base calling is closely linked to the sequencing system and is done using software that is provided by the supplier of the sequencing device. Alignment is a strategy that is chosen when the sample that is under examination has a reference sequence. Sequences produced are compared against the reference sequence and the most likely position in the genome for a given read is determined. Sequences that uniquely map to the genome and have no mismatches relative to the reference are most easy to position. Evidently sequences that differ by an acceptable degree from the reference are of most interest as these are used to identify variability. There is a battery of different tools that have been developed to do alignment, to enter into details of the pro's and contra's would explode this review. What needs to be said is that there are four balancing criteria in an aligner, (a) is an aligner deterministic or not, (b) is it exhaustive, (c) what percentage of reads is it capable of mapping to a reference, and (d) what computational effort is required to do so. Variant calling is intimately linked to alignment. The more reliable the alignment is the easier the task of variant calling is. Again there are several variant callers available and detailing them would go too far. Most aligners nowadays do reasonably well at identifying single nucleotide variants and small indels. However, the identification of larger indels and structural variants is more demanding and software to do this are still subject to a lot of change. Crucial in terms of computation effort required is that the alignment step is the one that requires most compute resources.

Alignment strategies rely on the hypothesis that the position most similar to the reference is the correct place for a sequence read. In many instances this will be correct; however, there is residual probability that it could be false and that the correct position is in a different place in the genome where the match is more distant. Evidently this influences the results. An alternative strategy to alignment is *de novo* sequence assembly. Several software for *de novo* assembly from short reads from 2nd generation sequencers have been presented. However, these approaches require hardware architectures that are not commodity and computation time, sophistication required and efforts are substantial. A lot of development is going on in this domain. Strategies using a combination of alignment and *de novo* assembly could also be beneficial. In any case as 3rd generation will become available with different read characteristics, data analysis strategies will need to be adapted. They will not necessarily be more demanding than they are for 2nd generation sequencing.

3rd generation sequencing – nanopores

The only notable systems that satisfy the criteria we set out above for 3rd generation nucleic acid sequencing are nanopores. The methods shown to date directly measure the DNA molecule without a need to subject it to an enzymatic replication system.

Introduction to nanopore technologies

The general principle of nanopore based DNA sequencing is extremely simple and was first demonstrated as early as 1996 [21]. Essentially the system consists of a nanopore embedded in a membrane (biological or synthetic) immersed in salt solution. A current is applied to the system that drives ions through the nanopore. The arrival of a charged biomolecule such as a DNA base will create a resistance in the flow of ions that will lead to changes in the electrical current that can be measured. Molecules of different size block the nanopore by a different degree and for a different time [22]. If the four DNA bases (T, C, G, and A) can be distinguished as they pass through the nanopore this would allow to measure a DNA sequence. In essence the principle highlighted above points towards a method capable of long read DNA reads (>2 kb) [23]. The key to pricing of nanopore sequencing is the nanoscale size of these pores, the possibility of parallelization and low consumables cost [24].

Currently two main paths are being pursued for nanopore sequencing, the first is the use of biological nanopores and the second is the use of solid state nanopores [25]. Below the focus is on biological nanopores as companies taking advantage of biological nanopores plan to commercialise nanopore sequencing platforms in 2012 and 2013 (Oxford Nanopore Technologies (ONT) and Genia).

Currently biological nanopores have the upper hand as two papers have been published that prove that DNA sequencing is possible using either an α-hemolysin pore [26] or a MspA pore [24]. The advantage of biological nanopores is that they have a well defined size, can be modified very easily and produced in mass remaining homogenous in size and structure. In the α-hemolysin nanopore experiments, it should be noted that to achieve a fully functional system capable of discriminating each of the four DNA bases the nanopore had to be genetically modified and in addition an adaptor molecule important in slowing the translocation of DNA bases through the nanopore had to be non-covalently attached within the pore [27]. Essentially mutagenesis experiments concluded that a specific mutation within the beta barrel was necessary to achieve base discrimination. It describes that attachment of a cyclodextrin (CD) derivative was essential for continuous detection of bases flowing through the engineered nanopore [28], in particular optimal separation of bases was achieved by attaching the CD to the L135C cysteine residue within the beta barrel of the pore [29]. Conditions for the engineered hemolysin/CD construct were optimized to the conditions required by the exonuclease to be active while at the same time maintaining base discrimination capability. Ultimately a balance was sought between the voltage applied and the salt concentration used to ensure that the exonuclease performs correctly and the majority of DNA bases do indeed travel all the way through the pore (to avoid double reads of bases) at a speed that is not too fast to allow detection of individual molecules. The MspA paper uses the same general concepts but takes advantage of the smaller constriction site in the MspA pore (diameter 1.2 nm and length 0.5 nm) compared to α-haemeolysin (diameter 1.5 nm and length 5 nm) that shows a better potential to discriminate between bases. A drawback of the method is that double stranded DNA molecules are used to block translocation of DNA molecules (compared to using the CD molecule in the α-hemolysin work), therefore DNA must be converted so that double stranded sections separate the single stranded nucleotides that need to be measured, resulting in heavy pre-sequencing preparation of the DNA. However in a recent development the integration of bacteriophage DNA polymerase (DNAP) phi29 into the pore acts as a motor to pull the single-stranded template through the aperture one nucleotide at a time, allowing a distinct signal to be generated and each of the nucleotides can be measured [30]. A similar result was published using the phi29 with the α-hemolysin pore and this work has been licensed by ONT [31].

As for solid state nanopores there is a huge focus on graphene being the material that may be used in the future [32] although in a crucial piece by Hagan Bayley there is still some way to go before solid state nanopores and in particular graphene are at a stage where they can be used to routinely sequence DNA. It should be mentioned that solid state nanopore technology if mastered will produce pores that are more stable, the length and diameter can be controlled, have adjustable surface properties and have a greater potential for integration into devices and arrays than their biological counterparts [33].

On the eve of commercialization of two nanopore systems (Oxford Nanopore Technologies and Genia Technologies, Inc.) it appears that many of the technological challenges have been resolved, although the eagerly awaited DNA sequencing results are the only way to decide if the nanopore challenge has been cracked and it will take another few years to have a DNA sequencing system that delivers the kind of data quality required by researchers.

Key commercial players in nanopore sequencing

Before focusing on READNA and its partner ONT, below is a short compilation of companies developing nanopore sequencing platforms.

Genia Technologies Inc.[*] – Nanopore sequencing using biological nanopores and electronic control for the detection of tagged pyrophosphate released from a nucleotide triphosphate during replication with a polymerase attached to the entrance of the nanopore (release date 2013).

IBM/Roche[*] – Nanopore sequencing using solid state nanopores and electronic control moves DNA through nanopores embedded in a silicon chip and detection of electron tunneling current (no release date).

NABsys Inc.[*] – Nanopore sequencing using solid state nanopores and electronic control moves DNA with hybridized oligonucleotide probes through nanopores embedded in a silicon chip and detection of ionic current (no release date).

NobleGen Biosciences Inc.[*] – ''Optipore'' platform – Nanopore Sequencing using solid state nanopores and electronic control moves synthetic DNA template with hybridized oligonucleotide beacons through nanopores in a silicon chip, displacing the beacons and detection by fluorescence (2014).

Oxford Nanopore Technologies[*] – Nanopore sequencing using biological nanopores and a processive enzyme moves DNA through nanopores embedded in a lipid bilayer and detection of ionic current (2012).

READNA and DNA sequencing

Overview of READNA

The REvolutionary Approaches and Devices for Nucleic Acid analysis (READNA) project is an EU Framework 7 funded project that began on the 1st June 2008. The READNA consortium includes projects to accelerate new breakthrough DNA sequencing technologies and methods to enhance existing nucleic acid analysis methods. Information of all developments of READNA can be found at (http://www.cng.fr/READNA). The ultimate aim will be to progress technologies enabling sequencing of an entire human genome for 1000€ in less than one day. Here we will focus on a 3rd generation DNA sequencing technology that has been developed by Oxford Nanopore Technologies within READNA.

Oxford nanopore sequencing (nanopore exonuclease sequencing and nanopore strand sequencing) – 3rd generation sequencing

Nanopore technologies are being developed by several READNA partners. Here we will focus on the efforts of two READNA

[*] Data above from publication by Tracy Baas Keeping up with Nanopores [34].

partners, Oxford Nanopore Technologies (ONT) and the University of Oxford (Professor Hagan Bayley, also cofounder of ONT).

Oxford Nanopore Technologies (ONT) has been pursuing two different types of 3rd generation DNA technology based on the use of nanopores to sequence DNA, with the help of an exonuclease nanopore and direct strand sequencing. In February 2012 at the annual Advances in Genome Biology & Technology meeting in Marco Island, ONT announced their new DNA ''strand sequencing'' platform entitled ''GridION'' and at the same time they presented a disposable version of the same platform ''MinION''. The MinION platform has attracted a lot of interest as when commercialised it will become the first miniaturized, disposable sequencer that is as small as a USB stick and the cost will be around $900. ONT plans to commercialise the ''1st generation'' of their GridION and MinION to a select number of customers in the second half of 2012.

Nanopore exonuclease sequencing

In the first method a DNA molecule is captured by a processive exonuclease that is covalently attached at the entrance of the nanopore. The exonuclease cleaves DNA base by base and drops it into the nanopore. The bases are identified as they hit a confinement in an α-hemolysin nanopore at a speed that ensures that single base resolution is achieved. Translocation speed of DNA bases travelling through a nanopore of 5 nm has been a major issue in the past as single base resolution was not feasible. Essentially a nanopore that is 5 nm in length will have between 10–15 nucleotides passing through the nanopore at a given time [23]. Furthermore the translocation speed is in the range of 1 nucleotide per μs which is too fast to resolve an individual base. To achieve single base detection the rate must be in the order of 1 nucleotide/ms. One way to slow down the flow is to use non-covalent adaptors and in particular cyclodextrin (CD) has been the most efficient adaptor in combination with the engineered α-hemolysin protein pore designed and used by ONT, another is adapting the salt concentration [35]. The landmark finding published by Clarke et al. shows that CD was attached to specific cysteines within the nanopore to achieve base identification [26]. The bases are cleaved by the exonuclease and immediately come into contact with the CD adaptor, binding briefly before moving on and through the nanopore due to a difference in salt concentration. The temporary binding creates a disruption event that is read by an electronic chip associated with the nanopore allowing the order of bases to be determined due to a change in current. The differences in current are such that the system is capable of differentiating the four DNA bases A, G, C and T and in addition the system is able to distinguish methyl cytosine bases that are important in the analysis of DNA methylation [26]. Although this is a proof-of-principle that nanopores under the correct conditions (modifications to the pore, addition of an adaptor, modification of salt concentration and temperature control) can distinguish the four bases, further developments will be required to bring this methodology into the commercial arena. In addition throughput will require arrays of these nanopores to ensure that nanopore sequencing fulfills current expectations of a complete human genome sequence in a single day.

Nanopore strand sequencing

The second method being developed by ONT is nanopore strand sequencing. The major difference to nanopore exonuclease sequencing is that instead of individualizing the bases using an exonuclease enzyme for detection by the nanopore, the DNA molecule is rendered single-stranded and translocated through the nanopore. The issue that needs to be resolved for this approach to work for sequencing is the resolution of an individual base in the nanopore. The size of the nucleotide in comparison to the size of the nanopore results in multiple bases being within the constriction area of the nanopore and the current is blocked by a group of bases. As the DNA strand translocates through the nanopores it needs to be translated into sequence. ONT have publicly presented that they have managed to use this approach to sequence λ phage, by deconvoluting the changes in current across the nanopore using a Viterbi algorithm. The method has been privileged by ONT as it is potentially faster and more accurate due to each base being physically attached to the previous one allowing the order of the bases to be read correctly. The major limiting factor will become the processivity of the enzyme used, but if enough array slots are available, throughput will definitely surpass that of any other sequencing platform currently available at a cost that is hoped will be $1000 for a complete genome. ONT have still to publish DNA sequencing results but have confirmed the sequencing of a 5 kb viral genome that had a base calling error rate of 4%. The company aims to have a maximum error rate of 2% upon commercialization sometime in 2012. An impressive feature of this technology is that individual reads are on the order of 10s of Kbases. Data of this type revolutionize DNA sequencing.

In situ *sequencing (sequencing individual cells) – 4th generation sequencing*

1st generation sequencing prevailed for nearly 30 years, 2nd generation has been going on for the last seven years, 3rd generation is on the near horizon and already 4th generation is on the further horizon. Neighbouring cells in the context of tumor are not identical and it will become increasingly more important to be able to measure this. The group of Mats Nilsson has provided examples of transcript analysis of many different cells in parallel in their cellular context [36]. This is a proof of concept of 4th generation sequencing.

Manipulation of complete chromosomes, optical mapping techniques

Because of the sample preparation procedures that are used in 2nd generation sequencing the contextual information is lost. Haplotypic information is particularly valuable in the assembly of new genomes, but also is important for biological issues. Whether two mutant variants are on the same or opposite chromosomes is biologically relevant. Methods to predict haplotypes either use familial context or statistical inference. In random reorganized DNA molecules, such as they are encountered in many tumor cells, these strategies break down. However, molecular determination is possible using methods such as optical mapping [37]. Within READNA we have spent substantial effort on techniques to manipulate individual DNA molecules and have developed micro- and nano-fluidic systems to stretch out and make them available for further investigation [38].

Conclusion

DNA sequencing has reached a point where it can be applied holistically. Transitions between generations of DNA sequencing methods have been disruptive with orders of magnitude changes in output and complete changes in the nature of the type of sequence a method provides and dramatic decrease of cost. Whereas 1st generation DNA sequencing prevailed for 30 years, the last few years have seen a flurry of changes with 2nd generation and already the appearance of 3rd and 4th generation nucleic acid analysis methods. Every new generation has not displaced the previous one. There is a substantial overlap of different methods that help complete the toolset required for comprehensive genomic analysis. The research or clinical question at hand determines the optimal device or usage thereof. 2nd generation that is the current state of the art and most widely spread methodology, by the nature of it sequence (many short reads) puts substantial strain on computation and data analysis procedures. As we advance into 3rd generation sequencing, with longer reads, computational demands will quite probably decrease. However, because of increased overall output the downstream bioinformatic treatment of results will become even more important. 4th generation nucleic acid analysis methods will challenge image processing procedures and the resolution that these can achieve.

Acknowledgement

The review was written and funded through the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° [HEALTH-F4-2008-201418] entitled READNA.

References

[1] Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America 1977;74(12):5463–7.
[2] Kircher M, Kelso J. High-throughput DNA sequencing – concepts and limitations. Bioessays 2010;32(6):524–36.
[3] Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature 2001;409(6822): 860–921.
[4] Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science 2001;291(5507):1304–51.
[5] Finishing the euchromatic sequence of the human genome. Nature 2004; 431(7011):931–45.
[6] Maxam AM, Gilbert W. A new method for sequencing DNA. Proceedings of the National Academy of Sciences of the United States of America 1977;74(2):560–4.
[7] Metzker ML. Sequencing technologies – the next generation. Nature Reviews Genetics 2010;11(1):31–46.
[8] Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing. Human Molecular Genetics 2010;19(R2):R227–40.
[9] Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. Science 2009;323(5910):133–8.
[10] Fuller CW, Middendorf LR, Benner SA, Church GM, Harris T, Huang X, et al. The challenges of sequencing by synthesis. Nature Biotechnology 2009;27(11): 1013–23.
[11] Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature 2005;437(7057):376–80.
[12] Ronaghi M, Uhlen M, Nyren P. A sequencing method based on real-time pyrophosphate. Science 1998;281(5375):363–5.
[13] Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyrén P. Real-time DNA sequencing using detection of pyrophosphate release. Analytical Biochemistry 1996;242(1):84–9.
[14] Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, et al. The complete genome of an individual by massively parallel DNA sequencing. Nature 2008;452(7189):872–6.
[15] Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 2008;456(7218):53–9.
[16] Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H, et al. A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. Genome Research 2008;18(7):1051–63.
[17] Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, et al. Accurate multiplex polony sequencing of an evolved bacterial genome. Science 2005;309(5741):1728–32.
[18] Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature 2011;475(7356):348–52.
[19] Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, Braslavsky I, et al. Single-molecule DNA sequencing of a viral genome. Science 2008;320(5872):106–9.
[20] Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science 2010;327(5961):78–81.
[21] Kasianowicz JJ, Brandin E, Branton D, Deamer DW. Characterization of individual polynucleotide molecules using a membrane channel. Proceedings of the National Academy of Sciences of the United States of America 1996;93(24): 13770–73.
[22] Bayley H. Sequencing single molecules of DNA. Current Opinion in Chemical Biology 2006;10(6):628–37.
[23] Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, et al. The potential and challenges of nanopore sequencing. Nature Biotechnology 2008;26(10):1146–53.
[24] Derrington IM, Butler TZ, Collins MD, Manrao E, Pavlenok M, Niederweis M, et al. Nanopore DNA sequencing with MspA. Proceedings of the National Academy of Sciences of the United States of America 2010;107(37):16060–65.
[25] Schneider GF, Dekker C. DNA sequencing with nanopores. Nature Biotechnology 2012;30(4):326–8.
[26] Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H. Continuous base identification for single-molecule nanopore DNA sequencing. Nature Nanotechnology 2009;4(4):265–70.
[27] Wu HC, Astier Y, Maglia G, Mikhailova E, Bayley H. Protein nanopores with covalently attached molecular adapters. Journal of the American Chemical Society 2007;129(51):16142–48.
[28] Astier Y, Braha O, Bayley H. Toward single molecule DNA sequencing: direct identification of ribonucleoside and deoxyribonucleoside 5′-monophosphates by using an engineered protein nanopore equipped with a molecular adapter. Journal of the American Chemical Society 2006;128(5):1705–10.
[29] Stoddart D, Heron AJ, Klingelhoefer J, Mikhailova E, Maglia G, Bayley H. Nucleobase recognition in ssDNA at the central constriction of the alpha-hemolysin pore. Nano Letters 2010;10(9):3633–7.
[30] Manrao EA, Derrington IM, Laszlo AH, Langford KW, Hopper MK, Gillgren N, et al. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. Nature Biotechnology 2012;30(4):349–53.
[31] Cherf GM, Lieberman KR, Rashid H, Lam CE, Karplus K, Akeson M. Automated forward and reverse ratcheting of DNA in a nanopore at 5-A precision. Nature Biotechnology 2012;30(4):344–8.
[32] Garaj S, Hubbard W, Reina A, Kong J, Branton D, Golovchenko JA. Graphene as a subnanometre trans-electrode membrane. Nature 2010;467(7312):190–3.
[33] Dekker C. Solid-state nanopores. Nature Nanotechnology 2007;2(4):209–15.
[34] Baas T. Keeping up with nanopores. In: SciBx; 2012. pp. 1–3.
[35] Banerjee A, Mikhailova E, Cheley S, Gu LQ, Montoya M, Nagaoka Y, et al. Molecular bases of cyclodextrin adapter interactions with engineered protein nanopores. Proceedings of the National Academy of Sciences of the United States of America 2010;107(18):8165–70.
[36] Larsson C, Grundberg I, Söderberg O, Nilsson M. In situ detection and genotyping of individual mRNA molecules. Nature Methods 2010;7(5):395–7.
[37] Aston C, Mishra B, Schwartz DC. Optical mapping and its potential for large-scale sequencing projects. Trends in Biotechnology 1999;17(7):297–302.
[38] Westerlund F, et al. Fluorescence enhancement of single DNA molecules confined in Si/SiO_2 nanochannels. Lab Chip 2010;10(16):2049–51.
[39] McKernan KJ, Peckham HE, Costa GL, McLaughlin SF, Fu Y, Tsung EF, et al. Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. Genome Research 2009;19(9):1527–41.