The molecular principles of next-generation sequencing

Bachelor thesis Rijksuniversiteit Groningen

Author: G.R. Veurman (S2812290)
Supervisor: prof. dr. O.P. Kuipers
February 2015

Abstract

Nowadays next-generation sequencing is one of the most important sequencing technologies. The technique is used in many fields. It is used in hospitals, to identify bacterial and viral species, and also to sequence the whole genome of different organisms. Today a lot of companies have developed a next-generation sequence method. The companies use different methods. One is fast, the other more reliable, but what are the advantages and disadvantages of each method. The development goes quite fast of the next-generation sequencing technology because each company will get the best method on the market. Companies have to continue improve their own method. If a company doesn't invest they will lose their position on the market.

Table of contents

Introduction

DNA sequencing began in 1968. The first paper was published by Wu and Kaiser. (Wu 1968) The method was laborious and only applicable to Lambda, but it was an important step. The modern DNA sequencing began in 1977, with two different procedures. The chain termination method from Sanger and the chemical degradation method from Maxam and Gilbert. The DNA sequencing was interesting because you could something tell about the genetic code. They could see that two different protein products were produce by the same piece of the genome. (Maxam 1977) (Sanger 1977)

Furthermore, the human genome was interesting to sequence. The idea to sequence the human genome began in 1985 when Robert Sinsheimer organized a meeting on human genome sequencing at the University of California, Santa Cruz. A genome research program was published in 1988. (Hutchison 2007) In 1990, the Human Genome Project (HGP) started with the goal an accurate sequence of the vast majority of the human genome. (International Human Genome Sequencing Consortium 2004) The project would develop technology for analysing the human, fruit, flies and mice genomes. It cost $3 billion and completed in 2003. The HGP sequenced 99% of the human genome's gene-containing regions. (National Human Genome Research Institute n.d.) The HGP used the Sanger method, the gold standard for nucleic acid sequencing. (Grada 2013) The time is quite long to sequence the genome, with the Sanger method. Companies developed other techniques, namely the next-generation sequencing method.

28 years after the Sanger sequence method, the first next-generation DNA sequence method, the Roche 454, came on the market. Nowadays a lot of companies have developed a DNA and RNA sequence method. (Morozova 2008) Some of the companies are Illumina, Life Technologies, Pacific Bioscience and Oxford Technologies. Next-generation sequence had an enormous impact on genomic research. The sequencing goes faster and is cheaper than the Sanger method. Next-generation sequencing is used for standard sequencing applications and for applications undiscovered by Sanger sequencing. (Morozova 2008) It is based on single DNA molecules that are spatially separated in a flow cell or massively parallel sequencing of clonally amplified. (Voelkerding 2009) Next-generation sequencing consists of 4 phases: sample collection, template generation, sequencing reactions and detection, and data analysis. (Rizzo 2012)

The next-generation sequence methods can be used to sequence the human genome or the whole genome of other organisms. A lot of next-generation sequence methods are used in hospitals or in the research. The human genome could be sequenced for genetic research, so they could genetic diseases investigate in a family. Also mutation discovery in individual human genomes is examined. (Wheeler 2008) It is also possible to sequence the cancer genome. So they could get more information about cancer. (Samorodnitsky 2014) Another research is large-scale polymorphism discovery. (Van Tassel 2008) So they could tell something about the polymorphism in an organism. Also possible the sequencing of bacterial and viral species to facilitate the identification of novel virulence factors. (Grada 2013) It is also possible to sequence RNA. The method is called RNA seq. An example of RNA seq is the effect of microRNA developed for gene regulation. (Morin 2008) RNA seq is also used for studying complex transcriptomes. The RNA is a short piece, so it could give information about how two exons are connected. Advantages of RNA seq is a low background signal. A challenge is the larger RNA molecules cannot directly, after preparation, be sequenced. They have to make in smaller pieces. So the larger and small RNA molecules have to be separate. (Wang 2009)

Great advantages of next-generation sequencing are the low cost per base, less labour and the short time to sequence a genome. Whereby the Sanger sequence method sequence a genome in a couple of years, now they could sequence a genome in a couple days. The cost per genome was $100,000,000 in 2001. In 2014 it dropped to $70,000. The cost are based on labour, administration, management, reagents, consumables, and sequencing instruments. (Wetterstrand 2014) The instrument price is lower but the financial investment remains high. (Kircher 2010) The field was dominated by Sanger, but now a lot of companies have invested in a next-generation sequence method.

The developments are still engaged of the next-generation sequence methods. An example of developments of the next-generation sequencing methods is the Oxford nanopore. In 2012, Oxford Technologies came with a new next-generation sequence method on the market and they came with new developments in October 2014. The other companies use other methods. All companies have to constantly development otherwise another company will take your customers, if they could do it cheaper or faster.

So there are a lot of different companies on the market with their own next-generation sequence method, but what are the advantages and disadvantages of each method. This research will discuss the different methods. How do they work. What is the price, time and reliability of the different next-generation sequence methods. Which method is the future.

Next-Generation DNA and RNA Sequence methods

Different companies have developed a next-generation sequence method. The companies are Roche, Illumina, Applied Biosystems, Helicos, Life Technologies, Pacific Bioscience and Oxford Technologies. This thesis will discuss Roche, Illumina, Life Technologies, Pacific Bioscience and Oxford technologies. The Applied Biosystems is not discussed because it used the similar PCR method as the Roche. Helicos use expensive reagents and is not use a lot.

The Sanger method was the basic method for the next-generation sequence method. The Sanger method is based on the principle that a polyacrylamide gel electrophoresis can separate the DNA molecules difference in length by just one nucleotide. The sample consists of identical single-stranded DNA molecules. Short oligonucleotides anneal to the same position of each DNA molecule. The four deoxyribonucleotide triphosphates (dNTPs) and the dideoxynucleotide (ddATP) are added to the reaction. A dNTP or a ddATP is built in the growing strand. If a ddATP is built in the growing strand, it will block further elongation of the strand. The DNA position can directly be read after the electrophorese, from the gel. (Brown 2002)

All the next-generation sequence methods begin with library preparation. Hereby the DNA is broken in small pieces. Adapters are bind to the end of the pieces DNA. After this the companies have their own method.

In this review I will have a look at the next-generation DNA sequencing of the following companies Roche, Illumina, Life Technologies, Pacific Bioscience and Oxford Technologies. Next-generation sequencing methods can also read RNA. RNA must first be converted to cDNA by the companies Roche, Illumina, Life Technologies and Pacific Bioscience. Hereafter the cDNA can be sequenced in the same way of the DNA sequencing. (Roche Diagnostic Corporation 2014) (Illumina 2014) (Thermo Fisher Scientific Inc. [1] 2014) (Underwood 2012) The RNA strand can directly analysed by the method of Oxford Technologies in the same way of the DNA sequencing. (Oxford nanopore technologies [1] 2014) Table 1 gives a comparison of the different companies. Table 2 gives the advantages and disadvantages of the methods of the different companies.

Table 1 Comparison of the different companies. The data is based on the different machines of a company, because the companies developed different machines.

	Roche 454[1]	Illumina[2]	Life Technology[3]	Pacific Bioscience[4]	Oxford Technologies[5]
Run time	10 - 23 hours	26 hours - 10 days	2 hours	0.5 – 2 hours	36 hour
Millions of reads per run	0.10 - 1	3.4 - 3000	0.10, 1, 4 or 8	0.01	5 400 -10 000
Bases per run	400 - 1000	500 - 5 billion	100	860 – 1100	*
Yield Mb per run	35 - 700	1020 - 600 000	10, 100 or 1000	5 – 10	*
Reagent cost per run	$1100 or $6200	$70 - $23 470	$500 - $925	$110 - $900	*
Reagent cost per Mb	$22, $12.4 or $7	$0.04 - $0.75	$0.93 - $50	$11 - $180	*
Purchase cost (list price in thousands of US dollars)	$108, $500 or $29.5	$125 - $690	$49.5	$695	$1
Error rate	1%	0.1%	1%	16%	*
DNA input	500 ng	500 ng – 4 µg	50 – 100 ng	250 -5000 ng (dependent of the amount base pairs)	1 µg

[1] References Roche: (Glenn 2011) (Roche Applied Science 2010) (Liu 2012)

[2] References Illumina: (Glenn 2011)

[3] References Life Technology: (Glenn 2011) (Thermo Fisher Scientific Inc. [2] 2014)

[4] References Pacific Bioscience: (Glenn 2011) (Pacific Bioscience 2014)

[5] References Oxford Technologies: (Mikheyev 2014) (Hayden 2014)

* Oxford Technologies use a relative new method, all the data is not known.

Table 2 Advantages and disadvantages of the methods of the different companies.

	Advantages	Disadvantages
Roche[1]	○ Long read length ○ Low capital cost ○ High speed	○ High reagent cost per Mb
Illumina[2]	○ Low error rate	○ Long running time
Life Technologies[3]	○ Low cost instruments ○ Very simple machine ○ Run time is short	○ High cost per Mb
Pacific Bioscience[4]	○ Long read length ○ Low cost per sample ○ Run time is short ○ Sample preparation time is short	○ High error rate ○ High cost per Mb
Oxford Technologies[5]	○ Low cost ○ Portability because you use a chip ○ Ability to directly identify modified bases ○ Sequence readout that does not require nucleotides, polymerases or ligases	○ High error rate ○ Sequencer makes the same mistakes

[1] References Roche: (Glenn 2011) (Liu 2012)

[2] References Illumina: (Glenn 2011)

[3] References Life Technology: (Glenn 2011)

[4] References Pacific Bioscience: (Glenn 2011) (Liu 2012)

[5] References Oxford Technologies (Bayley 2014) (Branton 2008) (Krol 2014)

Roche

Roche has developed the 454 Roche. The 454 Roche is the first generation next-generation sequence method. It is still on the market but Roche will stop supporting the Roche 454 sequencing instrument in 2016. (Karow 2013) The 454 Roche used an emulsion PCR (Fig. 1). Emulsion PCR is used to prepare sequencing templates in a cell-free system. The reaction mixture consists of emulsion oil, a PCR mix, beads and library DNA. A library of fragments is created. The adapters containing universal priming sites are ligated to the fragments ends. After ligation, the fragments are denatured in the PCR. One fragment, the reverse strand, anneals to the adaptersite on the bead. The complementary strand is extended from the bead toward the primer site. The reverse strand releases the bead. The forward strand is connected to the bead by the sugarphosphate backbone of DNA. The reverse strand anneals on another adaptersite on the bead. A primer anneals to the forward strand and the complementary strand is extended. Also the complementary strand of the original reverse strand is extended. The annealing, extension and the release steps are repeat for 30-60 PCR cycles. So the beads have a lot of strands. The beads can deposit into individual PicoTiterPlate (PTP) wells. A spectrometer read the samples. (Dressman 2003) (Bhattacharya 2012) (Margulies 2005) (Vierstraete 2012)

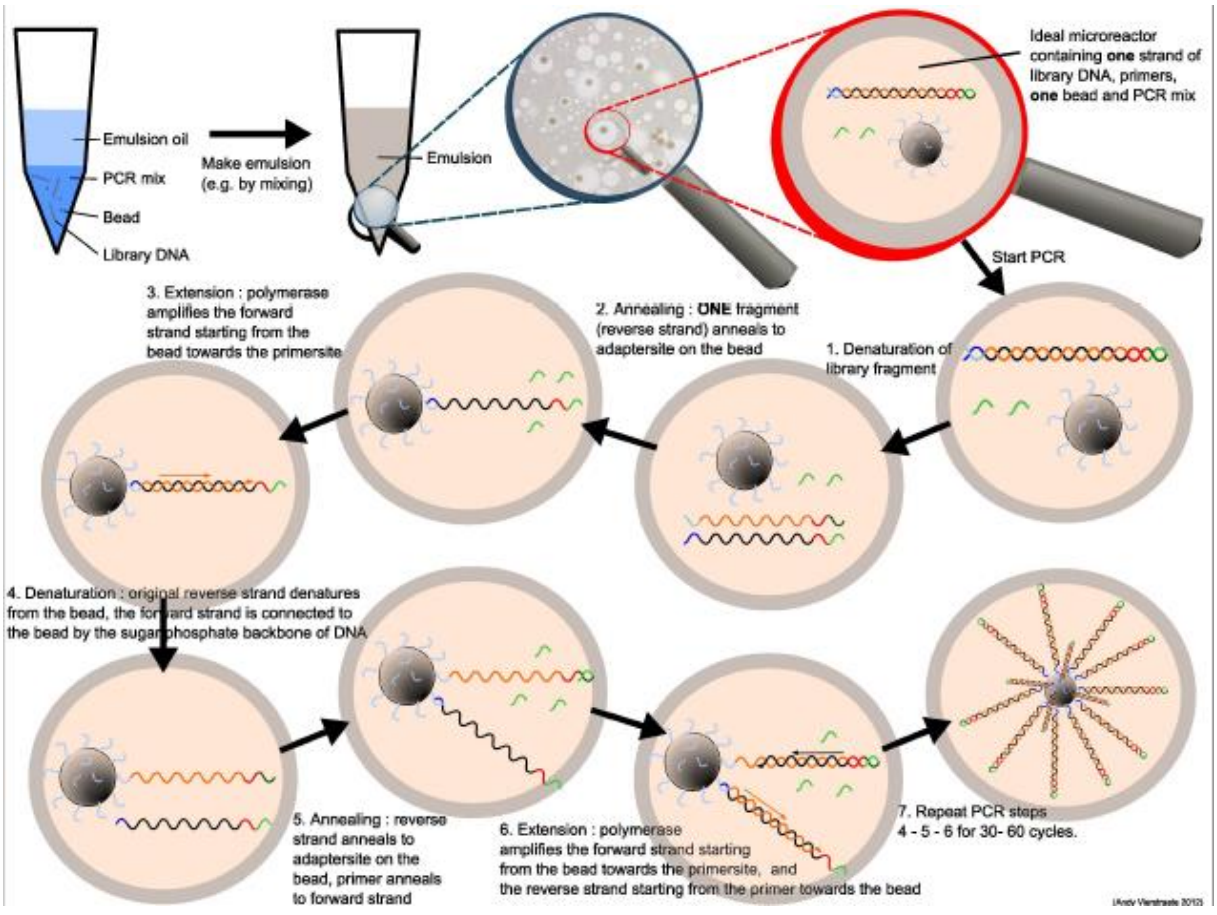

Fig. 1 Explanation of the emulsion PCR. First the DNA is denatured. One fragment (the reverse strand) anneals to the adaptersite on the bead in the second step. Hereafter the complementary strand is extended. Then the original reverse strand releases the bead. The forward strand is connected to the bead by the sugarphosphate backbone of DNA. The original reverse strand anneals to another adaptersite on the bead. Both complementary strands are extended. Step 4, 5 and 6 are repeated for 30 – 60 PCR cycles. (Vierstraete 2012)

Illumina

Illumina has lot next-generation sequence methods for different applications. The methods start in the follow way, DNA is broken into fragments. Adapters attach to both ends of the fragments (Fig. 2a). The fragments are bind to the adapters on the surface of the flow cell. The flow cell is a glass slide with lanes. Each lane is a channel coded with a lane composed of two types of oligos. The oligos are complementary to the adapter region on one of the fragment strands. The fragment strands are hybridized. After this the double stranded molecule is denatured and the originally template is washed away. The residual strand makes a bridge with another adapter on the surface. Unlabeled nucleotides and enzyme added to initiate solid-phase bridge amplification. The adapter region hybridized to the second type of oligo on the flow cell. The double bridge is denatured, what resulted in two single-stranded templates, a forward strand and a reverse strand. These strands make a new bridge, denatured, so we have now four strands. This process is repeated over and over again (Fig. 2b) 100-200 million separated template clusters are generated in each channel of the flow cell. The reverse strands are cleared off and washed away. (Illumina 2010) (Fedurco 2006) (Illumina inc 2013) (Bhattacharya 2012)

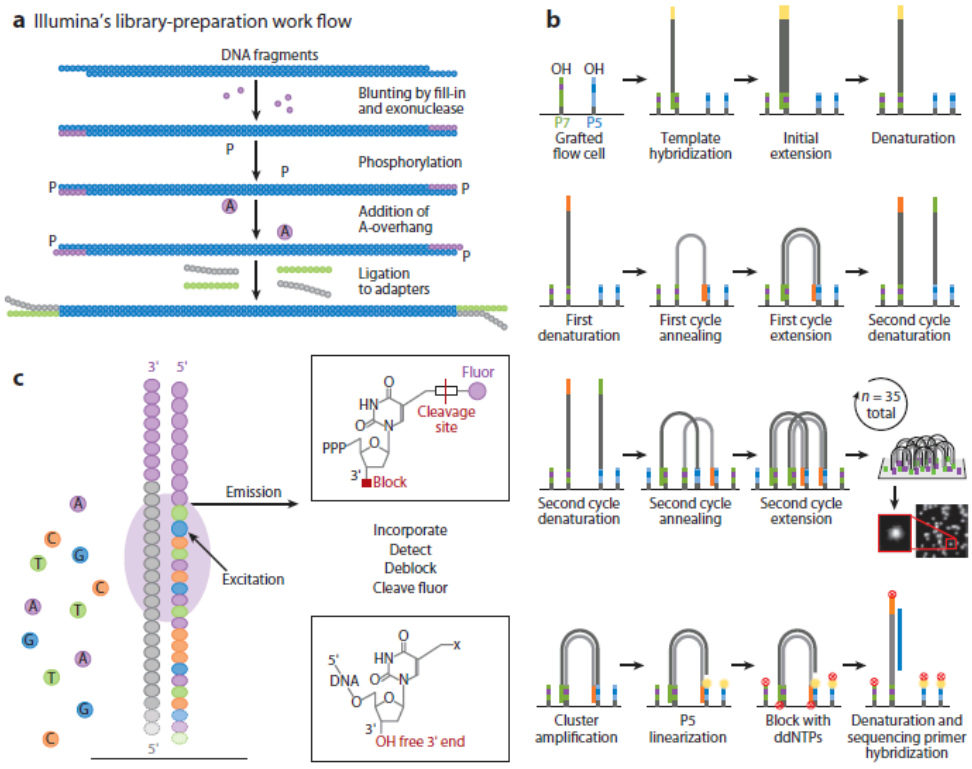

Fig. 2 Process of Illumina. (a) The library preparation. (b) The amplification on the flow cell. First the template is hybridized to the flow cell. The strand is extended and denatured. The original strand is washed away. The residual strand makes a bridge. This strand is extended. The strands go from bridge positon to upright position. We have two strands now. The steps are repeated and the strands make clusters. (c) The sequencing of the strand. The nucleotides have a unique fluorescent label, what gives a signal, so you can see where each nucleotide is in the DNA molecule. (Mardis 2013)

Sequencing begins by adding four labeled nucleotides, primers and DNA polymerase. The primer attached to the strand. A labeled nucleotide attach to the primer. Which nucleotide attach, is dependent on the template of the forward strand. You can identify the base with a laser. The labeled nucleotide sends a fluorescent signal, so you can identify the base (Fig. 2c). A camera records the

location of the clusters. The sequence becomes visible with the pictures, made by the camera. (Illumina 2010) (Illumina inc 2013) (Bhattacharya 2012)

Life Technologies

Life Technologies has developed the Ion Torrent. The Ion Torrent uses a chip for the sequencing. This chip is similar to your photo camera. It has millions of wells covering pixels. The Ion Torrent uses just, as the 454 Roche, an emulsion PCR. The sequencing started when DNA is cut in millions of fragments. Adapters are attached to the fragments. The sample is placed in the PCR. Each fragment is attached to his own bead and is copied around the bead. This process costs millions of beads with millions of different fragments. The beads with fragments flow to the chip, each bead comes in a well. The chip consists of several layers (Fig. 3). The upper surface consists of the wells with the beads. The well floods with one of the four nucleotides. If a nucleotide is in cooperate into a single stranded DNA, hydrogen is released. The hydrogen changed the pH in the well. The middle layer serves as a microfluidic conduit to provide the reactions needed for the sequencing reactions. The bottom layer detect the change in the pH in the well and coverage in voltage. The change in voltage goes directly to digital information. The process is repeated every 15 seconds with another nucleotide washed over the chip. If the nucleotide does not give a match, the voltage will not change and the base will not called. If two consecutive bases are the same, the voltage will double. (Life Technologies 2014) (Rothberg 2010) (Rothberg 2011)

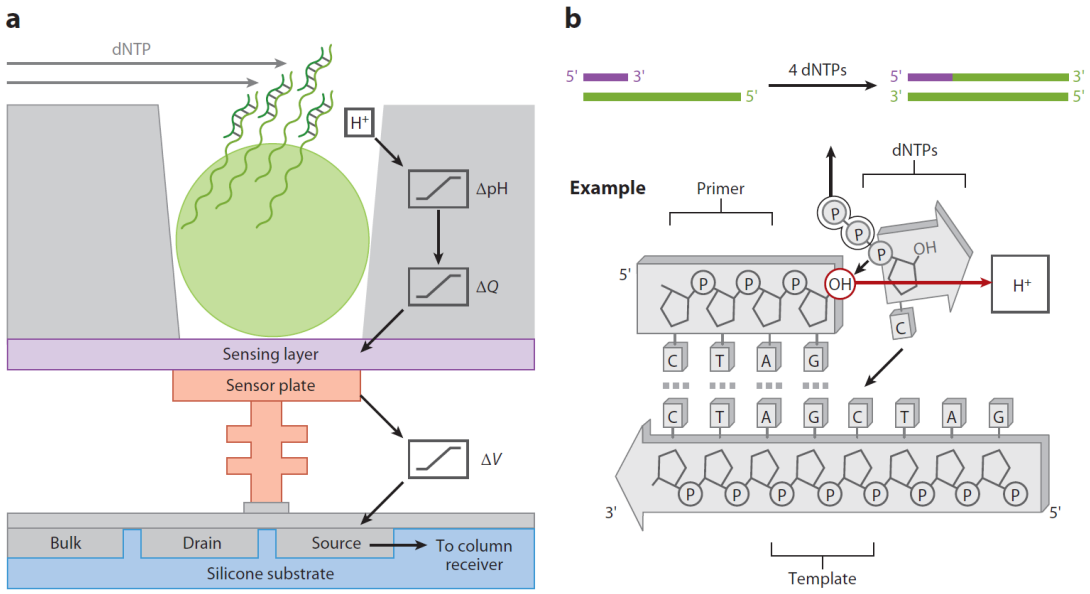

Fig. 3 (a) Structure of the ion chip. The upper layer is the well with the bead (green sphere). The purple layer servers as a microfluidic conduit to provide the reactions needed for the sequencing reactions. The orange layer detects the change in the pH. (b) The pH detection. If a dNTP is incorporated into the strand, hydrogen releases. The pH is changed and the ion detector detects the change in the well. (Mardis 2013)

Pacific Bioscience

Pacific Bioscience (PacBio) uses the single molecule real-time (SMRT) sequencing. The method is based on real-time imaging of fluorescently tagged nucleotides and do not use a PCR. The nucleotides are included into nascent DNA molecules from individual DNA templates. The four nitrogenous bases have a different fluorescence. The fluorescent label is attached to the phospholinked dNTPs in the active site of the polymerase. Single DNA polymerase molecules are attached to the bottom of the zero-mode waveguide (ZMW) (Fig.4) to which a primed template molecule is bound. The ZMW are essentially small pores that are surrounded by metal film and silicon dioxide. This enables the detection of single molecules while DNA polymerase replicates the chain inside the well. A fluorescence pulse is produced by the polymerase. The fluorescence pulse corresponded to a dNTP. Two lasers are used for the detection of the fluorescence. (Eid 2009) (Levene 2003) (Roberts 2013) (Bhattacharya 2012) PacBio can use a real-time sequencing because they did not use base-linked fluorescent nucleotides. Base-linked florescent nucleotides cannot be used in real-time sequencing because they are poorly absorbed in consecutive positions by DNA polymerase. (Eid 2009)

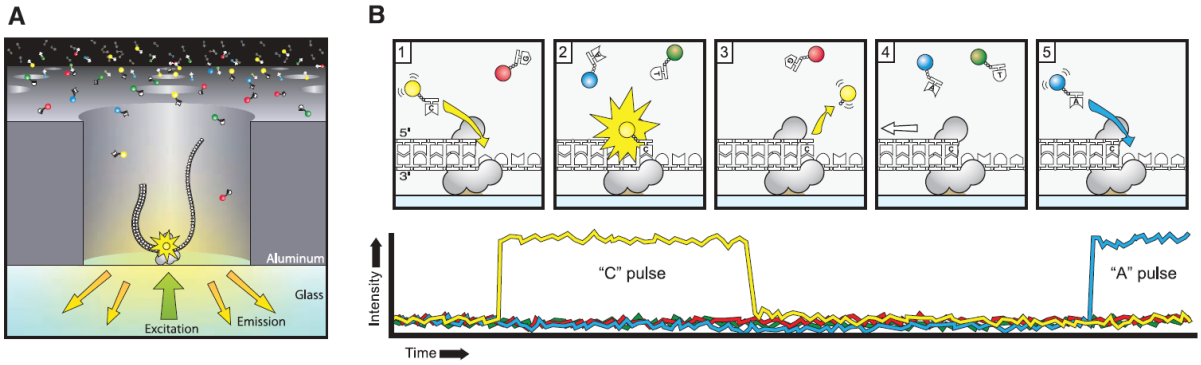

Fig. 4 (a) The zero-mode waveguide (ZMW). DNA polymerase is bound at the bottom of the ZMW. The DNA strand is in the ZMW. When a dNTP is built into the strand, a fluorescent pulse is released. The nucleotide can be detected. (b) The detection of the nucleotides in the ZMW and the intensity plotted versus the time. (1) A nucleotide with a fluorescent label is built in the template. (2) Hereby, the fluorescence output is increased. (3) Phosphodiester bond formation releases the dye-linker-pyrophosphate product, so the fluorescence pulse will ending. (4) The polymerase moved to the next position. (5) The next nucleotide is built in the template. (Eid 2009)

Oxford Technologies

Oxford Technologies has developed the Oxford nanopore, a relatively new method. (Fig. 5). The technology is on the market since 2014. The technology is based on an electrical fingerprint of each nucleotide, produced by the nucleotides passing through an alpha-hemolysin nanopore. (Frese 2013) The Oxford nanopore uses a protein nanopore. This protein is at the heart in strand sequencing. The hole is a few nanometres in diameter in the protein. The nanopore inserted into a membrane created by synthetic polymer. The membrane has very high electronic resistance. A potential is applied across the membrane resulting in a current flowing only through the aperture of the nanopore. Single molecules that end the nanopore cause characteristic disruptions. The molecule can be identified by measuring the disruptions. (Oxford nanopore technologies [2] 2014)

The DNA strand is broken in pieces. The pieces are mixed with buffer and a 'fuel mix' and loaded into the sequencer. An enzyme is designed to pull the DNA strand through the nanopore one base at a time. The enzyme binds to the end of the double-stranded DNA. The enzyme makes a single strand from the double strand. The single strand can through the nanopore. Then the characteristic disruption is created by the present of particular combinations of bases. So you can call the base. The

length of the DNA strand has no effect on the accuracy of the sequence. The system can read both strands. A number of nanopore experiments can be conducted at the same time by using an array chip. (Oxford nanopore technologies [2] 2014) (Mikheyev 2014)

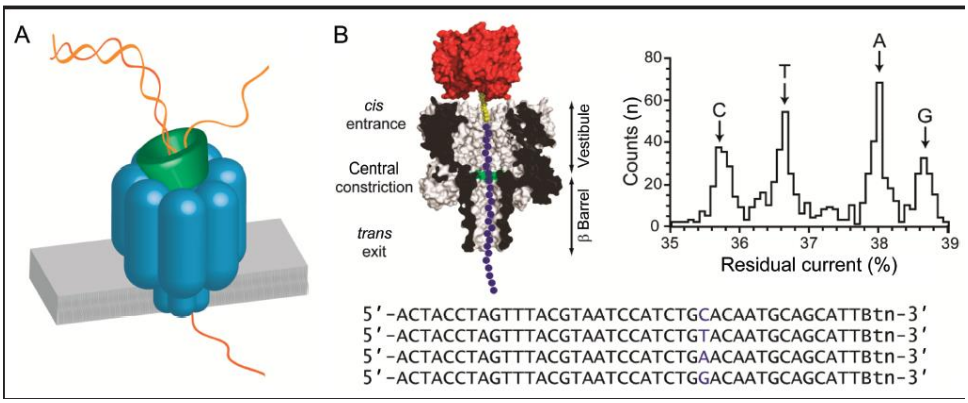

Fig. 5 (a) Schematic overview of the nanopore. The enzyme makes single DNA from the double DNA. It pulls the strand through the nanopore and read one base by one. (b) The strand inside the nanopore. The blue circles are the homopolymeric DNA oligonucleotides inside an αHL pore (grey) through the use of a biotin (yellow)-streptavidin (red) linkage. The central constriction is green. (Stoddart 2009) (Bayley 2014)

Discussion

Next-generation sequence is becoming increasingly important. A lot of companies have developed a next-generation sequence method. The companies use different methods. The next-generation sequence methods are cheaper, faster and less laborious as compared to the Sanger sequence method. But there are differences between the different next-generation sequence methods. Some are faster than another or more reliable than another.

Oxford Technologies has developed the latest method. The technique is on the market since 2014. Oxford Technologies could be a large player on the sequence market because they use a whole other technique than all other companies. Pacific Bioscience uses also another technique than the other companies. At the moment Illumina dominate the sequence market. Illumina has the most difference methods for many applications. They also have the most reliable method but the time it takes to read a sequence is quite long. If Oxford Technologies and Pacific Bioscience improve their reliability, they can be great players on the market. The sequence time is shorter of the Oxford Technologies and Pacific Bioscience than of Illumina. Oxford Technologies and Pacific Bioscience have to improve their reliability to get great players on the market. This will take a while because Illumina has a great name and the most companies know the name Illumina. Pacific Bioscience and Oxford Technologies have something to do, but also Illumina. If Illumina wants to stay the biggest player on the market, they have to keep innovating.

The Roche 454 is still on the market but Roche will stop supporting the Roche 454 sequencing instrument in 2016. An advantage of the Roche 454 was the long-read length but Pacific Bioscience and Oxford Technologies can read more bases. The advantage of the Roche 454 is gone. Ion Torrent, the fastest method, will get on the market but will not the biggest player because the method is expensive. Pacific Bioscience and Oxford Technologies use a very different technique than the Ion torrent. The Ion torrent is not a very different technique than the sequence methods on the market because the Ion torrent also uses a PCR.

At the moment everyone is curious about the method of the Oxford Technologies. It is a new technique and a lot of people are enthusiastic. The technique is not working optimal because of the high error rate. If Oxford Technologies improves the high error rate, they can get a great player on the market of sequencing. The method is cheap, fast and you do not use a large apparatus but a small chip. So you do not need a lot of space in your laboratory.

References

Bayley, H. “Nanopore Sequencing: from imagination to reality.” *Clinical chemistry*, 2014: 25-31.

Bhattacharya, T., Elmore, W.M., Qian, Z., & Zhang, L. *Impacts of genome sequencing technologies [video file].* Worcester Polytechnic Institute. 29 April 2012. https://www.youtube.com/watch?v=_ApDinCBt8g (accessed December 29, 2014).

Branton, D., Deamer, D.W., Marziali, A., Bayley, H., Benner, S.A., Butler, T., Ventra, M.D., et al. “The potential and challenges of nanopore sequencing.” *Nature biotechnology*, 2008: 1146-1153.

Brown, Terence. *Genomes.* Oxford: Wiley-Liss, 2002: 165.

Dressman, D., Yan, H., Traversco, G., Kinzler, K. W. & Vogelstein, B. “Transfroming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations.” *Proceedings of the National Academy of Sciences of the United States of America*, 2003: 8817-8822.

Eid, J., Fehr, A., Gray, J, Luong, K., Lyle, J., Otto, G. et al. “Real-time DNA sequencing from single polymerase molecules.” *Science*, 2009: 133-138.

Fedurco, M., Romieu, A., Williams, S., Lawrence, I. & Turcatti, G. “BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies.” *Nucleic acids research*, 2006: e22.

Frese, K.S., Katus, H.A., Meder, B. “Next-generation sequencing: from understanding biology to personalized medicine.” *Biology*, 2013: 378-398.

Glenn, T.C. “Field guide to next-generation DNA sequencers.” *Molecular ecology resources*, 2011: 759-769.

Grada, A., Weinbrecht, K. “Next-generation sequencing: methodology and application.” *journal of investigative dermatology*, 2013: 248.

Hayden, E.C. *Data from pocket-sized genome sequencer unveiled.* 14 February 2014. http://www.nature.com/news/data-from-pocket-sized-genome-sequencer-unveiled-1.14724 (accessed December 31, 2014).

Hutchison, C.A. “DNA sequencing: bench to bedside and beyond.” *Nucleic acids research*, 2007: 6227-6237.

Illumina . *Targeted RNA sequencing.* 2014. http://www.illumina.com/applications/sequencing/rna/targeted-rna-expression.html (accessed December 29, 2014).

Illumina. “Illumina Sequencing Technology - Highest data accuracy, simple workflow, and a broad range of applications.” 2010.

Illumina inc. *Illumina Sequencing Technology [video file].* 23 October 2013. https://www.youtube.com/watch?v=womKfikWlxM (accessed December 29, 2014).

International Human Genome Sequencing Consortium. “Finishing the euchromatic sequence of the human genome.” *Nature*, 2004: 931-945.

Karow, J. *Genome web - Following Roche's Decision to Shut Down 454, Customers Make Plans to Move to Other Platforms.* 22 October 2013. https://www.genomeweb.com/sequencing/following-roches-decision-shut-down-454-customers-make-plans-move-other-platform (accessed December 27, 2014).

Kircher, M., Kelso, J. „High-throughput DNA sequencing - concepts and limitations.” *Bioessays*, 2010: 524-536.

Krol, A. “Nanopore sequencing is her to stay.” *Bio-IT world.* 2014. http://www.bio-itworld.com/2014/12/22/nanopore-sequencing-here-stay.html (accessed December 29, 2014).

Levene, M.J., Korlach, J., Turner, S.W., Foguet, M. Craighead, H.G. and Webb, W.W. “Zero-mode waveguides for single-molecule analysis at high concentrations.” *Science*, 2003: 682-686.

Li, H. “A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data.” *Bioinformatics*, 2011: 2987-2993.

Life Technologies. *Ion Torrent next-gen sequencing technology [video file].* 19 May 2014. https://www.youtube.com/watch?v=WYBzbxIfuKs (accessed December 29, 2014).

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., et al. “Comparison of next-generation sequencing systems.” *journal of biomedicine and biotechnology*, 2012.

Mardis, E.R. “Next-generation sequencing platforms.” *Annual review of analytical chemistry*, 2013: 295-297.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., et all. “Genome sequencing in open microfabricated high density picoliter reactors.” *Nature*, 2005: 376-380.

Maxam, A.M., Gilbert, W. “A new method for sequencing DNA.” *Proceedings of the National Academy of Sciences of the United States of America*, 1977: 560-564.

McKenna, A., Hanna, M., Banks, E., et al. “The genome analysis tool-kit: a MapReduce framework for analyzing next-generation DNA sequencing data.” *Genome Research*, 2010: 1297-1303.

Mikheyev, A.S., Tin, M.M.Y. “A first look at the Oxford Nanopore MinION sequencer.” *Molecular Ecology Resources*, 2014: 1097-1102.

Morin, R.D., O'Connor, M.D., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A.L., et al. “Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells.” *Cold spring harbor laboratory press*, 2008: 610-621.

Morozova, O., Marra, M. A. “Applications of next-generation sequencing technologies in functioonal genomics.” *Genomics*, 2008: 255-264.

National Human Genome Research Institute. *Timeline of the human genome.* n.d. http://unlockinglifescode.org/timeline?tid=4 (accessed December 24, 2014).

Oxford nanopore technologies [1]. *RNA.* 2014. https://nanoporetech.com/technology/analytes-and-applications-dna-rna-proteins/rna (accessed December 29, 2014).

Oxford nanopore technologies [2]. "DNA: an introduction to nanopore sequencing." *Oxford nanopore technologies [video file].* 2014. https://nanoporetech.com/technology/analytes-and-applications-dna-rna-proteins/dna-an-introduction-to-nanopore-sequencing (accessed December 29, 2014).

Pacific Bioscience. "PacBio RS II Sequencing System." *PacBio.* 2014. http://files.pacb.com/pdf/PacBio_RS_II_Brochure.pdf (accessed December 29, 2014).

Rizzo, J.M, Buck, M.J. "Key principles and clinical applications of "Next-generation" DNA sequencing." *cancer prevention research*, 2012: 887-900.

Roberts, R.J., Carneiro, M.O. & Schatz, M.C. "The advantages of SMRT sequencing." *Genome Biology*, 2013: 405.

Roche Applied Science. "454 sequencing - get the complete picture." *454 sequencing.* 2010. http://my454.com/downloads/454SystemsBrochureSpread_Final.pdf (accessed December 28, 2014).

Roche Diagnostic Corporation. *Metagenomics.* 2014. http://454.com/applications/metagenomics/index.asp (accessed December 29, 2014).

Rothberg. *Ion Torrent Chemistry [video file].* 12 June 2010. https://www.youtube.com/watch?v=ZL7DXFPz8rU (accessed December 29, 2014).

Rothberg, J.M., Hinz, W., Rearick, T.M., Schultz, J., Milexki, W., Davey, M., Leamon, J.H., et al. "An integrated semiconductor device enabling non-optical genome sequencing." *Nature*, 2011: 348-352.

Samorodnitsky, E., Datta, J., Jewell, B.M. Hagopian, R., Miya, J. Wing, M.R., et al. "Comparision of custom capture for targeted next-generation DNA sequencing." *the journal of molecular diagnostics*, 2014: 64-75.

Sanger, F., Nicklen, S. and Coulson, A.R. "DNA sequencing with chain-terminating inhibitors." *Proceedings of the National Academy of Sciences of the United States of America*, 1977: 5463-5467.

Stoddart, D., Heron, A.J., Mikhallova, E., Maglia, G., Bayley, H. "Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore." *PNAS*, 2009: 7702-7707.

Thermo Fisher Scientific Inc. [1]. *Targeted RNA sequencing by Ion Torrent next-generation sequencing.* 2014. http://www.lifetechnologies.com/nl/en/home/life-science/sequencing/rna-sequencing/targeted-rna-sequencing-ion-torrent-next-generation-sequencing.html (accessed December 29, 2014).

Thermo Fisher Scientific Inc. [2]. *Ion Torrent next-generation sequencing construct library.* 2014. https://www.lifetechnologies.com/nl/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-construct-library.html#dna (accessed December 28, 2014).

Underwood, J.G., Lee, L., Clark, T.A., Brown, M., Olson, S., Graveley, B., Korlach, J., Travers, K. *Full length cDNA sequencing on the PacBio RS.* 2012. http://www.pacificbiosciences.com/pdf/Poster_Full_Length_cDNA_Sequencing_PacBioRS.pdf (accessed December 29, 2014).

Van Tassel, C.P., Smith, T.P., Matukumalli, L.K., Taylor, J.F., Schnabel, R.D., Lawley, C.T. et al. "SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries." *Nature Mehtods*, 2008: 247-252.

Vierstraete, A. "Next-generation sequencing for dummies." June 2012. http://users.ugent.be/~avierstr/nextgen/Next_generation_sequencing_web.pdf (accessed December 2014).

Voelkerding, K. V., Dames, S.A., Durtschi, J.D. "Next-generation sequencing: from basic research to diagnostics." *clinical chemistry*, 2009: 641-658.

Wang, Z., Gerstein, M., Snyder, M. "RNA-seq: a revolutionary tool for transcriptomics." *Nature Review Genetics*, 2009: 57-63.

Wetterstrand, K. „DNA sequencing cost." *National Human Genome Research Institute.* 31 October 2014. http://www.genome.gov/sequencingcosts/ (accessed February 6, 2015).

Wheeler, D.A, et al. "The complete genome of an individual by massively parallel DNA sequencing." *Nature*, 2008: 872-876.

Wu, R & Katser, A.D. "Structure and base sequence in the cohesive ends of bacteriophage lambda DNA." *journal of Molecular Biology*, 1968: 523-527.