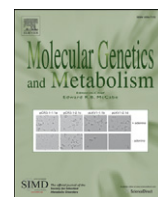




Contents lists available at ScienceDirect

Molecular Genetics and Metabolism

journal homepage: www.elsevier.com/locate/ymgme



Minireview

A glimpse into past, present, and future DNA sequencing



Marcos Morey^a, Ana Fernández-Marmiesse^a, Daisy Castiñeiras^a, José M. Fraga^b,
María L. Couce^{b,*}, José A. Cocho^a

^a Unidad de Diagnóstico y Tratamiento de Enfermedades Metabólicas Congénitas, Hospital Clínico Universitario de Santiago, A Choupana s/n, 15706 Santiago de Compostela, Spain

^b Unidad de Diagnóstico y Tratamiento de Enfermedades Metabólicas Congénitas, Departamento de Pediatría, Hospital Clínico Universitario de Santiago, Universidad de Santiago de Compostela, A Choupana s/n, 15706 Santiago de Compostela, Spain

ARTICLE INFO

Article history:

Received 27 February 2013
Received in revised form 30 April 2013
Accepted 30 April 2013
Available online 11 May 2013

Keywords:

Next generation sequencing
DNA sequencing
Genome sequencing
High throughput genomics
Third generation sequencing
Sequencing technologies

ABSTRACT

Current advances in DNA sequencing technologies are dropping down sequencing cost while increasing throughput at a pace never shown before. Past-decade great milestones, as the establishment of a reference human genome (amongst others) and large-scale human genetic variation study in the 1000 Genome project are, in conjunction with the use of these techniques, triggering advances in many areas of basic and applied science. These tools, stored in and combined with the vast amount of information present in biological online databases are, with the use of automated interpretation and analysis tools, allowing the fulfillment of increasingly ambitious studies in many areas and also are democratizing the access to information, interpretation and technologies, being the first opportunity for researchers to assess the influence of genetics in complex events as multifactorial diseases, evolutionary studies, metagenomics, transcriptomics, etc. In this review, we present the current state of the art of these technologies, focusing on second generation sequencing, from sample and library preparation to sequencing chemistries and bioinformatic software available for final data analysis and visualisation, with its possible applications. We also make an overview of first and third generation, due to its historical importance and for being the upcoming future tools for genetic analysis, respectively.

© 2013 Elsevier Inc. All rights reserved.

Contents

| | | |
|--------|--|----|
| 1. | Introduction | 4 |
| 2. | Classical sequencing: first generation | 4 |
| 2.1. | Sanger sequencing | 4 |
| 2.2. | Maxam and Gilbert sequencing | 4 |
| 2.3. | Limitations of classic sequencing techniques | 4 |
| 3. | Next generation sequencing: second generation | 6 |
| 3.1. | Introduction | 6 |
| 3.2. | Targeted enrichment strategies | 7 |
| 3.2.1. | Enrichment by PCR | 7 |
| 3.2.2. | Enrichment by hybridisation capture | 9 |
| 3.2.3. | Enrichment by circularisation | 10 |
| 3.3. | Clonal amplification | 10 |
| 3.3.1. | Bridge-PCR | 10 |
| 3.3.2. | Emulsion PCR | 10 |
| 3.4. | Sequencing chemistries: platforms | 10 |
| 3.4.1. | 454 Pyrosequencing | 10 |
| 3.4.2. | Qiagen-intelligent bio-systems sequencing-by-synthesis | 12 |
| 3.4.3. | Illumina sequencing-by-synthesis | 13 |
| 3.4.4. | Polony sequencing–Danaher Motion Polonator | 13 |
| 3.4.5. | Sequencing-by-ligation: SOLiD | 14 |

* Corresponding author.

E-mail addresses: marcosmorey84@gmail.com (M. Morey), amarmiesse@gmail.com (A. Fernández-Marmiesse), Daisy.Castineiras.Ramos@sergas.es (D. Castiñeiras), pdneonat@usc.es (J.M. Fraga), maria.luz.couce.pico@sergas.es (M.L. Couce), cochoja@yahoo.es (J.A. Cocho).

| | | |
|--------|--|----|
| 3.4.6. | Ion-torrent semiconductor sequencing | 15 |
| 3.4.7. | DNA nanoball sequencing: complete genomics | 15 |
| 4. | NGS bioinformatics | 16 |
| 5. | Applications of next-generation sequencing | 16 |
| 6. | Third Generation Sequencing: an insight into the near future | 17 |
| 6.1. | Single-molecule real-time sequencing | 18 |
| 6.2. | Real-Time DNA sequencing using fluorescence resonance energy transfer | 19 |
| 6.3. | Direct imaging of DNA using electron microscopy | 19 |
| 6.4. | Direct imaging of DNA sequences using scanning tunneling microscope tips | 19 |
| 6.5. | DNA sequencing with nanopores | 19 |
| 6.6. | Transistor-mediated DNA sequencing | 20 |
| 6.7. | Conclusions and future perspectives | 20 |
| | Conflict of interest | 20 |
| | References | 20 |

1. Introduction

As we move into an era of personalised medicine and complex genomic databases, the demand for development of new and existing sequencing technologies is constant. Although we cannot yet sequence an individual genome for \$1000, novel approaches are reducing the cost per base and increasing throughput on a daily basis [1]. The establishment of the human reference genome in 2001 [2,3] and the achievements of large sequencing projects such as the 1000 Genome Project [4] are catalysing advances in human genetics. A broad range of areas of basic and applied research are benefitting from such advances, notably genetic diagnostics[5], biotechnology [6], microbiological studies [7], forensic biology [8], and systematics and taxonomy [9,10]. Advances in sequencing methodologies, moreover, are changing the ways in which scientists analyse and understand genomes, while the results that they yield are disseminated increasingly widely through science news magazines [11].

In the medical sciences, advances in our knowledge of the genetic basis of pathologies have changed the way in which such entities are understood. Thus, disease has gone from being an individual-specific to a familial phenomenon, in which genetic alterations (mutations) can be genealogically traced [12]. Such progress has led to the development of numerous applications for disease prevention and diagnosis, prenatal screening [13], and genetic counseling.

Historically, distinct generations of sequencing chemistries can be distinguished. Classical methods began in the 1970s [14,15], when the first enzymatic methods became widely used for most applications [16]. Due to their inherently limited throughput, however, large studies are not feasible that methodologies. New technologies, known as Next-Generation Sequencing (NGS) technologies, or Second Generation Sequencing, have recently been developed [17]. Using these, it is possible to analyse complete genomes [18], exomes [19], and large gene panels [20], complete RNA transcripts [21], and also to carry out a range of other techniques, including chromatin immunoprecipitation [22]. In medical practice, complex or rare diseases that previously were assessed in terms of one or a few putative causal genes can now be more thoroughly analysed. Using modern technologies it is thus possible to carry out more rigorous research with unambiguous outcomes and to diagnose genetic-based diseases [23,24].

In this review, we describe the state of the art of NGS technologies and their potential applications. We also briefly summarise the current state of Third Generation Sequencing technologies, still very much in their infancy [25].

2. Classical sequencing: first generation

2.1. Sanger sequencing

In the 70s, Sanger and Coulson developed what is known as enzymatic DNA sequencing, or Sanger Sequencing, which utilises DNA

polymerase [16], in contrast to previous non-enzymatic approaches [15]. In 1975 they published their sequencing method, known as ‘plus and minus’ and the sequence of bacteriophage ϕ X174 [14]. Two years later the same authors presented the ‘chain termination method’, a more efficient and easier variation on the ‘plus and minus’ method [16] which utilised radioactively or fluorescently [26] labeled dideoxynucleotides acting as chain terminators [27]. The template DNA to be sequenced is obtained by in vivo (Fig. 1A) or in vitro cloning (Fig. 1B) and added to four parallel reactions. Each reaction contains four unmarked deoxynucleotides and one dideoxynucleotide marked by its lack of 3’OH, which impedes chain elongation. The template elongates by incorporation of deoxynucleotides in random positions: this process generates amplicons of any size which can then be spatially separated in parallel acrylamide gels to obtain a sequence of up to a few hundred bases [16].

This technique has been since been optimised as follows: 1) using a nucleotide-specific fluorescent dye, which allows the chain-termination reaction for the four dideoxynucleotides to be carried out in a single reaction (Fig. 1C) [26]; 2) replacement of conventional acrylamide gels with polyacrylamide gels in capillary electrophoresis, to shorten run times [28]; 3) automatic laser fluorescence detection (Fig. 1D) [29].

Even today, multiplexing and miniaturisation are being used to reduce reagent volumes and consumable costs while increasing throughput. Thus, in microfluidic Sanger sequencing the entire thermocycling amplification of DNA fragments and separation by electrophoresis is done on a single slide of approximately 10 cm diameter. While this reduces reagent usage as well as cost [30], further work is needed to make this variation of Sanger Sequencing commercially viable [31].

2.2. Maxam and Gilbert sequencing

In the mid 70s, Maxam and Gilbert developed a non-enzymatic DNA sequencing method. Samples to be sequenced require less complex preparation than in Sanger Sequencing, but reads are shorter (maximum 100 bp). The process itself is technically complex and involves the use of hazardous chemicals.

The method consists of selectively fragmenting the region to be sequenced, for each of the four nitrogenous bases or combinations thereof, and loading the resulting fragments onto a polyacrylamide gel. Depending on the sizes of the fragments and how they are cut, the radioactively labeled fragments can be read to obtain the sequence in a similar way to Sanger sequencing [15].

2.3. Limitations of classic sequencing techniques

The main limitation of both these classical sequencing techniques is their low throughput, due to template preparation and in the case of Sanger Sequencing also to carrying out the enzymatic reaction. Each run in Sanger Sequencing can sequence up to 1000 bp, and with

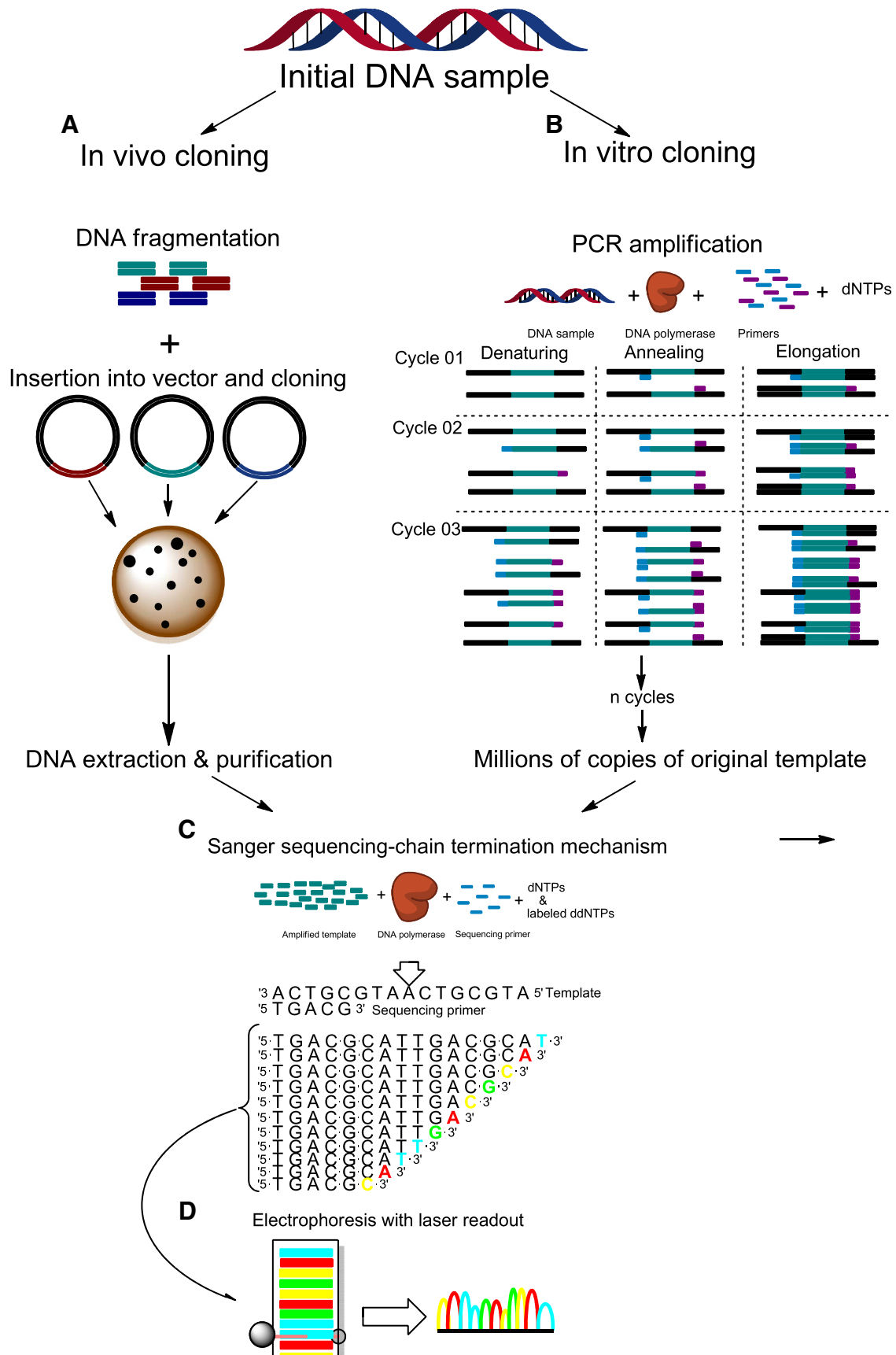


Fig. 1. Schematic representation of Sanger sequencing workflow. A1) The sample is first enzymatically digested and cloned into bacteria. Many copies of a given initial DNA fragment are collected from each bacterial clone. A2) With the appearance of the PCR in mid 80s the process of DNA amplification were significantly shortened. Millions of copies of the initial target are obtained in a few hours. B) Chain termination method relies on the ability of ddNTP's to stop chain elongation. The differently-sized products are separated in an electrophoresis. In this schema each ddNTP is fluorescently-labeled. Signal detection can be automated by the use of a laser/computer.

an automated sequencer 384 sequences can be run in parallel with a throughput of 80–100 kb per hour. Due its singleplex nature, Sanger Sequencing is not a hardly scalable process. In 1985, reading a single base cost \$10, while in 2005, the various improvements reading 10,000 bases cost the same. However, large projects such as the Human Genome Project still required vast amounts of time and resources [2]. Another limitation of First Generation Sequencing is that variants present at low frequency, such as mosaics, are difficult to detect due to high background levels. Finally, compared with modern technologies, the cost per base is still high [32].

The above limitations are important to consider when carrying out studies using these older technologies. However, due to the development of new approaches to DNA sequencing in which millions or even billions of reads can be carried out in parallel, the scenario is changing, to allow a large number of sequences to be obtained in a short time.

3. Next generation sequencing: second generation

3.1. Introduction

The first NGS equipment became available in 2004. The chance to sequence an entire genome at an unprecedented speed and for a reasonable cost was used to address previously intangible questions, for example differential susceptibility to disease among individuals [33], the influence of the microbiome on health [34], understanding cancer genomes [35], the roles of rare variants in common diseases

[36]. These new avenues are also revealing previously unknown genomic variation [37,38].

Indeed, NGS DNA sequencing is somewhat of a tipping point in research. Previously, research questions were ahead of what technology could address. But now, technology can not only address previously formulated biological questions, but is rapidly giving rise to new ones.

All NGS chemistries have in common the cyclic parallel reading of clonally amplified and spatially separated amplicons [39]. There is a significant conceptual change with these new approaches: in classical sequencing, a single amplicon from a single sample is amplified and a unique sequence is obtained. If 100% of the sample is amplified, 100% can be sequenced each time. NGS chemistry, on the other hand, is a massive representation of many random events. While it is impossible to drive them all in one direction, the enormous size of the process means that all sequences are likely to be represented. These processes usually follow a Poisson distribution.

There are key factors in the definition and evaluation of any Next-Generation technology platform: 1) read length, 2) throughput, 3) read accuracy (measurable using the Phred Score [40]), 4) read depth (number of times each base is sequenced in independent events), and 5) cost per base. These methodologies have brought down the costs of sequencing per base in recent years, from \$0.01 in 2004 to about \$0.0001 in 2006, and these continue to fall as capacity rises [1,32] (Fig. 2 [41]). Thus, the next bottleneck to consider will be the management and interpretation of the vast amount of data generated [17].

The main weakness of NGS is that the read length is not as long as with the previous techniques, due mainly to the progressive decline

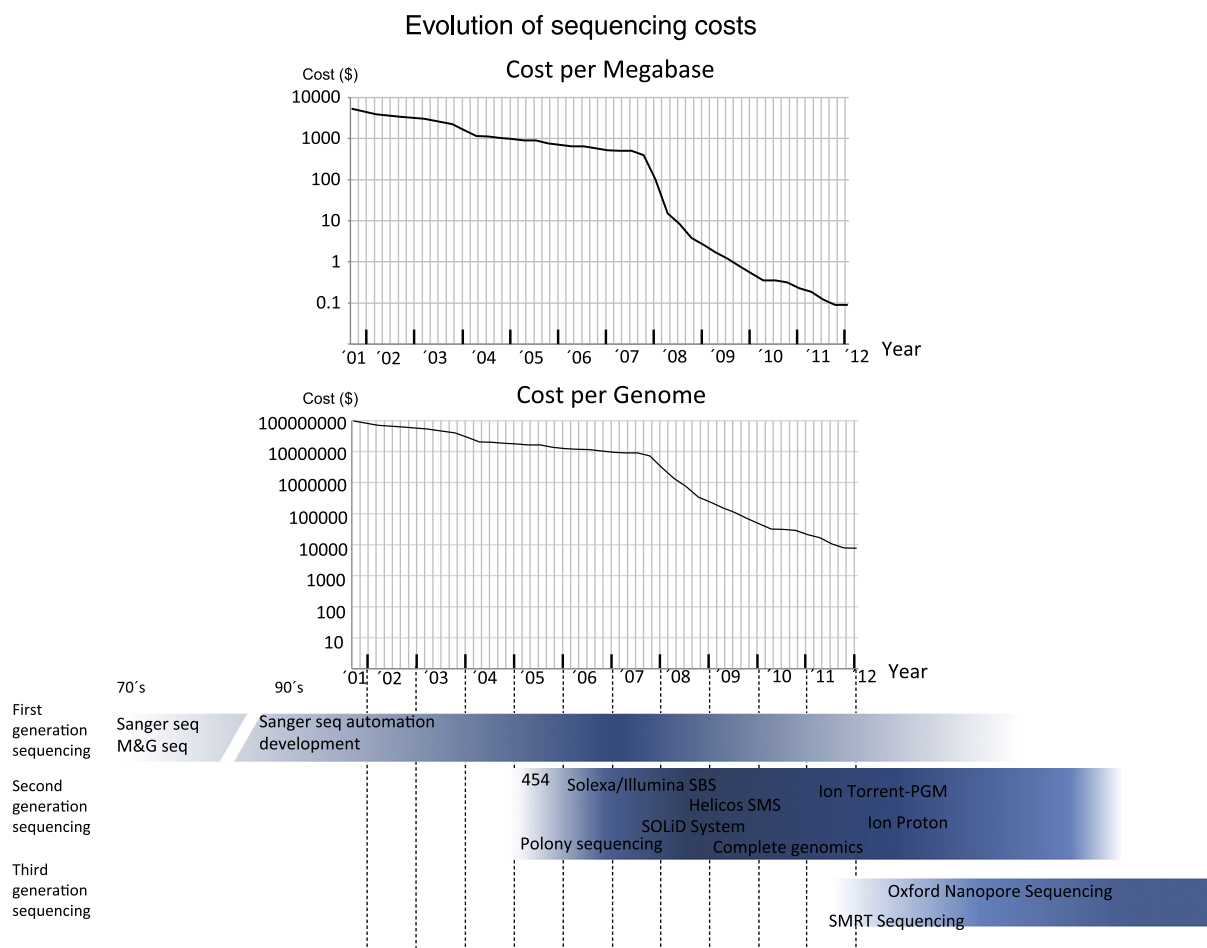


Fig. 2. Representation of the drop-down cost per Mb sequenced. Notice that in the last ten years, costs have move down five orders of magnitude, being the cost per Mb since Mid-08 less than 10\$.

in efficiency of the sequencing chemistry during the run. This also results in an asynchronous read elongation for any given amplicon clone (dephasing or prephasing) and a correspondingly ambiguous fluorescent signal [42]. This asynchrony increases with the number of sequencing cycles, finally reaching a point at which the data of those cycles are unusable. Short reads are a problem because they are difficult to align with the reference genome. Another weakness of NGS is the use of PCR which can itself potentially introduce bias [43].

NGS equipment provides a streamlined sample pipeline in comparison to previous technologies. In fact, one of the key advances of NGS sample preparation is its capacity to perform highly multiplexed reactions. Classical DNA sequencing, due to its 'one sample-one PCR-one sequence' nature only allows uniplex reactions.

The objective of NGS sample preparation is clonal amplification of single strands of the initial DNA. Thus, relatively significant amounts (μg) of initially high molecular weight DNA, with high purity levels (260 nm/280 nm = 1.8, 260 nm/230 nm = 2), are required. However, these levels depend on the application, and newer protocols tend to be less strict, enabling the use of DNA from a wide range of sources (paraffin-waxes, tissue, frozen blood, stained paper, etc.).

This review focuses mainly on DNA sequencing, with the techniques described below being applied mainly to those ends, rather than to the wide range of other sequencing applications. The different types of DNA libraries used in NGS will be described (Fig. 3). These are usually obtained by annealing specific DNA adaptors/bases to fragmented and size-selected DNA, although the precise mechanism of preparation depends on sequencing chemistry and the size of the region of interest.

Regardless of technology, there are several common steps in DNA sequencing. First, the DNA sample is usually fragmented to a target size (150 to many hundred base pairs), depending on the platform read length and chemistry used. This fragmentation is usually carried out by mechanical methods, i.e. sonication or nebulisation. The resulting fragment sizes follow a Poisson distribution. Currently, this step is the first bottleneck in processing a large number of samples. Methods of enzymatic sample fragmentation are currently being commercialised, offering the possibility of easily parallelizing the process [44]. DNA fragment-size distribution can be optionally refined via gel-based size selection [45] or SPRI bead-size selection [46], to improve library quality [47]. Universal DNA adaptors or bases can be ligase-bound at both ends of each DNA fragment, to allow PCR amplification to be performed either with a single pair of primers or to attach library fragments to surfaces by using adaptor-complementary oligos. Unlike in traditional sequencing, bacterial cloning is therefore no longer required. An additional variant is the addition of poly-A tails to one end of the fragment; this is used mainly in single molecule sequencing (SMS) technologies, which will be described below [48]. To minimise the introduction of PCR-based biases and maintain coverage uniformity, during sample preparation, PCRs are performed with as few cycles as possible [49]. Thus, each technology prepares DNA fragments from the initial sample as a so-called fragment library, on which to perform single-strand clonal amplification and sequencing reactions.

Currently, different types of DNA fragment libraries can be distinguished: 1) a fragment Library; this consists of variable-size DNA fragments with universal adaptors attached at both ends (Fig. 3A). From each DNA fragment in the library a single read of variable length is obtained (Fig. 3A1). Oversampling at both ends instead of only at one, to generate paired-end reads can be considered an improvement on the Fragment library (Fig. 3A2). Fragment libraries and Paired-end libraries are similarly built, but Paired-end can be used to obtain the information necessary to improve the detection of small indels. Repetitive regions are difficult to process in either case [39]. 2) Barcoded Library; the addition of a specific DNA sequence ('barcode') to each fragment allows sample multiplexing (Fig. 3B) [50]. These can be added during library construction, being attached as adaptors to the DNA fragments by ligation or polymerisation during an amplification process with barcoded primers. Barcodes are thus sample-specific

sequences that allow an increase in the number of samples per run by allowing the platform to distinguish them spatially. The platform throughput is divided between the total number of samples [51,52]. As with Fragment Libraries, a single tag (Fig. 3B1) or a pair-end tag can be obtained (Fig. 3B2). Barcodes also allow sample pooling during library preparation, saving time and reagents. 3) Mate-Pair Library; this consists of two DNA fragments, commonly separated by an internal adaptor and flanked by universal adaptors (Fig. 3C). Due to its construction mechanism both fragments are 'mates', being derived from the ends of the same initial fragment, and the distance between them is known. This method therefore allows colocalisation in the reference sequence and comparison of distances between readings, and is suited to detection of large rearrangements and analysis of repetitive regions [53].

The utility of each type library depends on the experiment. For targeted resequencing of specific genome regions a Fragment Library (with or without pair-end sequencing) is the adequate. However, for the sequencing of complex or repetitive regions, detection of large rearrangements, or *de novo* genome assembly, Mate-Pair Libraries are preferred because of their capacity for sequence colocalisation.

While the major advantage of these new technologies is their high throughput, for most applications it is neither necessary nor cost-effective to sequence the entire genome. Targeted enrichment, in which selected genomic regions are sequenced, are therefore often used for sequencing gene panels on demand. As less data is required per sample, Barcoded Libraries are used to process many samples per run.

Several approaches to target enrichment have been developed, which can differ with respect to several parameters: 1) sensitivity, or percentage of bases represented; 2) specificity, or percentage of the total sequence output that maps to the target regions; 3) uniformity, i.e. variability in sequence coverage across target regions; 4) reproducibility, or how well results correlate between replicates; 5) cost; 6) ease of use; and 7) the amount of DNA required [47,54].

3.2. Targeted enrichment strategies

3.2.1. Enrichment by PCR

Due to its ability to amplify a single DNA molecule with high specificity, PCR has been the most widely used technique for processing samples for sequencing; indeed Sanger sequencing relies on the reading of single amplicons [55]. For its application to NGS it is necessary to increase PCR throughput.

Long range uniplex PCR (up to 20 kb) is commercialised by several companies (e.g. Invitrogen Sequelprep [56], Qiagen SeqTarget [57]). These have the inconvenience that, before fragment library preparation, each reaction has to be cleaned, carefully normalised and pooled (Fig. 4A). Depending on the application, either short (100–200 bp) or long amplicons can be prepared (Fig. 4B); the latter are required for example in rapid screening of mutational hot spots.

Nano-scale uniplex PCR is commercialised by Fluidigm [58] and Raindance [59]. The Fluidigm platform comprises thousands of singleplex PCRs in parallel on a microfluidic PCR chip. Raindance, through its platform Rainstorm, is based on emulsion PCR [60] and uses microdroplet PCR. Each droplet supports an independent PCR, and a droplet population can thus carry out thousands of simultaneous reactions each with specific primer-pairs. After thermal cycling, the emulsion is broken down and the products are subjected to library construction [61].

An increasingly popular alternative to the options above is multiplex PCR, in which, the number of reactions is kept as low as possible for a high throughput. Multiplex PCR requires a careful primer design in order to allow several dozen primers per reaction without any primer interactions or non-specific amplifications.

Compared with other enrichment techniques, such as those which are probe-hybridisation based, targeted enrichment via PCR does not require such a high quality or quantity of DNA for sequencing regions of up to several hundred kilobases. It provides a very high enrichment

DNA libraries

Sequencing

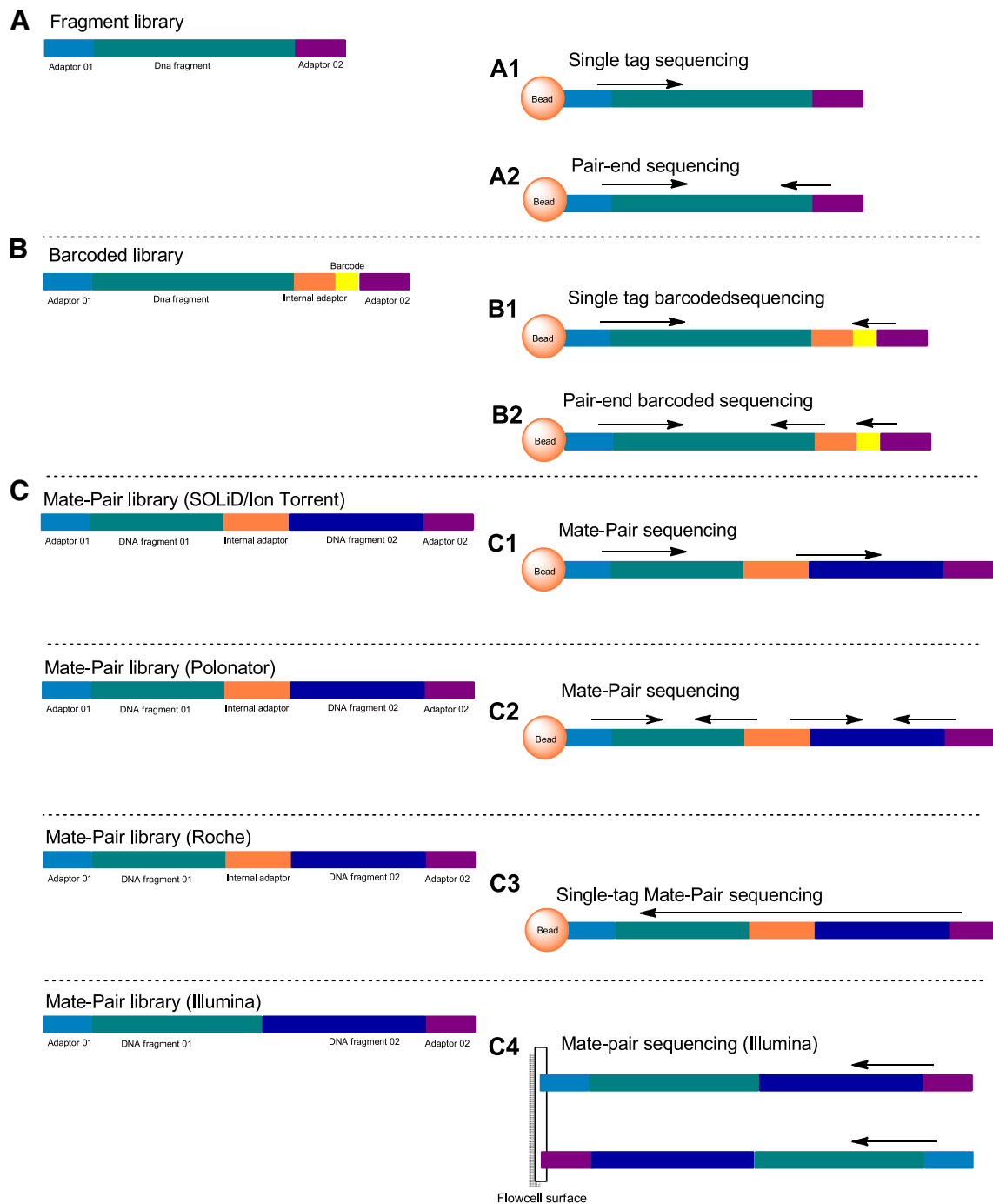


Fig. 3. Representation of the most used DNA libraries configuration for Next Generation Sequencing. A) A fragment library consist in the representation of the initial sample in DNA fragments of variable size (starting in 150 pb up to 1000) flanked by universal adaptors. One or two tags of sequence can be obtained per fragment (A1 and A2 respectively). B) A barcoded library includes an indexing, specific sequence in order to multiplex sample preparation and sequencing. Sequencing tags can be single or paired end (B1 and B2). C) A mate paired library consist in the representation of the initial sample in pairs of DNA fragments which separation in the reference sequence is previously known. Sequencing approach depends on the platform used (C1, C2, and C3).

ratio with few off-target reads. In the case of small amplicons it is also possible to bypass the initial step of sample DNA fragmentation [54]. If the regions to enrich are larger, the amount of DNA required grows linearly with the number of PCRs required.

The main weakness of target enrichment by PCR is that coverage uniformity may not be as good as with other techniques, being the

correct normalisation of PCR products crucial to improve it [47]. In the beginning of this methodology, it was difficult to perform enrichment by this approach if the region of interest involved many dispersed amplicons, due to issues with PCR-multiplexing, so it was more suitable to enrich kb-range contiguous regions due to the possibility to overlap the amplicons [62]. Currently, as the multiplexing capacity is being

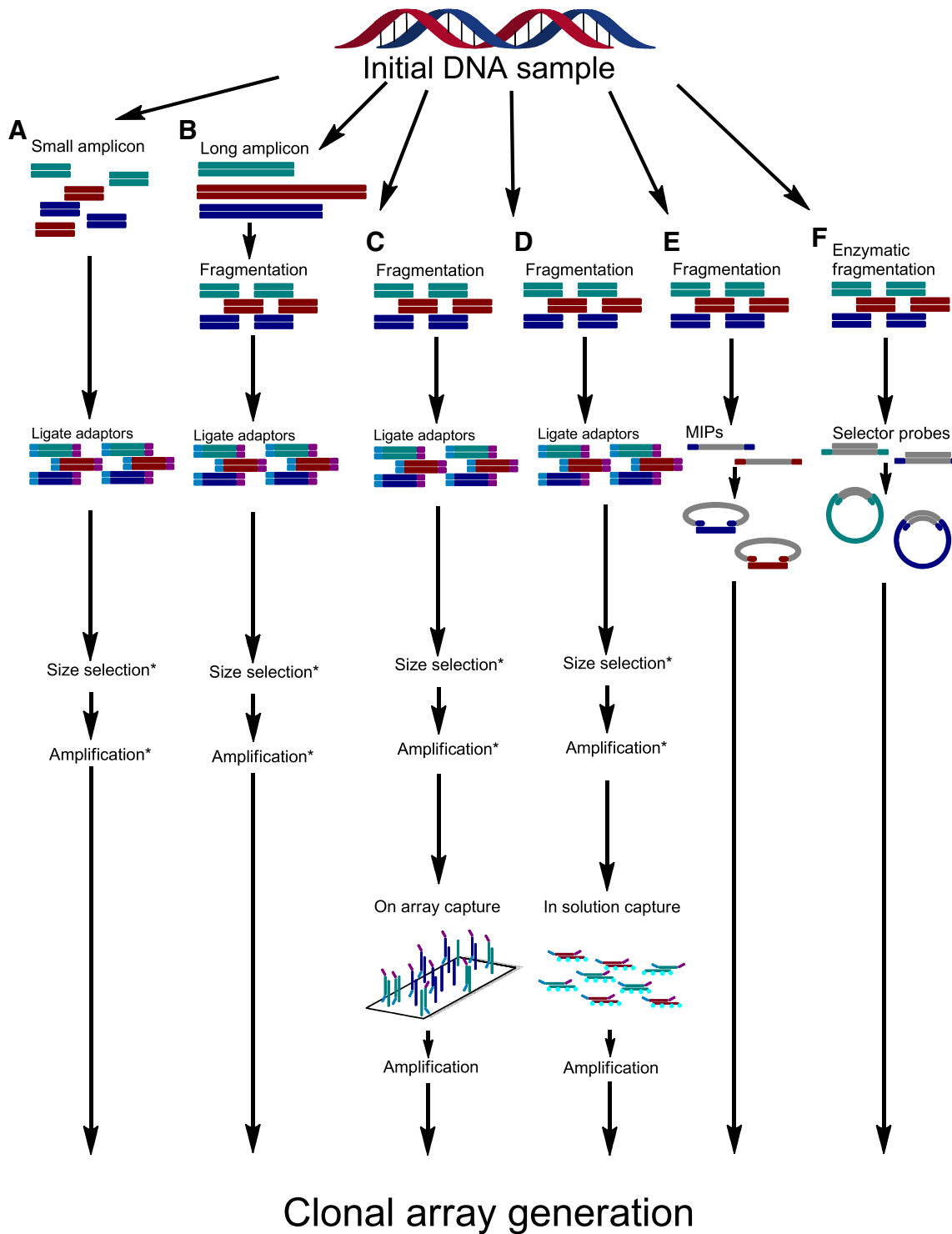


Fig. 4. DNA library preparation scheme in current Next Generation Sequencing. * Indicates non mandatory steps depending on the application.

increased by the improvement and automation of primer-designing, enrichment by PCR of numerous dispersed genomic regions is possible, being it commercialised by different companies, e.g.: Ampliseq, by Life Technologies [63], or Truseq amplicon [64], by Illumina.

3.2.2. Enrichment by hybridisation capture

Enrichment by hybridisation is based on knowledge acquired through microarray research [65] and applied to the enrichment of samples for NGS [66]. Thus, a DNA library is hybridised with probes

which represent the target region. Non-specific hybrids are removed by washing and the targeted DNA obtained is eluted.

Two alternative methods exist: array hybridisation and in-solution hybridisation. The main advantages of both, compared with multiplex PCR are the ability to capture a larger number of targets, and that the coverage obtained is more homogeneous. But the major drawbacks are, firstly, that working with microarrays usually necessitates additional equipment and, secondly, that due to the chemical nature of the process and regardless of the length of the captured region a large amount of starting DNA (10–15 µg) is usually required (Fig. 4C) [67,68].

Solution-based capture overcomes many of these disadvantages: much less DNA is needed, special equipment is not required, and the procedure is and is more scalable. The overall performance of solution-based enrichment has also been reported to be better (Fig. 4D) [69].

Roche Nimblegen was the first company to commercialise hybridisation enrichment methods, initially array-based, with capture regions of about 4–5 Mb. Later, Agilent Technologies also commercialised a similar approach, with similar performance. Currently, the major commercialisers are still Agilent (SureSelect [70]) and Roche Nimblegen (SeqCapEZ [71]), but a large number of protocols and vendors are also beginning to offer new capture options, for example Flexgen [72] and MYcroArray [73]. Applied Biosystems and Illumina (TargetSeq [74] and TruSeq [64], respectively) are also offering protocols integrated into the sample preparation platform workflow.

3.2.3. Enrichment by circularisation

Capture by DNA circularisation consists in the use of molecular inversion probes (MIP) [62] or selector probes [75]. It is based on the use of DNA probes, which are single-stranded oligonucleotides consisting of a common linker flanked by target-specific sequences. In MIPs the probes are annealed to the initial DNA sample, undergoing circularisation in the process. The hybridised ends of the probe flank the target region (Fig. 4E). The gap between the two ends is then filled and non-circularised DNA is removed by exonucleases.

One specific application of enrichment by circularisation, which differs from the previous approaches in terms of hybridising with target regions, lies in Selector Probes. These probes are designed to accommodate the restriction pattern of the target region (Fig. 4F) [76]. Using these, there is no need to perform downstream steps of library preparation, because the sequencing information for NGS application can be included in the probe [77], which reduces processing time.

Initially, despite being very specific, this approach had low coverage uniformity due to inefficiencies in the capturing reaction itself [47]. Due to implementations mainly by the company Halo genomics [78], acquired by Agilent Technologies, enrichment by circularisation is now being commercialised.

3.3. Clonal amplification

After library preparation, NGS technologies need to clonally amplify the sample prior to its being sequenced, due to the need for an amplification of the signal generated in order to enable its detection during the sequencing process. This process aims to produce several copies of each DNA library fragment, but must also keep them separated in order to produce an unambiguous, monoclonal signal in the sequencer [39,42]. Several *in vitro* techniques for clonal amplification have been developed.

3.3.1. Bridge-PCR

Bridge-PCR, also called two-dimensional PCR, is a variant of traditional PCR, developed during the 1990s and used first by Solexa [79] in its NGS platforms. Currently this clonal amplification approach is used by Illumina, which acquired Solexa in 2008.

Bridge amplification allows clonal amplification of a large number of DNA fragments simultaneously, using a solid oligonucleotide-coated surface (known as a 'flow cell') (Fig. 5A) [80]. The oligonucleotides are complementary to the linkers added to the DNA sample during library construction. The flow cell is a multi-channel (currently up to 8) sealed glass device. A different library can be added to each channel or the same one can be used in each, depending on the size of the region to be sequenced. DNA fragments from the library are denatured and then attached to the flow cell surface. Oligonucleotides of the flow cell are linked to the surface at its 5' end, leaving the 3' end free for the polymerase to act on. The resulting double-stranded-DNA is covalently attached to the surface. This double-stranded DNA is then denatured (Fig. 5A1) and the single strand flips over to hybridise to

adjacent primers, thus forming a bridge (Fig. 5A2). This newly closed amplicon is extended by polymerases to form a double-stranded bridge (Fig. 5A3). After denaturing, two copies of covalently bound single-stranded templates are obtained (Fig. 5A4) [81]. This cyclical process is repeated several times, producing approximately one million clonal copies of each initial fragment in the form of clonal clusters (Fig. 5A5). No primers are required in the reaction solution and clusters are spatially separated.

The separation between clusters depends directly in the initial quantity of DNA in the reaction. Higher concentrations allow larger numbers of clusters to be obtained, but also carry the risk of clusters being so close to each other that they interfere with each other. On the other hand, smaller amounts of DNA lead to a lower PCR yield [82].

After bridge amplification, clusters are denatured to give single strands. The sample is then ready for subsequent sequencing steps (Fig. 5C).

3.3.2. Emulsion PCR

Emulsion PCR (emPCR) is a variant of conventional PCR that was introduced into commercial use first by 454 Life Sciences, in the 454 sequencer. [83] Subsequent technologies have included it in their protocols, e.g. Applied Biosystems [84] in the SOLiD (Sequencing by Oligonucleotide Ligation and Detection) platform [85], Ion Torrent Personal Genome Machine (PGM), and Proton. This technology is also used in the open-source NGS platform Polonator [85].

emPCR is an elaboration of the classical amplification reaction, in which each DNA fragment is isolated in independent aqueous microreactors surrounded by an oil phase, and templates are amplified on primer-coated beads (Fig. 5B). An emPCR complex therefore contains millions of compartments separated from each other and each acting as microreactors, in which an independent single-molecule PCR reaction is performed. For its application to NGS, emPCR is usually carried out with beads carrying reaction primers on their surfaces, in order to localise the clonally amplified fragments of each reaction on the surface of one bead (Fig. 5B1) [86,87]. After the reaction, the emulsion is broken down with organic solvents and the beads are isolated by extraction of the aqueous phase.

Each bead will ideally contain multiple copies of a unique monoclonal fragment. However, depending on the amount of DNA added initially to the reaction, there might be more than one DNA fragment in each of the microreactors, leading to generation of polyclonal beads. These are unsuitable for obtaining specific base-interrogation signals, and thus for sequencing. Beads with no DNA fragment amplified on their surface can be eliminated via 'bead enrichment'. This process consists of the physical separation of amplified and non-amplified beads (Fig. 5B2) [88]. Polyclonal beads, however, cannot be eliminated using enrichment methods, but only by limiting the amount of DNA added initially to the reaction.

In subsequent steps, beads are deposited on a solid surface (Fig. 5B3) prior to sequencing (Fig. 5C) [89–91].

3.4. Sequencing chemistries: platforms

There are diverse chemistries commercially available. Despite the differences between them, they all rely in the same principles: base interrogation is performed on all DNA fragments cyclically and in parallel. Signal detection is performed by an optical system composed by a microscope with a Charge-Coupled-Device (CCD) camera plus a computer and storage system, or, in the case of semiconductor sequencing, by a semiconductor chip. Read length of these technologies is usually short, ranging between 35 and 700 bp, being the most usual currently 100 bp.

3.4.1. 454 Pyrosequencing

Developed in 1996 by the Stockholm Royal Institute of Technology [92] and commercialised for NGS in 2005 by 454 Life Sciences [89]

Clonal array generation

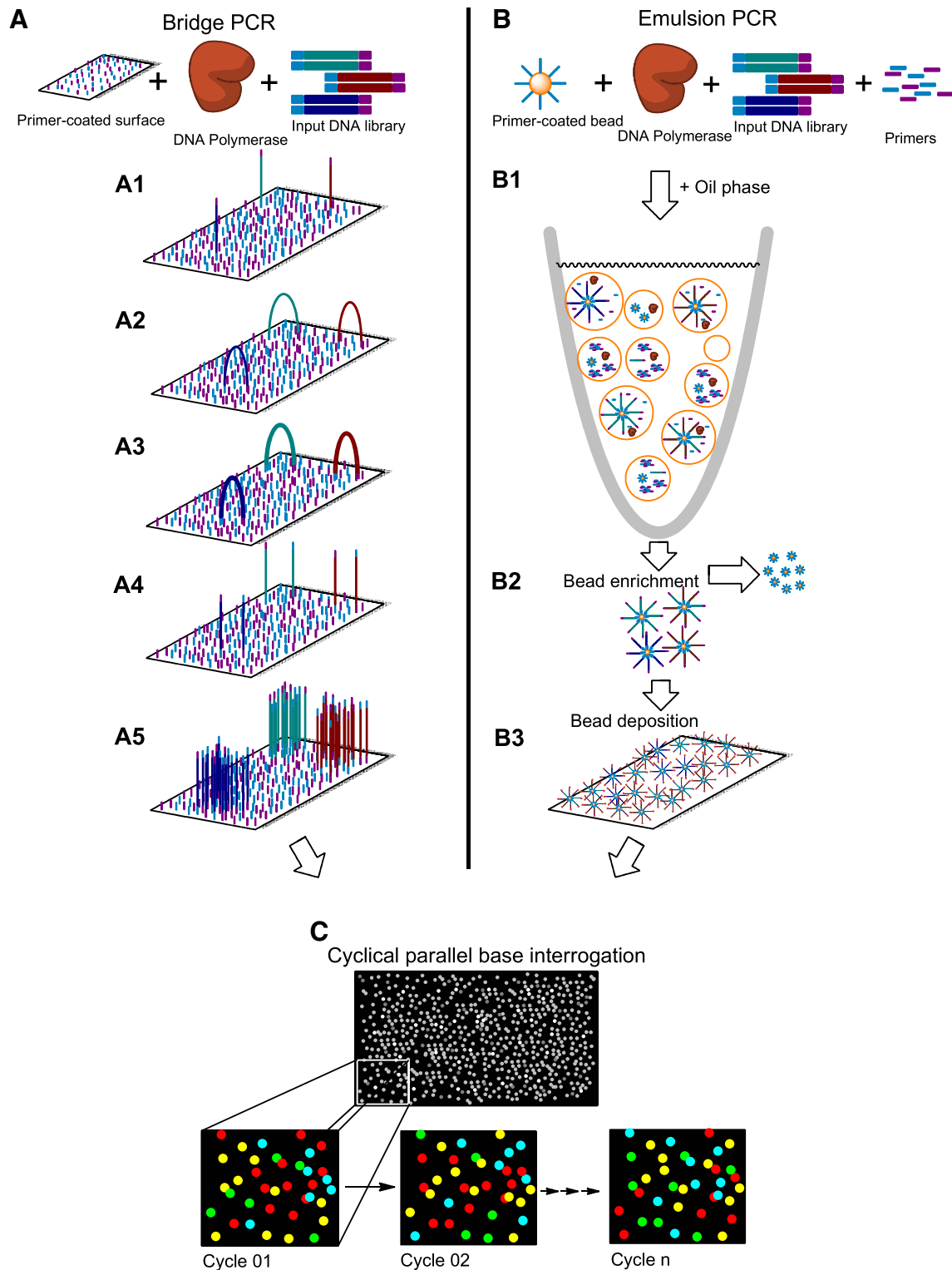


Fig. 5. Schema of the current main approaches for clonal array generation.

(currently part of Roche Diagnostics), pyrosequencing is based on detection and quantification of DNA polymerase activity, which is carried out using the enzyme luciferase. Thus, pyrosequencing can be considered a sequencing-by-synthesis method.

The library is first attached to beads via emulsion PCR and then deposited in a PicoTiterPlate which contains millions of wells of 44 μm diameter. Each well fits only one bead, of diameter 28 μm (Fig. 6A), although not all wells contain a bead [83].

Pyrosequencing

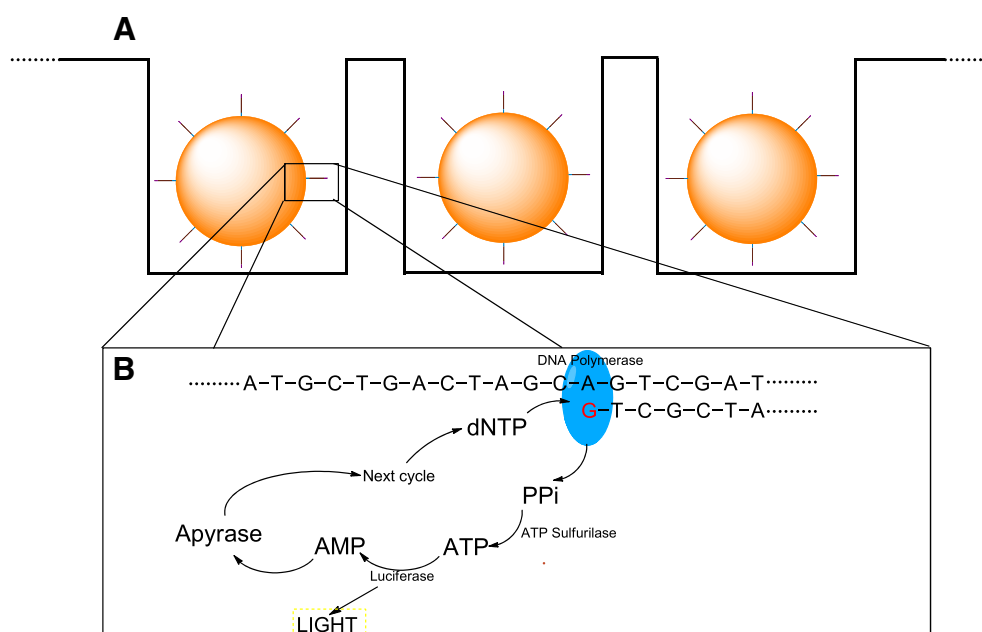


Fig. 6. Pyrosequencing steps. A) The DNA library is attached to beads and deposited in wells on a PicoTiterPlate. B) The objective of pyrosequencing is to measure the light emitted by the luciferase. To link it to the base addition to the growing strand of DNA a number of chemical reactions are necessary. First a known dNTP is added to the reaction and binds by a polymerase to the complementary 3' positions adjacent to a universal sequencing primer. This generates one free pyrophosphate per base added. ATP sulfurylase will generate ATP by combining this pyrophosphate with adenosine 5' phosphosulfate. This ATP will be the luciferase substrate to generate visible light. After light measurement over each bead, the enzyme apyrase will degrade the non-incorporated nucleotides and a new sequencing cycle will begin.

To start the sequencing reaction, a universal primer (complementary to one of the library adaptors), together with DNA polymerase, ATP sulfurylase, luciferase and apyrase are added to the plate. Then, a modified variant of ATP, which acts as a substrate for luciferase) is added to the reaction. This binds to the complementary 3' positions adjacent to the universal primer previously added to the reaction. The union is catalysed by DNA polymerase and generates one free pyrophosphate per base added. ATP sulfurylase then uses the released pyrophosphate to generate ATP by combining it with adenosine 5' phosphosulfate. The ATP is then used by the luciferase to generate visible light with an intensity equivalent to the quantity of bases incorporated (and thus available ATP) (Fig. 6B). After each ATP-3' binding and light-generating step, imaging takes place. This allows discrimination of the deposited library fragments that having a base complementary to any of those added. Apyrase then degrades the non-incorporated nucleotides, permitting another reaction cycle to begin, with incorporation of another dideoxynucleotide [93].

The data output of this sequencing platform is not as high as other approaches, having a yield of 700 Mb of sequence per run (23 h), approximately one million reads and a consensus accuracy at 15× coverage of 99.997%. The relatively low output is compensated for by the high read length obtained (up to 1000 bases) (Table 1) [39,83].

This read length is useful for performing de novo sequencing on unknown genomes due to its capacity to align complex regions and relatively low computational power that that requires. Moreover, it is possible to sequence up to 132 samples per run with the use of barcodes [94].

3.4.2. Qiagen-intelligent bio-systems sequencing-by-synthesis

Founded in 2005 at Columbia University (New York), Intelligent Bio-Systems (IBS) has developed its own sequencing-by-synthesis approach. Template preparation is based either on polony amplification or emulsion PCR.

The sequencing chemistry is very similar to Illumina's approach (see below), except that only a small fraction of the total nucleotides added to the reaction are labeled (1:100), which lower the costs of reagents while still detecting signal.

The IBS platform, known as Mini-20, claims to be able of sequence 20 samples in 20 individual flowcells, with a maximum output of 20 Mb reads per flowcell and reads of 2×100 (80 Gb of total output), in approximately 2 days [95]. A larger scale version of Mini-20, MAX-seq, has also been commercialised, although its status has been unclear since IBS was acquired by Qiagen in 2012 [96].

Table 1
Compared characteristics of the main NGS platforms.

| Platform | Clonal amplification | Sequencing chemistry | Read length (bp) | Throughput (Mb/run) | Run time |
|---------------------------|-------------------------------|--------------------------------------|--------------------|---------------------|----------|
| Roche-454 | emPCR | Pyrosequencing | 400–1000 | 700 | 23 h |
| Illumina | Bridge PCR | Synthesis | 150+150 paired end | 700 | 11 days |
| Danaher Motion Polonator | emPCR | ligation | 13+13 paired end | ~9000 | – |
| Applied Biosystem's SOLiD | emPCR | Ligation | 75+35 paired end | 300,000/flowcell | 14 days |
| Nanoball sequencing | Template circular replication | ligation | 35+35 paired end | 60,000/slide | 12 days |
| Ion Torrent-PGM | emPCR | Synthesis (semiconductor sequencing) | 100–200 | 1000 | 2.5 h |
| Ion Proton | emPCR | Synthesis (semiconductor sequencing) | 100–200 | 100,000 | 3.5 h |
| Helicos (HeliScope) | None (SMS) | Synthesis | 25–55 | 14,000/flowcell | 10 days |

While the above IBS platforms have not been widely used, Qiagen's GeneReader has recently reappeared at the Advances in Genome Biology & Technology Meeting (AGBT, Florida, 2013).

3.4.3. Illumina sequencing-by-synthesis

Illumina's sequencing-by-synthesis method is performed after a bridge-PCR amplification process. Sequencing takes place on the solid surface of the flowcell (Fig. 7A). Sequencing-by-synthesis consists of the polymerase-catalysed addition of reverse-terminator fluorescently labeled bases. Bases are added simultaneously to the reaction and compete to form a union with oligo-primed cluster fragments (Fig. 7B). Once a base is added, it prevents addition of subsequent bases, meaning that only one base will be attached in any one cycle.

An imaging step is performed after base incorporation, to record cluster-specific fluorescence (Fig. 7C). Each flowcell lane is divided in 'panels' or 'tiles', for a given cluster density (several thousand). Each image represents one tile. After each imaging step, 3' blocking is chemically removed and the process is restarted (Figs. 7D and E) [17,39,97].

This technology was developed in the early 2000s by the US company Solexa, becoming commercially available in 2006. Later, as Solexa became part of Illumina, the chemistry became part of their platforms. Currently, the maximum output available with this NGS sequencing chemistry is approximately 600 Gb per flowcell with 6 billion reads (read length 2×100 bases on a paired-end library) and an accuracy of 99.9%, in approximately 11 days (Table 1); there is also the option of using barcodes (up to 96 samples) [98].

3.4.4. Polony sequencing–Danaher Motion Polonator

Polony (polymerase colony) sequencing is an open-source sequencing chemistry developed at Harvard Medical School [99]. Protocols and software are freely accessible. The open-source implementation of Polony sequencing is the Danaher Motion Polonator G007 [90], but it has fallen increasingly into dis-use.

The protocol begins with the preparation of a DNA library (in this case a mate-paired library) and template amplification via emulsion-PCR [100].

Sequencing chemistry [101] begins by attaching a universal 'anchor' primer which is complementary to one of the library adaptors and

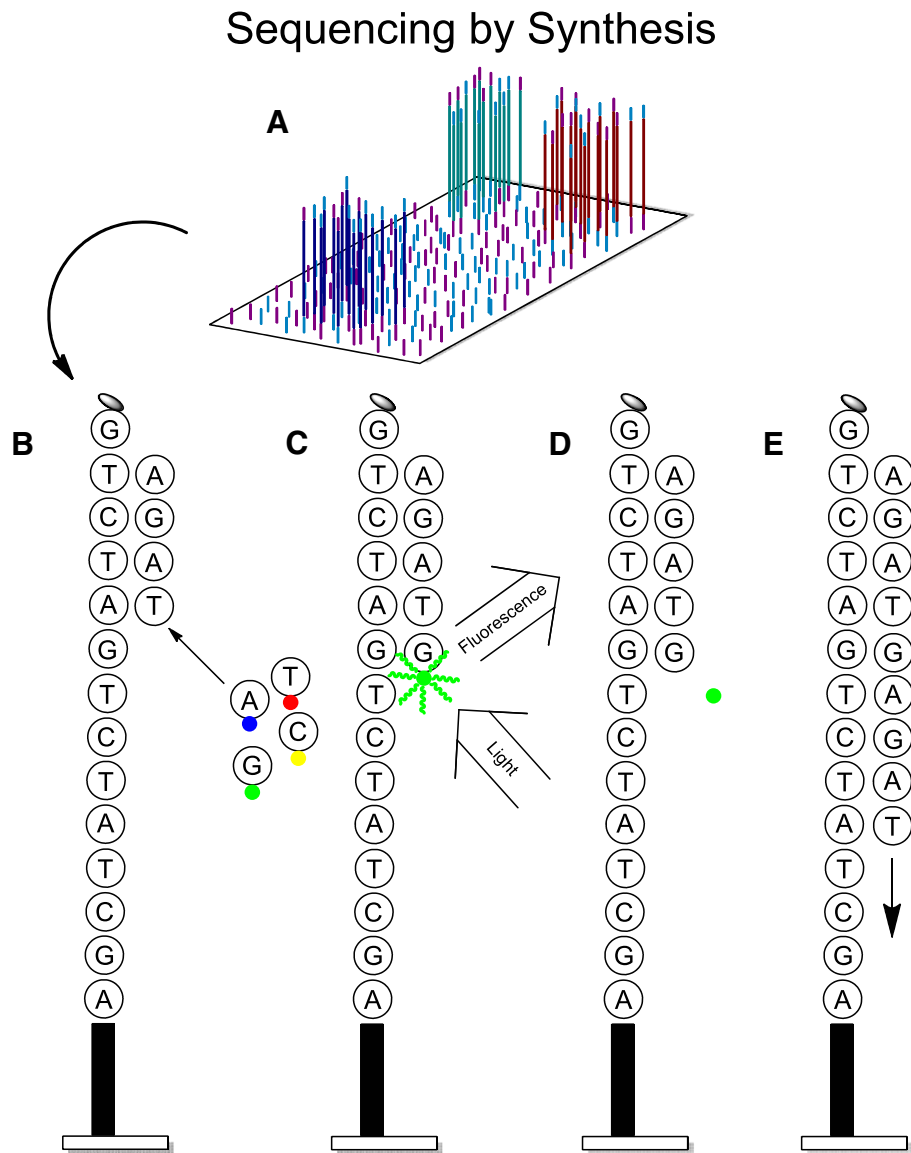


Fig. 7. Sequencing by Synthesis steps. A) Sequencing reactions take place in the clonal clusters on the flowcell surface. B) First, bases are added to the reaction, competing for the union to a primer attached to a universal adaptor on each library fragment. Bases are blocked at its 3' end, so only one is added. C) After incorporation, fluorescence is emitted and registered by the imaging system. D) 3' blocking is removed. E) The previous steps are cyclically repeated in order to sequence the desired number of bases.

leaving a free 3'OH or 5'P end. A ligase then attaches a base-specific fluorescently labeled nonamer in the position adjacent to that primer. The nonamer bases are degenerate, with the exception of one, whose position will be the one interrogated in the particular ligation cycle. Fluorescence is emitted and recorded by the imaging system after each ligation cycle. The universal anchor primer and nonamer are then separated from the DNA and a new sequencing cycle begins, with addition of an identical universal anchor primer and another nonamer. The position of the non-degenerated base changes each time, so the interrogated position is also different. The result is the generation of millions of 26 bases-reads (2×13 bases from each tag).

Due to its open-source nature, Polony sequencing is a very flexible technique with variable applications, protocols and reagents. However, while a lot of raw data are produced, only a small proportion are useful due to limitations in performance of the technique (Table 1) [85,102].

It is noteworthy that Polony ligase sequencing has been of significant importance in establishing the basis of other sequencing chemistries, including SOLiD sequencing [103].

3.4.5. Sequencing-by-ligation: SOLiD

Sequencing-by-ligation has its beginnings and basis in older sequencing strategies such as Massively Parallel Signature Sequencing [104] and the aforementioned Polony sequencing [85,101,102], and has been commercially available through Applied Biosystems, as the SOLiD system, since 2008.

In this technology, an emulsion-PCR is performed to generate clonal amplicons on the surface of beads, starting from a previously

constructed DNA library. Beads are enriched and attached to a glass surface in a random pattern (according to a Poisson distribution) and forming a dense array on which sequencing-by-ligation is then performed (Fig. 8A). First, a universal primer is attached to one of the library adaptors. Fluorescently labeled di-base probes of eight nucleotides in length compete for ligase binding to the DNA fragment adjacent to the universal primer (3' or 5', depending on the tag being sequenced). Di-base probes interrogate the two first bases simultaneously in each ligation cycle. The remaining base probes are degenerate. After ligation, the three final bases of the probe are cleaved so that the length of the probe is reduced to five nucleotides and di-base-specific fluorescence is emitted. After each ligation-cleaving step, an imaging step is performed to detect fluorescence. As in other NGS technologies, the glass array is divided into panels that are each represented as an image. After imaging, a new di-base probe is added and a new ligation cycle begun. By this process, the first two bases of each group of five are interrogated, at three-base interval.

After many cycles of ligation, the extended product and the universal primer are cleaved from the DNA fragment, thus resetting the sequencing reaction. A new universal primer, displaced one base towards the adaptor, is added, such that the sequence interrogation performed by the di-base probes is correspondingly displaced by one base [105]. This process of reset is repeated five times with five different universal primers (Fig. 8B). Finally, every base on each clonally amplified fragment will have been interrogated twice in independent ligations [39,103]. This two-base encoding system increases sequencing accuracy (Fig. 8C).

Recently, an optional sequencing chemistry add-on has been commercialised, introducing a kind of third-base encoding based on the

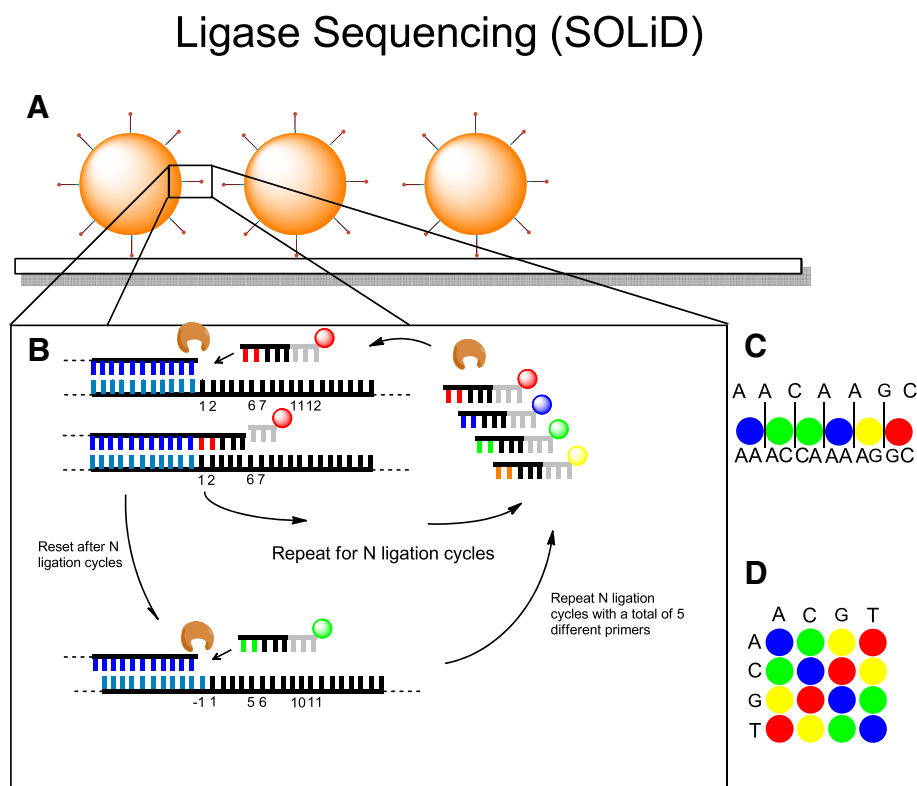


Fig. 8. SOLiD sequencing by ligation steps. A) Sequencing is performed on the surface of DNA-coated beads attached to a glass slide (also known as Flowchip). B) First, a universal primer is attached to one of the library adaptors. Di-base probes compete for the union and subsequent ligation to the primer-adjacent position. After ligation, the three final bases of the octamer is cleaved and fluorescence is emitted. After N ligation cycles (usually 10 but may vary) the reaction is reset and a new displaced primer is added to the reaction, so different bases are interrogated. This process of resetting and primer addition is performed five times per sequencing tag. By this, each base of the sequenced DNA fragments will be interrogated two times in independent reactions. C) This two base encoding system entails the need of decoding the information obtained by the double interrogation. D) Color decoding matrix.

addition of new probes in which certain positions are already known. This module increases accuracy up to 99.99%. A SOLiD System sequencing run takes 7–14 days (depending on the chemistry used), and currently has a maximum output of 300 Gb per run with approximately 5 billion reads. The read length is 35–75 bp with a paired end library (Table 1), and up to 96 barcoded samples can be used [106].

3.4.6. Ion-torrent semiconductor sequencing

Semiconductor sequencing can be considered a variant of pyrosequencing in which the detection entity is hydrogen ions generated after dNTP insertion, rather than light. Detection is performed by a semiconductor chip. The concept has been developed by Ion Torrent Systems Inc. [107], which since 2010 has been part of LifeTechnologies Corporation [91]. The aim of this technology is to enable faster sequencing times by omitting time-consuming imaging steps.

There are currently two Ion Torrent platforms that use this technology: the Personal Genome Machine (PGM) and the Ion Proton. The main differences between these are the size of the sample-deposition surface and the distance between wells, which together determine the sequencing power of the platform. In different platforms there are also different capacities, depending on the size of the chip used. Ion PGM currently has a maximum output of one gigabase of raw data in 2–3 h, generating up to five and a half million reads with an average read length of 200 bp. Ion Proton has announced a maximum raw data output of a hundred gigabases in 2–4 h, generating more than three hundred million reads with an average read length of 200 bp (Table 1) [107]. Therefore, Ion PGM is focused mainly on the targeted resequencing of gene panels, while Ion Proton is geared towards sequencing exomes and whole genomes.

Clonal amplification of the library is performed through an emulsion-PCR with non-paramagnetic beads. The product of the reaction is enriched (beads that have no amplified product attached are removed) and then deposited on a microwell surface, with each microwell able to accommodate only one bead and positioned above a sensor plate sensitive to changes in pH (Fig. 9A).

To perform the sequencing process, first, an unmodified deoxynucleotide (A, T, G, or C) is flooded into the microwells. Its union with a DNA fragment by a polymerase causes the release of a hydrogen ion, which leads to a pH change inside the microwell. This pH alteration is

detected on the sensor plate (ISFET: Ion-Sensitive field-effect transistor [108]). Ion release is proportional to the number of bases added and thus will be the pH change (Fig. 9B).

The main advantages of this sequencing technology lie in the absence of an imaging system which is usually expensive and for which data acquisition is time-consuming in the data acquisition process, and in the ability to use unmodified deoxynucleotides. This, allows the attachment of more than one base in the case of repetitive regions. The signal generated does not increase perfectly linear with the number of bases added. As in the case of pyrosequencing technology, that makes it difficult to sequence correctly homopolymers longer than seven bases [109].

3.4.7. DNA nanoball sequencing: complete genomics

DNA nanoball sequencing was developed by Complete Genomics (recently acquired by the Beijing Genomics Institute [110]) and released commercially in early 2009. It is specifically designed to sequence whole genomes [111]. This technology consists of the construction of large numbers of compact DNA nanoballs for sequencing. The processing of the DNA sample and library construction differs slightly from the other NGS technologies described above.

Initially, the DNA sample is fragmented and selected to a 400–500 bp size. After that, universal adaptors (known as Ad1L and Ad1R) are attached to both ends of the DNA fragments. The sample is then amplified with Ad1-complementary primers and circularised. The product is processed with a restriction enzyme which cuts 13 pb internally from Ad1R, making the product linear again. A second set of universal adaptors (Ad2L and Ad2R) is added to the reaction; these also attach to the ends of the DNA fragments, and the sample is again amplified with Ad2-complementary primers and circularised. The product is again processed with a restriction enzyme, this time cutting 13 pb internally from Ad2L. After this, a third set of adaptors (Ad3L and Ad3R) is added, and PCR and circularisation repeated. This time, the restriction enzyme is EcoP15, which cuts DNA at 26 pb internally to Ad3L, and 26 pb from Ad2R. Thus, a fragment of approximately 350 bp is eliminated (depending on the size of the initial insert). Finally, the last adaptor set is added (Ad4L and Ad4R), a PCR is performed, and the products circularised, but they are not linearised as before. The sample is then amplified by circular replication using DNA Phi29 polymerase [112].

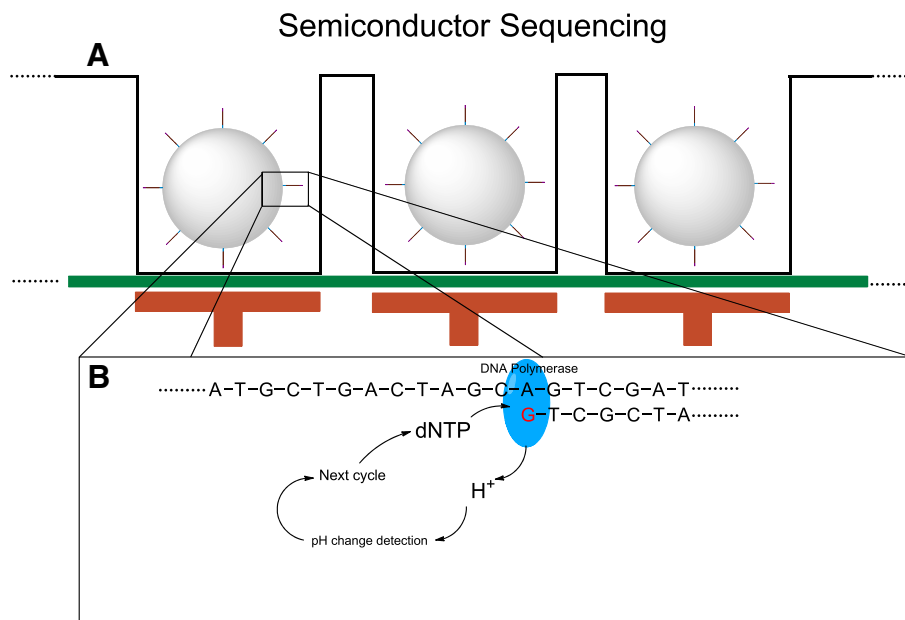


Fig. 9. Semiconductor sequencing. A) Sequencing reactions are performed inside a microwell above a sensor plate. B) First, an unmodified, known dNTP is added to the reaction. It is attached to the 3' end of a sequencing primer by a polymerase. By this, a hydrogen ion is released with each base added. These ions -pH alteration- will be detected by the sensor plate, so no imaging steps are needed. After detection and washing a new deoxynucleotide can be added to the reaction, repeating the process.

The DNA strand that emerges from the circular template is a continuous copy (in the same strand) of that template. Due to that the adaptors added to process, the DNA (Ad1, Ad2, Ad3, and Ad4) possesses palindromic sequences and the new strand will collapse onto itself, forming a DNA 'ball' of approximately 300 nm (nanoball). Nanoballs are unable to bind to each other by base complementarity. For base-reading DNA nanoballs are deposited in an ordered pattern onto a slide (in contrast to other NGS technologies in which this process is performed in a random pattern) [113].

Sequencing itself is performed using 'anchor' oligonucleotides, which are complementary to each end of the Ad1, Ad2, Ad3 or Ad4 adaptors (the technique is therefore also known as 'combinatorial probe-anchor ligation, or cPAL™ technology [114]). Next, a ligase T4 and a pool of 10mer DNA sequences with degenerate nucleotides in all but the interrogative position are added to the reaction. The base interrogated with the 10mer probe is correlated with the fluorophore attached to that probe. A probe will only bind to the DNA if it is complementary in the interrogated position. Non-binding probes are washed away and, fluorescence is detected in an imaging step. After each probe reading and base interrogation, the used anchor and 10mer are removed, and new ones added. The process is repeated until all ten interrogative positions next to the anchor sequence have been recorded. Once that has been achieved, a new anchor that binds to a different adaptor end is added and the process of interrogation repeated. A total of 70 bases can be sequenced using this technology. Due to template processing with Ad1, Ad2, Ad3 and Ad4 adaptors, the bases sequenced are the first and last 35 with respect to the initial DNA size-selected fragment of 400–500 bp.

The main advantages of this technology with respect to other sequencing platforms are: 1) The use of very high-density arrays by nanoball dispersion in an ordered gap-free pattern; and 2) the non-progressive nature of the sequencing. Having been designed specifically to sequence whole genomes, nanoball sequencing also claims to be cheaper than other NGS platforms for that specific application. Disadvantages to consider are the short read length of sequences obtained, which complicate the bioinformatic process of aligning the sequence with the reference genome, and the multiple rounds of PCR necessary [115,116].

The Complete Genomics sequencing platform produces 60 gigabases of usable data from each slide in a 12-day run—therefore to sequence a complete genome with 40× coverage, 120 usable gigabases are needed. 18 slides can be processed at a time, and accuracy is said to be up to 99.999% for 90× coverage (Table 1).

4. NGS bioinformatics

The development of NGS platforms makes huge demands on statistical methods and bioinformatics tools for the analysis and management of the data generated. Indeed, bioinformatics analysis has become one of the main bottlenecks for NGS. As described in previous sections, the data throughput of these platforms has increased by two orders of magnitude compared with older technologies. It is therefore necessary to continue to develop informatics algorithms to translate the results of DNA sequencing into bites of information that can be managed by researchers, physicians, patients, and others [117]. Such algorithms can be organised into several categories: 1) alignment of sequence reads to a reference; 2) base-calling and/or polymorphism detection [118,119]; 3) de novo assembly; and 4) genomic viewers [120–125].

Alignment algorithms have to allow for the fact that the reads obtained with these technologies are short; widely used solutions such as BLAST, BLAT, or CLUSTALW do not therefore constitute the best options. A large number of new algorithms specifically designed to manage short reads have now been developed or adapted from older algorithms [126,127]. These new tools outperform the previously mentioned traditional aligners. In recent years, several dozen

short-read alignments have been published, example Maq [119], Bowtie [128], SSAHA [127], BWA [129], SOAP2 [130]. Short unpaired reads represent a particular challenge for de novo assembly [131,132], with mate-paired reads being the most appropriate for that purpose [133]. Alignment software has to take into account the estimated quality of the underlying data in order to make alignments correctly [119]. Many new algorithms have also been developed for sequencing de novo: AbySS [134], ALLPATHS [131], Edena [133,135], SOAPdenovo [126], Velvet [134]. Challenges still remain in the reconstruction of repetitive regions.

A handful of tools have been developed for base-calling and polymorphism detection, which are important applications of NGS technologies. Some tools filter reads based on the initial alignment quality, such that the conditional likelihood of nucleotides at each position can be computed, following a Bayesian approach and giving the probability of a particular variation to be real. Correct annotation of the variant found is essential to correctly understand its function. Many tools [136–140] and databases [4] are also being used and developed to carry out fast filtering of functionally important variants, with a specific language and ontology [141]. A complete list of algorithms and software is available online [142].

For RNA sequencing, some additional aspects have to be considered, to meet challenges such as transcriptome sequencing, RNA splicing, and the extremely short lengths of reads in small RNAs. In contrast to other approaches, therefore, alignment algorithms have, not only to map reads to the genome, but also to take into account the possible presence of splice junctions and apply algorithms to determine and quantify the most likely splicing isoform [143,144].

One major type of algorithm for RNA sequence alignment comprises unspliced read aligners. These do not allow gaps between exons) and can be divided into two sub-categories: Seed methods (MAQ [119] and Stampy [145]) and Burrows-Wheeler transformation methods (BWA [129] and Bowtie [128]). The other major type of algorithms consists of spliced aligners, which do allow gaps between exons) and can be further divided into: exon-first (MapSplice [146], SpliceMap [147] and TopHat [148]) and seed-extend (GSNAP [149] and QPALMA [150]). For transcriptome reconstruction against the reference genome, the process can be either genome-guided or be genome-independent.

For Chip-Seq, the analytical requirements are distinct from the previously described approaches, because it is important to correctly calculate the confidence of the 'peak' of the binding domain. Common tools for this analysis are the PeakSeq tool [151] and the SPP tool [152].

For data visualisation, a genome browser is required; this is a graphical interface designed to show NGS genomic data (reads) aligned to a reference genome. Such browsers usually allow the acquisition of annotated data from databases (dbSNP, 1000 Genomes, etc.), and thus comparative analysis, prediction of pathogenicity of variants, expression changes, or protein folding. Such tools include Eagleview, Mapview, Maqview, Tablet, and IGV.

It is also important to consider the huge computational power needed to align the millions of RNA reads generated by a standard platform, which necessitate large dedicated centers or cloud computing. Data storage is also an important consideration, with the sequencing cost per base dropping faster than the cost of storing a byte on a hard disc. NGS throughput, in fact, has overcome Moore's Law of Informatics [153].

5. Applications of next-generation sequencing

In recent years, the number of publications on NGS has increased substantially, demonstrating its capabilities in many areas and to many applications. Important applications include: whole genome resequencing [154], targeted resequencing [155], de novo sequencing [156], gene expression analysis with whole transcriptome analysis [157], small RNA sequencing [158], methylation analysis [159], the sequencing of DNA extracted by antibodies targeting DNA-binding

proteins in proteome research (ChIP-Seq) [160], and nuclease fragmentation and sequencing. These applications are currently used to conduct studies in a large number of areas: 1) genetic diagnosis of common and rare diseases and the discovery of causal genes by targeted resequencing [20,53,161–170]; 2) cancer research [22,35,171–175]; 3) microbiological studies [7,156,158]; 4) evolutionary and population studies, including the use of ancient DNA [21,33,36,176–178]; 5) prenatal diagnosis from maternal blood samples [179]; 6) transcriptome studies [157]; 7) personal identification/forensics [8]; 8) metagenomic studies [159,180]; 9) mitochondrial genome studies [177,181]; 10) immunology studies [170]; and 11) plant biology [10].

The ultimate aim of all these applications, as for any gene-mapping technology, is to understand the functional consequences of DNA sequence variation and how that leads to disease. There is the additional difficulty of our limited understanding of non-coding regions of the genome, which previous methods could not address. Now, due to worldwide projects such as ENCODE, this information about the functional organisation of the genome is beginning to be uncovered [182,183].

Due to the affordable cost and high sensibility of NGS, there is currently an increase in the use of exome-sequencing. This is especially useful in family inheritance analysis, to find a causative gene in rare Mendelian disorders. NGS allows the laborious process of searching for a causal mutation via linkage analysis and fine-mapping by Sanger sequencing to be bypassed. This has led to the discovery of causative genes in many diseases in which causal alleles disrupt protein-coding (exonic) sequences [12,184–186].

There is also the need to analyse multiple genes that show overlap in structure and clinical phenotype, without having to sequence the entire exome. Customisable genomic enrichment techniques mentioned in previous sections offer such an approach. There are in fact many commercially available options, such as predefined gene panels, to study diverse types of related diseases or groups of genes (cancer, inherited diseases, the kinome) [20]. It would, however, take several decades to identify the underlying factors of many complex diseases, since they are influenced by many subtle environmental and genetic factors [187].

Such problems are often viewed as an impasse to the ability to perform whole-genome sequencing at an affordable cost. As complex diseases are driven by both coding and non-coding variation, the dataset needed for its study comprises the complete sequence of the genome of an individual and its comparison with databases containing thousands of genomes. In principle, this would make it possible to compare all types of DNA variation (SNP, indels), frequencies, and functions (coding, non-coding). While being very powerful, such a project would also be extremely complex due to the added difficulties inherent in sequencing repetitive or degenerated regions of the genome. Due to the relatively short read-lengths of NGS technologies, such regions are still excluded. Using the complete sequence of the human genome, NGS chemistries can currently make alignments of approximately 85% against the reference sequence (which is increased to 93% if only coding sequences are considered). Longer reads will improve these values [4]. It is also possible to perform whole-genome haplotyping of family genome sequences, using appropriate analysis algorithms [188]. The extra information can be used to increase the power of analysis in disease-association studies, recombination studies, allelic exclusion, and in population genetics (migration studies and evolutionary selection). Finally, such information can facilitate the discovery of modifying genes in Mendelian disease, together with their assignment to a haplotype. The only weakness of this approach is that it is limited to individuals who have access to the genomic sequences of their parents and siblings. There is also the possibility, described by Peters et al. [189], to perform this kind of haplotyping using a very small number of cells (1–20) as the starting material.

The high sensitivity of these techniques is giving new research insights that, with time, will not only increase the resolution of routine diagnostic procedures but reveal new dimensions in diseases

previously classified exclusively by traditional parameters (morphology, pathophysiology, etc.) [169] and lead to the development of new diagnostic methodologies. It is, for example, possible for the first time to track the evolution of a cancer by sequencing it at many developmental stages, by performing micro-biopsies and whole genome sequencing and haplotyping [189], or exome sequencing [190,191]. It is even possible to detect mutations in tumour cells directly from the circulation, or to sequence cell-free fetal nucleic acids (DNAs and RNAs) from maternal blood during pregnancy [13,192]. Several approaches using this latter technique have been developed to determine fetal aneuploidy, by counting the number of reads that map to each chromosome and to calculate the relative dosage of each.

Transcriptome sequencing, on the other hand, is the complete sequencing of all the transcripts in a cell, and their quantification at any specific developmental stage; it is essential for interpreting the functional elements of the genome and for tracking both development and disease. Transcriptome sequencing studies have been used for applications ranging from gene-expression profiling across many cell types and/or conditions, genome annotation, and non-coding RNA discovery and quantification. The most usual application is the sequencing of poly(A)-containing RNA molecules, principally the messenger RNA fraction of the transcriptome [157]. Alternative protocols include strand-specific RNA-seq, and sequencing of poly(A)negative RNA. The analysis of small RNA (transfer RNA, ribosomal RNA, small nuclear, small nucleolar, microRNA and small interfering RNA) is also becoming an important application of these technologies, as they produce appropriate read lengths for it. High-throughput sequencing of small RNAs provides great potential for the identification of novel small RNAs and for profiling those that are already known.

NGS can also be applied to pathogen detection via small RNA sequencing [158], or to metagenomic-based strategies used for the discovery of novel microorganisms and characterisation of microbial communities. Because it is possible to study a complete microbial population, there is no longer a need to isolate and culture individual microbial species, and prior knowledge of their sequences is not necessary [193]. It is interesting also to note the versatility of the data, which can be analysed to provide insight into the level of gene expression, structure of genomic loci and sequence variation.

In mitochondrial disorders, NGS can be used to sequence an entire mitochondrial genome (16.5 kb), determine percentage mutation heteroplasmy, and analyse mitochondrial-related nuclear genes, all in a single run [194].

The association between DNA and proteins is also a fundamental biological phenomenon which plays a key part in the regulation of gene expression and replication, as well as many other processes. In this context, chromatin immunoprecipitation can also be adapted for investigation using NGS. This has been demonstrated by Johnson et al. [195], who have generated a genome-wide, highly precise positional profile of chromatin following degradation of the genomic DNA by nucleases. Two of the main areas of interest for Chip-sequencing are the study of transcription factor binding sites [196,197] and in profiling histone modification [198].

However, in spite of the very wide range of possible applications and uses for NGS technologies, together with the advances being made in this area, their transition into routine clinical diagnosis is still in the early stages of development. Undoubtedly, this is partly due to the enormous computational requirements and the vast amount of data needing analysis.

6. Third Generation Sequencing: an insight into the near future

While in many respects NGS is now in full swing, other technologies, known as 'Third Generation Sequencing technologies' (TGS) are starting to appear. The main advantage of TGS is their ability to sequence single molecules of DNA [199] with no need for clonal amplification prior to

sequencing. This avoids the introduction of artefacts from PCR and requires less manipulation of the sample in comparison with NGS. TGS also usually involve sequencing-by-synthesis chemistries, but detection techniques are based mainly on the physical recognition of DNA bases in an unmodified DNA strand, rather than on detection of chemical incorporation [17]. In these newer technologies, moreover, the sequencing reaction is not paused for ‘wash and scan’ steps after the incorporation of each base, so time and reagent-consumption steps are avoided [200]. TGS also represents a new challenge for bioinformatics; the data obtained with these technologies is of a different kind compared with SGS or Sanger sequencing, due to their ability to detect kinetic activity in the polymerase and/or the influence of secondary structures on sequencing speed.

This type of sequencing is currently being developed by a large number of companies, each using different chemistries, and can be categorised as follows.

6.1. Single-molecule real-time sequencing

Developed by Pacific Biosciences [201] (previously known as Nanofluidics Inc), this technology is based on the detection of natural DNA synthesis by a single DNA polymerase. Specific phosphate-labeled nucleotides (known as phospholinked nucleotides) are added simultaneously, so that the polymerisation occurs in a continuous and processive manner [202]. To understand this approach, two main innovations should be considered: the aforementioned phospholinked nucleotides and the Single-Molecule Real-Time Sequencing chip (SMRT), a nanophotonic visualisation chamber that contains thousands of zero-mode waveguides (ZMWs) [199,203]. These latter are cylindrical holes of approximately 45 nm diameter in a 100 nm thick aluminum film, which together provide a fluorescence detection volume of about 20 zeptoliters. This, confines the enzyme to a tiny volume which

increases the signal-to-noise ratio [204]. When phosphate is cleaved as the nucleotide is incorporated into the DNA strand, base-specific fluorescence is released. After incorporation, the phosphate and dye diffuse quickly out of the tiny detection volume, maintaining the background signal at low levels. As the dye is eliminated when phosphate is cleaved, the result of polymerisation is a single long, natural DNA strand (Fig. 10). The sequencing instrument detects base incorporation as a specific-color flash at a rate of multiple bases per second.

Each SMRT chip has approximately 75,000 ZMWs, each with a DNA polymerase loaded with a different strand of DNA sample. Thus, 75,000 single-molecule-sequencing reactions occur in parallel, with no time-consuming scanning or manual detection steps (signal detection is performed by video recording). Sequencing runs therefore last minutes or hours instead of days [205].

Average read length is above 1000 bp and up to 10 kb, enabling easy de novo assembly. Sample preparation is also easy, consisting of fragmenting DNA, blunting the ends and ligating hairpin adaptors. It is also possible to make independent reads of the same molecule via circular consensus sequencing, which increases the read quality of linearly with the number of reads. However, sequence length decreases to 250 bp in this case [206].

SMRT sequencing instruments allow applications that are not possible with other technologies, such as real-time enzyme activity observation (e.g. of the ribosome as it translates mRNA, RNA-dependent polymerases, etc.) [207]. Also, by observing polymerase activity, kinetic information on nucleotide incorporation can be obtained, which in turn allows data on DNA modification and the structural nature of the strands to be gathered. Thus, the data obtained using TGS are different from those using SGS.

While read accuracy has been reported to be about 83%, this can be improved to more than 99% with 15× coverage. Throughput is also not as high as in SGS [25].

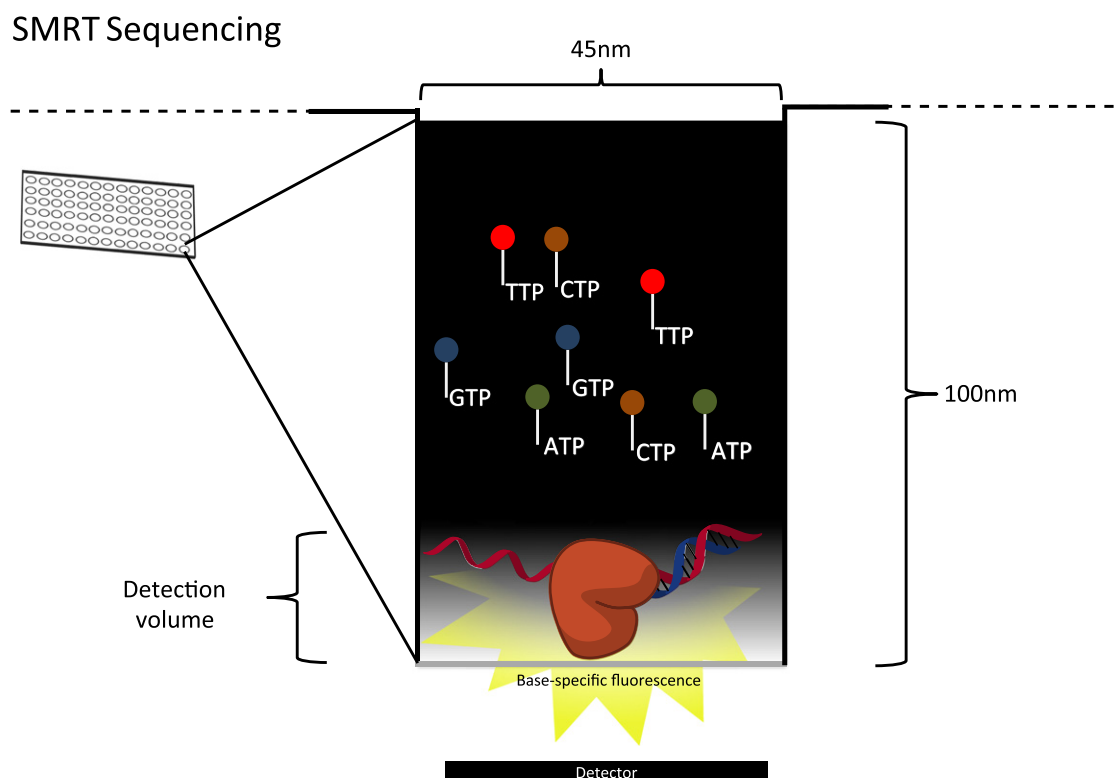


Fig. 10. Single Molecule Real Time (SMRT) Sequencing. A single DNA polymerase enzyme is fixed to the bottom of each zero-mode waveguide (ZMW) of the array. Each of the four DNA bases is attached to a specific fluorescent dye. A single DNA molecule is interrogated on each ZMW. When a nucleotide is incorporated, the fluorescent tag is cleaved and diffuses out of the detection volume, being no longer observable. The detector (optical) registers the fluorescence signal at nucleotide incorporation.

6.2. Real-Time DNA sequencing using fluorescence resonance energy transfer

This chemistry is still in its early developmental stages and few data are available. In this approach, DNA polymerase with an attached fluorophore is brought into close proximity with a phospholinked nucleotide, which is itself tagged with an 'acceptor fluorophore' [208]. The interaction causes emission of a fluorescence resonance energy transfer (FRET) signal, and the fluorophore label on the nucleotide is released.

Due to the high speed of the process, sequencing can be performed at millions of bases per second using this technique [200].

6.3. Direct imaging of DNA using electron microscopy

Halcyon Molecular [209], and ZS Genetics [210] are developing this approach. It allows direct imaging and chemical detection of atoms to identify nucleotides [211]. Its ability to detect atoms in a planar surface by annular dark-field imaging in an aberration-corrected scanning (TEM) has recently been demonstrated as proof-of-concept [212]. Halcyon is developing supporting technologies in parallel, to allow TEM DNA sequencing work, as for example the use of functionalised needles to attach stretched molecules of DNA to a substrate for the imaging.

The main drawbacks of this technology is that a one-cycle PCR step is needed, specifically to label each of the four nucleotides with heavy elements to allow their detection it on the planar surface. Even with one cycle, PCR-based modifications can be introduced. Good sample preparation is critical, and there is a limit on speed due to imaging steps [25].

6.4. Direct imaging of DNA sequences using scanning tunneling microscope tips

There is no proof-of-concept yet published by the main companies involved in the development of this approach (Reveo Inc. [213]). In this technique, DNA is placed on a conductive surface from which bases are detected by Scanning Tunnelling Microscope Tips (STM Tips). Individual bases will probably be detected by tunnelling current measurements. This is expected to become a cheap and rapid technology due to a lack of labeling steps [200].

6.5. DNA sequencing with nanopores

Nanopore sequencing is one of the most advanced third generation sequencing techniques. It consists of the transit of a DNA molecule or its component bases through a pore [214] and its sequence interrogation due to their effect on an electric current or optical signal [215,216]. The system is contained in a synthetic lipid bilayer such that when DNA is loaded onto one face and a voltage applied across the bilayer (by changing the concentration of salt), DNA strands can be mobilised through the pore (Fig. 11) [200]. One of the most important advantages of nanopore sequencing is its inexpensive sample preparation due to the use of unmodified DNA and the employment of a nanopore sensor which eliminates the need for nucleotides and polymerases or ligases. Very long read lengths are also expected.

Many approaches to nanopore sequencing exist;

- 1) Strand-sequencing using ionic current blockage through an alpha hemolysin pore.

Nanopore Sequencing

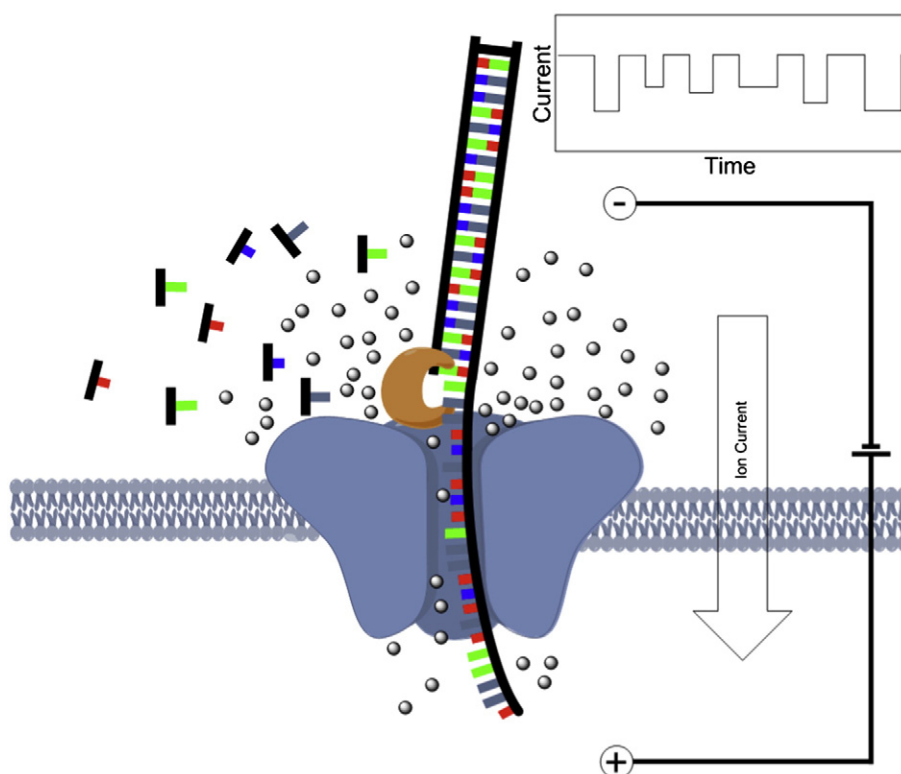


Fig. 11. Nanopore Sequencing. A difference of voltage is applied across a membrane, so an ion current is generated through the nanopore. As the amount of current that pass along the pore is very sensitive to the size and shape of the molecules that move through it, different molecules create characteristic changes in the current.

- 2) Exonuclease sequencing by modulation of the ionic current. The exonuclease attached to the pore cleaves dNMPs from the end of a strand. Base recognition is determined by the level of current blockade of each nucleotide.
- 3) Nanopore sequencing using synthetic DNA and optical readout [217].
- 4) Strand-sequencing using transverse electron currents [218].
- 5) Sequencing using solid-state graphene nanopores. Research is currently being carried by the Cees Dekker Laboratory (Delft, Holland) [219]. This technique has the advantage of the nanopores being narrower than the distance between two bases of DNA (0.3 nm and 0.5 nm respectively), and graphene being mechanically robust and a very good electrical conductor. It has already been shown that single molecules of DNA can be pulled through graphene nanopores and simultaneously detected [220,221].
- 6) Nanopore sequencing-by-synthesis (Nano-SBS). In this approach, a polymerase is attached to the pore, and nucleotides added to the reaction are specifically tagged. As each nucleotide is incorporated into the template DNA, the tag is released, producing a unique ionic blockade (as tags are of different sizes), thus determining the DNA sequence [222].

Oxford Nanopore Technologies has recently introduced its nanopore sequencing platform, GridION, together with a disposable nanopore sequencing device the size of a USB memory stick, called MinION [223].

6.6. Transistor-mediated DNA sequencing

IBM is currently developing a nanostructured sequencing device to electronically detect individual bases in a single molecule of DNA. The structure consists of metallic pores with alternating layers of metal and dielectric materials. As DNA passes through the pores, each base is specifically recognised by its specific interaction with the device [224]. Due to the nature of the technique, it is expected to achieve extremely high throughput (millions of bases per second). As other TGS approaches, this one is claimed to be inexpensive due to its label-free optics-free characteristics [225].

6.7. Conclusions and future perspectives

Revolutionary changes that have occurred in the life sciences coincide temporally with technical advances and major breakthroughs in sequencing technologies, thus demonstrating their utility and value in addressing complex biological questions. The spread and popularisation of NGS due to its reduction in cost-per-base is democratizing access to these technologies, with investigators able to carry out projects at scales previously only accessible in large genome centers. This has been accompanied by an increased interest in sequencing, reflected in the growing number of research groups using the technologies in their research and companies working on future technologies. In this respect, TGS still has much to demonstrate in terms of translating technology into useful data. NGS, therefore, has not only been crucial on the beginning to elucidate genome variability, but continues to be critical in advance the understanding of living systems and complex phenotypes.

None of these advances would be possible without also developing appropriate mathematical algorithms to transform the sequence reads and associated data into meaningful information. In many cases algorithms are still being developed. Despite this, due to current NGS technologies major advances are being possible in many areas, especially, as mentioned before, in the fields related to the understanding and diagnosis of rare and complex diseases [226]. NGS technologies therefore constitute the first opportunity for researchers and clinicians to understand the genetic variability underlying such problems, and provide an unprecedented tool in their investigation,

as reflected in the year-on-year increase in the number of relevant publications.

In conclusion, in this review we have described the current state of sequencing technologies. While it is difficult to predict even the near future of sequencing, it is clear that NGS will become an increasingly routine tool. Moreover, as our knowledge of variability in the human genome increases so too will our functional annotation of the genome. In turn, this will enable sequencing data to be much more readily interpreted by professionals, in terms of potential impact. Currently it is one of the major impediments of these technologies in order to be of routine use.

Conflict of interest

The authors declare that they have no competing interests.

References

- [1] R.F. Service, The race for the \$1000 genome, *Science* 311 (2006) 1544–1546.
- [2] J.C. Venter, M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, et al., The sequence of the human genome, *Science* 291 (2001) 1304–1351.
- [3] E.S. Lander, L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, et al., International human genome, initial sequencing and analysis of the human genome, *Nature* 409 (2001) 860–921.
- [4] R.M. Durbin, D.L. Altshuler, R.M. Durbin, G.R. Abecasis, D.R. Bentley, A. Chakravarti, et al., A map of human genome variation from population-scale sequencing, *Nature* 467 (2010) 1061–1073.
- [5] K. Neveling, R. Collin, C. Gilissen, R. Van Huet, L. Visser, M. Kwint, S. Gijzen, M. Zonneveld, N. Wieskamp, J. De Ligt, A. Siemiatkowska, L. Hoefsloot, M. Buckley, U. Kellner, K. Branham, A. Den Hollander, A. Hoischen, C. Hoyng, B. Klevering, D.B. Van, J. Veltman, F. Cremers, H. Scheffer, Next-generation genetic testing for retinitis pigmentosa, *Hum. Mutat.* 33 (2012) 963–972.
- [6] P. Kapranov, L. Chen, D. Dederich, B. Dong, J. He, K.E. Steinmann, A.R. Moore, J.F. Thompson, P.M. Milos, W. Xiao, Native molecular state of adeno-associated viral vectors revealed by single-molecule sequencing, *Hum. Gene Ther.* 23 (2012) 46–55.
- [7] S. Lee, P.F. Markham, J.F. Markham, I. Petermann, A.H. Noormohammadi, G.F. Browning, N.P. Ficorilli, C.A. Hartley, J.M. Devlin, First complete genome sequence of infectious laryngotracheitis virus, *BMC Genomics* 12 (2011) 197.
- [8] J. Irwin, R. Just, M. Scheible, O. Loreille, Assessing the potential of next generation sequencing technologies for missing persons identification efforts, *Forensic Sci. Int.: Genet. Suppl. Series* 3 (2011) 447–448.
- [9] S. Gravel, B. Henn, R. Gutenkunst, A. Indap, G. Marth, A. Clark, et al., Demographic history and rare allele sharing among human populations, *Proc. Natl. Acad. Sci.* 108 (2011) 11983–11988.
- [10] N.J. Sucher, J.R. Hennell, M.C. Carles, DNA fingerprinting, DNA barcoding, and next generation sequencing technology in plants, *Methods Mol. Biol.* 862 (2011) 1322.
- [11] P. Broadwith, Sequencing in the fast lane, *Chem. World* 9 (2012) 54–58.
- [12] J.C. Roach, G. Glusman, A.F. Smit, C.D. Huff, R. Hubley, P.T. Shannon, L. Rowen, K.P. Pant, N. Goodman, M. Bamshad, J. Shendure, R. Drmanac, L.B. Jorde, L. Hood, D.J. Galas, Analysis of genetic inheritance in a family quartet by whole-genome sequencing, *Science* 328 (2010) 636–639.
- [13] R.W. Chiu, H. Sun, R. Akolekar, C. Clouser, C. Lee, K. McKernan, D. Zhou, K.H. Nicolaides, Y.M. Lo, Maternal plasma DNA analysis with massively parallel sequencing by ligation for noninvasive prenatal diagnosis of trisomy 21, *Clin. Chem.* 56 (2010) 459–463.
- [14] F. Sanger, A.R. Coulson, A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase, *J. Mol. Biol.* 94 (1975) 441–448.
- [15] A.M. Maxam, W. Gilbert, A new method for sequencing DNA, *Proc. Natl. Acad. Sci. U.S.A.* 74 (1977) 560–564.
- [16] F. Sanger, S. Nicklen, A.R. Coulson, DNA sequencing with chain-terminating inhibitors, *Proc. Natl. Acad. Sci. U.S.A.* 74 (1977) 5463–5467.
- [17] E. Pettersson, J. Lundeberg, A. Ahmadian, Generations of sequencing technologies, *Genomics* 93 (2009) 105–111.
- [18] H.Y. Lam, M.J. Clark, R. Chen, R. Chen, G. Natsoulis, M. O'Huallachain, F.E. Dewey, L. Habegger, E.A. Ashley, M.B. Gerstein, A.J. Butte, H.P. Ji, M. Snyder, Performance comparison of whole-genome sequencing platforms, *Nat. Biotechnol.* 30 (2012) 78–82.
- [19] M.J. Clark, R. Chen, H.Y. Lam, K.J. Karczewski, R. Chen, G. Euskirchen, A.J. Butte, M. Snyder, Performance comparison of exome DNA sequencing technologies, *Nat. Biotechnol.* 29 (2011) 908–914.
- [20] I. Audo, K.M. Bujakowska, T. Leveillard, S. Mohand-Said, M. Lancelot, A. Germain, A. Antonio, C. Michiels, J. Saraiva, M. Letexier, J. Sahel, S.S. Bhattacharya, C. Zeit, Development and application of a next-generation-sequencing (NGS) approach to detect known and novel gene defects underlying retinal diseases, *Orphanet J. Rare Dis.* 7 (2012) 8.
- [21] J.G. Gibbons, E.M. Janson, C.T. Hittinger, M. Johnston, P. Abbot, A. Rokas, Benchmarking next-generation transcriptome sequencing for functional and evolutionary genomics, *Mol. Biol. Evol.* 26 (2009) 2731–2744.
- [22] S.M. Tarjei, K. Manching, J.B. David, I. Biju, L. Erez, G. Georgia, A. Pablo, B. William, K. Tae-Kyung, P. Koche, L. William, Eric Mendenhall, Aisling O'Donovan, Aviva Presser, Carsten Russ, Xiaohui Xie, Alexander Meissner,

- Marius Wernig, Rudolf Jaenisch, Chad Nusbaum, Eric S. Lander, Bradley E. Bernstein, Genome-wide maps of chromatin state in pluripotent and lineage-committed cells, *Nature* 448 (2007) 7553.
- [23] M. Choi, U.I. Scholl, W. Ji, T. Liu, I.R. Tikhonova, P. Zumbo, A. Nayir, A. Bakkaloglu, S. Ozen, S. Sanjad, C. Nelson-Williams, A. Farhi, S. Mane, R.P. Lifton, Genetic diagnosis by whole exome capture and massively parallel DNA sequencing, *Proc. Natl. Acad. Sci. U.S.A.* 106 (2009) 19096–19101.
 - [24] E. Mardis, The impact of next-generation sequencing technology on genetics, *Trends Genet.* 24 (2008) 133–141.
 - [25] M. Xu, D. Fujita, N. Hanagata, Perspectives and challenges of emerging single-molecule DNA sequencing technologies, *Small* 5 (2009) 2638–2649.
 - [26] L.M. Smith, S. Fung, M.W. Hunkapiller, T.J. Hunkapiller, L.E. Hood, The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis, *Nucleic Acids Res.* 13 (1985) 2399–2412.
 - [27] M.R. Atkinson, M.P. Deutscher, A. Kornberg, A.F. Russell, J.G. Moffatt, Enzymatic synthesis of deoxyribonucleic acid. XXXIV. Termination of chain growth by a 2',3'-dideoxyribonucleotide, *Biochemistry* 8 (1969) 4897–4904.
 - [28] H. Swerdlow, R. Gesteland, Capillary gel electrophoresis for rapid, high resolution DNA sequencing, *Nucleic Acids Res.* 18 (1990) 1415–1419.
 - [29] L.M. Smith, J.Z. Sanders, R.J. Kaiser, P. Hughes, C. Dodd, C.R. Connell, C. Heiner, S.B. Kent, L.E. Hood, Fluorescence detection in automated DNA sequence analysis, *Nature* 321 (1986) 674–679.
 - [30] Y. Chen, E.E. Roller, X. Huang, DNA sequencing by denaturation: experimental proof of concept with an integrated fluidic device, *Lab Chip* 10 (2010) 1153–1159.
 - [31] C.A. Emrich, H. Tian, I.L. Medintz, R.A. Mathies, Microfabricated 384-lane capillary array electrophoresis bioanalyzer for ultrahigh-throughput genetic analysis, *Anal. Chem.* 74 (2002) 5076–5083.
 - [32] E.R. Mardis, A decade's perspective on DNA sequencing technology, *Nature* 470 (2011) 198–203.
 - [33] S.C. Schuster, W. Miller, A. Ratan, L.P. Tomsho, B. Giardine, L.R. Kasson, R.S. Harris, D.C. Petersen, F. Zhao, J. Qi, C. Alkan, J.M. Kidd, Y. Sun, D.I. Drautz, P. Bouffard, D.M. Muzny, J.G. Reid, L.V. Nazareth, Q. Wang, R. Burhans, C. Riemer, N.E. Wittekindt, P. Moorjani, E.A. Tindall, C.G. Danko, W.S. Teo, A.M. Buboltz, Z. Zhang, Q. Ma, A. Oosthuisen, A.W. Steenkamp, H. Oosthuisen, P. Venter, J. Gajewski, Y. Zhang, B.F. Pugh, K.D. Makova, A. Nekrutenko, E.R. Mardis, N. Patterson, T.H. Pringle, F. Chiaromonte, J.C. Mullikin, E.E. Eichler, R.C. Hardison, R.A. Gibbs, T.T. Harkins, V.M. Hayes, Complete Khoisan and Bantu genomes from southern Africa, *Nature* 463 (2010) 943–947.
 - [34] I. Cho, M.J. Blaser, The human microbiome: at the interface of health and disease, *Nat. Rev. Genet.* 13 (2012) 260–270.
 - [35] The Cancer Genome Atlas Research Network, Integrated genomic analyses of ovarian carcinoma, *Nature* 474 (2011) 609–615.
 - [36] A. Coventry, L. Bull-Ottersen, X. Liu, A.G. Clark, T.J. Maxwell, J. Crosby, J.E. Hixson, T.J. Rea, D.M. Muzny, L.R. Lewis, D.A. Wheeler, A. Sabo, C. Lusk, K.G. Weiss, H. Akbar, A. Cree, A.C. Hawes, I. Newsham, R.T. Varghese, D. Villasana, S. Gross, V. Joshi, J. Santibanez, M. Morgan, K. Chang, W.H. Iv, A.R. Templeton, E. Boerwinkle, R. Gibbs, C.F. Sing, Deep resequencing reveals excess rare recent variants consistent with explosive population growth, *Nat. Commun.* 1 (2010) 131–136.
 - [37] D. MacArthur, C. Tyler-Smith, Loss-of-function variants in the genomes of healthy humans, *Hum. Mol. Genet.* 19 (2010) R125–R130.
 - [38] D.C. Page, V. Cheung, D. Haussler, Human genome: what's been most surprising? *Cell* 147 (2011) 9–10.
 - [39] E. Mardis, Next-generation DNA sequencing methods, *Annu. Rev. Genomics Hum. Genet.* 9 (2008) 387–402.
 - [40] B. Ewing, L. Hillier, M.C. Wendt, P. Green, Base-calling of automated sequencing traces using phred. I. Accuracy assessment, *Genome Res.* 8 (1998) 175–185.
 - [41] K.A. Wetterstrand, DNA sequencing costs: data from the NHGRI large-scale genome sequencing program. Available at www.genome.gov/sequencingcosts, Accessed [2012/08/13].
 - [42] J. Shendure, H. Ji, Next-generation DNA sequencing, *Nat. Biotechnol.* 26 (2008) 1135–1145.
 - [43] S. Acinas, R. Sarma-Rupavtarm, V. Klepac-Ceraj, M. Polz, PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample, *Appl. Environ. Microbiol.* 2005 (2005) 8966–8969.
 - [44] E. Knierim, B. Lucke, J.M. Schwarz, M. Schuelke, D. Seelow, Systematic comparison of three methods for fragmentation of long-range PCR products for next generation sequencing, *PLoS One* 6 (2011) e28240.
 - [45] E. Borgström, S. Lundin, J. Lundeberg, Large scale library generation for high throughput sequencing, *PLoS One* 6 (2011) e19119.
 - [46] M. DeAngelis, D. Wang, T. Hawkins, Solid-phase reversible immobilization for the isolation of PCR products, *Nucleic Acids Res.* 23 (1995) 4742–4743.
 - [47] L. Mamanova, A. Coffey, C. Scott, I. Kozarewa, E. Turner, A. Kumar, E. Howard, J. Shendure, D. Turner, Target-enrichment strategies for next-generation sequencing, *Nat. Methods* 7 (2010) 111–118.
 - [48] J.F. Thompson, P.M. Milos, The properties and applications of single-molecule DNA sequencing, *Genome Biol.* 12 (2011) 217.
 - [49] Y. Asan, H. Xu, C. Jiang, Y. Tyler-Smith, T. Xue, J. Jiang, M. Wang, X. Wu, G. Liu, J. Tian, J. Wang, H. Wang, X. Yang, Zhang, Comprehensive comparison of three commercial human whole-exome capture platforms, *Genome Biol.* 12 (2011) R95.
 - [50] Z. Brownstein, L.M. Friedman, H. Shahin, V. Oron-Karni, N. Kol, A.A. Rayyan, T. Parzefall, D. Lev, S. Shalev, M. Frydman, B. Davidov, M. Shohat, M. Rahile, S. Lieberman, E. Levy-Lahad, M.K. Lee, N. Shomron, M.C. King, T. Walsh, M. Kanaan, K.B. Avraham, Targeted genomic capture and massively parallel sequencing to identify genes for hereditary hearing loss in middle eastern families, *Genome Biol.* 12 (2011) R89.
 - [51] A. Smith, L. Heisler, R. St-Onge, E. Farias-Hesson, I. Wallace, J. Bodeau, A. Harris, K. Perry, G. Gaeveer, N. Pourmand, C. Nislow, Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples, *Nucleic Acids Res.* 38 (2010) e142.
 - [52] J. Kilzer, X. Xun, J. Bodeau, H. Breu, A. Harris, A balanced barcoding system for multiplexed DNA library and SOLiD SAGE "Sequencing", *J. Biomol. Tech.* 21 (2010) 528.
 - [53] A.L. Feldman, A. Dogan, D.I. Smith, M.E. Law, S.M. Ansell, S.H. Johnson, J.C. Porcher, N. Ozsan, E.D. Wieben, B.W. Eckloff, G. Vasmatazis, Massively parallel mate pair DNA library sequencing for translocation discovery: recurrent t (6;7)(p25.3;q32.3), *Translocations In ALK-Negative Anaplastic Large Cell Lymphomas*, *ASH Annual Meeting Abstracts*, 116, 2010, p. 633.
 - [54] F. Mertes, A. Elsharawy, S. Sauer, J. Van Helvoort, D.Z. Van, A. Franke, M. Nilsson, H. Lehrach, A. Brookes, Targeted enrichment of genomic DNA regions for next-generation sequencing, *Brief. Funct. Genomics* 10 (2011) 374–386.
 - [55] R. Saiki, D. Gelfand, S. Stoffel, S. Scharf, R. Higuchi, G. Horn, K. Mullis, H. Erlich, Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase, *Science* 239 (1988) 487–491.
 - [56] www.products.invitrogen.com/ivgn/product/A10498.
 - [57] <http://www.qiagen.com/products/seqtargetsystem.aspx>.
 - [58] <http://www.fluidigm.com>.
 - [59] www.raindancetech.com.
 - [60] R. Williams, S. Peisajovich, O. Miller, S. Magdassi, D. Tawfik, A. Griffiths, Amplification of complex gene libraries by emulsion PCR, *Nat. Methods* 3 (2006) 545–550.
 - [61] R. Tewhey, J. Warner, M. Nakano, B. Libby, M. Medkova, P. David, S. Kotsopoulos, M. Samuels, J. Hutchison, J. Larson, E. Topol, M. Weiner, O. Harismendy, J. Olson, D. Link, K. Frazer, Microdroplet-based PCR enrichment for large-scale targeted sequencing, *Nat. Biotechnol.* 27 (2009) 1025–1031.
 - [62] G. Porreca, K. Zhang, J. Li, B. Xie, D. Austin, S. Vassallo, E. Leproust, B. Peck, C. Emig, F. Dahl, Y. Gao, G. Church, J. Shendure, Multiplex amplification of large sets of human exons, *Nat. Methods* 4 (2007) 931–936.
 - [63] www.invitrogen.com/site/us/en/home/Products-and-Services/Applications/Sequencing/Semiconductor-Sequencing/ampliseq_target_selection.html.
 - [64] www.illumina.com/truseq.ilmn.
 - [65] M. Schena, R.A. Heller, T.P. Theriault, K. Konrad, E. Lachenmeier, R.W. Davis, Microarrays: biotechnology's discovery platform for functional genomics, *Trends Biotechnol.* 17 (1998) 301–306.
 - [66] T. Albert, M. Molla, D. Muzny, L. Nazareth, D. Wheeler, X. Song, T. Richmond, C. Middle, M. Rodesch, C. Packard, G. Weinstock, R. Gibbs, Direct selection of human genomic loci by microarray hybridization, *Nat. Methods* 4 (2007) 903–905.
 - [67] D. Okou, K. Steinberg, C. Middle, D. Cutler, T. Albert, M. Zwick, Microarray-based genomic selection for high-throughput resequencing, *Nat. Methods* 4 (2007) 907–909.
 - [68] E. Hodges, Z. Xuan, V. Balija, M. Kramer, M. Molla, S. Smith, C. Middle, M. Rodesch, T. Albert, G. Hannon, W. McCombie, Genome-wide in situ exon capture for selective resequencing, *Nat. Genet.* 39 (2007) 1522–1527.
 - [69] M.N. Bainbridge, M. Wang, D.L. Burgess, C. Kovar, M. Rodesch, M. D'Ascenzo, J. Kitzman, Y.Q. Wu, I. Newsham, T. Richmond, J.A. Jeddleloh, D. Muzny, T. Albert, R. Gibbs, Whole exome capture in solution with 3 Gbp of data, *Genome Biol.* 11 (2010) R62.
 - [70] www.genomics.agilent.com/CollectionOverview.aspx?PageType=Application&SubPageType=ApplicationOverview&PageID=2094.
 - [71] www.nimblegen.com/products/seqcap/ez/choice/index.html.
 - [72] www.flexgen.nl.
 - [73] www.microarray.com/myreads/myreads.html#Capture.
 - [74] www.invitrogen.com/site/us/en/home/Products-and-Services/Applications/Sequencing/Next-Generation-Sequencing/Sample-Prep-Library-Generation-for-Next-Generation-Sequencing/next-generation-sequencing-reagents-by-workflow/Sample-Enrichment-Reagents/TargetSeq-Exome-Enrichment-System.html.
 - [75] F. Dahl, M. Gullberg, J. Stenberg, U. Landegren, M. Nilsson, Multiplex amplification enabled by selective circularization of large sets of genomic DNA fragments, *Nucleic Acids Res.* 33 (2005) e71.
 - [76] H. Johansson, M. Isaksson, E. Sorqvist, F. Roos, J. Stenberg, T. Sjöblom, J. Botling, P. Micke, K. Edlund, S. Fredriksson, H. Kulitima, O. Ericsson, M. Nilsson, Targeted resequencing of candidate genes using selector probes, *Nucleic Acids Res.* 39 (2011) e8.
 - [77] H. Emily, L. Choli, B. Sarah, A. Deborah, J. Shendure, Massively parallel exon capture and library-free resequencing across 16 individuals, *Nat. Methods* 16 (2009) 315–316.
 - [78] www.halogenomics.com.
 - [79] www.illumina.com/technology/solexa_technology.ilmn.
 - [80] C. Adessi, G. Matton, G. Ayala, G. Turcatti, J. Mermod, P. Mayer, E. Kawashima, Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms, *Nucleic Acids Res.* 28 (2000) E87.
 - [81] www.illumina.com.
 - [82] D. Bing, C. Boles, F. Rehman, Bridge amplification: a solid phase PCR system for the amplification and detection of allelic differences in single copy genes, *Genetic Identity Conference Proceedings, Seventh International Symposium on Human Identification*, 1996.
 - [83] M. Margulies, M. Egholm, W. Altman, S. Attiya, J. Bader, L. Bemben, J. Berka, M. Braverman, Y. Chen, Z. Chen, S. Dewell, L. Du, J. Fierro, X. Gomes, B. Godwin, W. He, S. Helgesen, C.H. Ho, C. Ho, G. Irzyk, S. Jando, M.L. Alenquer, T. Jarvie, K. Jirage, J. Kim, J. Knight, J. Lanza, J. Leamon, S. Lefkowitz, M. Lei, J. Li, K. Lohman, H. Lu, V. Makhijani, K. McDade, M. McKenna, E. Myers, E. Nickerson, J. Nobile, R. Plant, B. Puc, M. Ronan, G. Roth, G. Sarkis, J. Simons, J. Simpson, M. Srinivasan, K. Tartaro, A. Tomasz, K. Vogt, G. Volkmer, S. Wang, Y. Wang, M. Weiner, P. Yu, R.

- Begley, J. Rothberg, Genome sequencing in microfabricated high-density picolitre reactors, *Nature* 437 (2005) 376–380.
- [84] www.appliedbiosystems.com.
- [85] J. Shendure, G. Porreca, N.B. Reppas, X. Lin, J.P. McCutcheon, A.M. Rosenbaum, M.D. Wang, K. Zhang, R.D. Mitra, G. Church, Accurate multiplex polony sequencing of an evolved bacterial genome, *Science* 309 (2005) 1728–1732.
- [86] D. Dressman, H. Yan, G. Traverso, K.W. Kinzler, B. Vogelstein, Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations, *Proc. Natl. Acad. Sci. USA* 100 (2003) 8817–8822.
- [87] M. Nakano, J. Komatsu, S. Matsuura, K. Takashima, S. Katsura, A. Mizuno, Single-molecule PCR using water-in-oil emulsion, *J. Biotechnol.* (2003) 117–124.
- [88] K. Shao, W. Ding, F. Wang, H. Li, D. Ma, H. Wang, Emulsion PCR: a high efficient way of PCR amplification of random DNA libraries in aptamer selection, *PLoS One* 6 (2011) e24910.
- [89] www.454.com.
- [90] www.polonator.org.
- [91] www.lifetechnologies.com.
- [92] www.kth.se/en.
- [93] M. Ronaghi, M. Uhlen, P. Nyrén, A sequencing method based on real-time pyrophosphate, *Science* 281 (1998) 363–365.
- [94] www.454.com/products/gx-flx-system/index.asp.
- [95] J. Ju, D.H. Kim, L. Bi, Q. Meng, X. Bai, Z. Li, X. Li, M.S. Marma, S. Shi, J. Wu, J.R. Edwards, A. Romu, N.J. Turro, Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators, *Proc. Natl. Acad. Sci.* 103 (2006) 19635–19640.
- [96] www.intelligentbiosystems.com/index-1%20mod%201.html#.
- [97] C.W. Fuller, L.R. Middendorf, S.A. Benner, G. Church, T. Harris, X. Huang, S.B. Jovanovich, J.R. Nelson, J.A. Schloss, D.C. Schwartz, D.V. Vezenov, The challenges of sequencing by synthesis, *Nat. Biotechnol.* 27 (2009) 1013–1023.
- [98] www.illumina.com/documents/%5Cproducts%5Ctechnote%5Ctechnote_Q-Scores.pdf.
- [99] www.hms.harvard.edu.
- [100] R. Mitra, G.M. Church, In situ localized amplification and contact replication of many individual DNA molecules, *Nucleic Acids Res.* 27 (1999) e34–e39.
- [101] R. Mitra, Polony sequencing: a DNA sequencing technology. Ph.D. Thesis Institute of Technology, Massachusetts, 2000.
- [102] R.D. Mitra, J. Shendure, J. Olejnik, Edyta-Krzyszanska-Olejnik, G. Church, Fluorescent in situ sequencing on polymerase colonies, *Anal. Biochem.* 320 (2003) 55–65.
- [103] V. Pandey, R.C. Nutter, E. Prediger, Applied Biosystems SOLiD™ System: ligation-based sequencing, *Next Generation Genome Sequencing: Towards Personalized Medicine*, 2008, pp. 29–41. (Book).
- [104] S. Brenner, M. Johnson, J. Bridgman, G. Golda, D.H. Lloyd, D. Johnson, S. Luo, S. McCurdy, M. Foy, M. Ewan, R. Roth, D. George, S. Eletr, G. Albrecht, E. Vermaas, S.R. Williams, K. Moon, T. Burcham, M. Pallas, R.B. DuBridge, J. Kirchner, K. Fearon, J. Mao, K. Corcoran, Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays, *Nat. Biotechnol.* 18 (2000) 630–634.
- [105] www.solid.appliedbiosystems.com.
- [106] www.appliedbiosystems.com/absite/us/en/home/applications-technologies/solid-next-generation-sequencing/next-generation-systems.html.
- [107] www.iontorrent.com.
- [108] P. Bergveld, ISFET, theory and practice, IEEE Sensor Conference, Toronto, 2003, p. 26.
- [109] J. Rothberg, W. Hinz, T.M. Rearick, J. Schultz, W. Mileski, M. Davey, J.H. Leamon, K. Johnson, M.J. Milgrew, M. Edwards, J. Hoon, J.F. Simons, D. Marran, J.W. Myers, J.F. Davidson, A. Branting, J. Nobile, B. Puc, D. Light, T.A. Clark, M. Huber, J.T. Branciforte, I.B. Stoner, S.E. Cawley, M. Lyons, Y. Fu, N. Homer, M. Sedova, X. Miao, B. Reed, J. Sabina, E. Feisterstein, M. Schorn, M. Alanjary, E. Dimalanta, D. Dressman, R. Kasinskas, T. Sokolsky, J.A. Fidanza, E. Namsaraev, K.J. McKernan, A. Williams, G.T. Roth, J. Bustillo, An integrated semiconductor device enabling non-optical genome sequencing, *Nature* 475 (2011) 348–352.
- [110] www.genomics.cn.
- [111] www.completegenomics.com.
- [112] L. Blanco, A. Bernad, J.M. Lázaro, G. Martín, C. Garmendia, M. Salas, Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication, *J. Biol. Chem.* 264 (1989) 8935–8940.
- [113] L. Chrisey, G. Lee, C. O'Ferrall, Covalent attachment of synthetic DNA to self-assembled monolayer films, *Nucleic Acids Res.* 24 (1996) 3031–3039.
- [114] www.completegenomics.com/services/technology/.
- [115] R. Drmanac, A. Sparks, M. Callow, A. Halpern, N. Burns, B. Kermani, P. Carnevali, I. Nazarenko, G. Nilsen, G. Yeung, F. Dahl, A. Fernandez, B. Staker, K. Pant, J. Baccash, A. Borcharding, A. Brownlee, R. Cedeno, L. Chen, D. Chernikoff, A. Cheung, R. Chirita, B. Curson, J. Ebert, C. Hacker, R. Hartlage, B. Hauser, S. Huang, Y. Jiang, V. Karpinchyk, M. Koenig, C. Kong, T. Landers, C. Le, J. Liu, C. McBride, M. Morensoni, R. Morey, K. Mutch, H. Perazich, K. Perry, B. Peters, J. Peterson, C. Pethiyagoda, K. Pothuraju, C. Richter, A. Rosenbaum, S. Roy, J. Shafro, U. Sharanhovich, K. Shannon, C. Sheppy, M. Sun, J. Thakuria, A. Tran, D. Vu, A. Zaranek, X. Wu, S. Drmanac, A. Oliphant, W. Banyai, B. Martin, D. Ballinger, G. Church, C. Reid, Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays, *Science* 327 (2010) 78–81.
- [116] G. Porreca, Genome sequencing on nanoballs, *Nat. Biotechnol.* 28 (2010) 43–44.
- [117] E.E. Schadt, M.D. Linderman, J. Sorenson, L. Lee, G.P. Nolan, Computational solutions to large-scale data management and analysis, *Nat. Rev. Genet.* 11 (2010) 647–657.
- [118] A.R. Quinlan, D.A. Stewart, M.P. Stromberg, G.T. Marth, PyroBayes: an improved base caller for SNP discovery in pyrosequences, *Nat. Methods* 5 (2008) 179–181.
- [119] H. Li, J. Ruan, R. Durbin, Mapping short DNA sequencing reads and calling variants using mapping quality scores, *Genome Res.* 18 (2008) 1851–1858.
- [120] www.broadinstitute.org/igv.
- [121] www.bamview.sourceforge.net.
- [122] www.evolution.sysu.edu.cn/mapview.
- [123] H. Bao, H. Guo, J. Wang, R. Zhou, X. Lu, S. Shi, MapView: visualization of short reads alignment on a desktop computer, *Bioinformatics* 25 (2009) 1554–1555.
- [124] www.maq.sourceforge.net/maqview.shtml.
- [125] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome Project Data Processing Subgroup, The sequence alignment/Map format and SAMtools, *Bioinformatics* 25 (2009) 2078–2079.
- [126] R. Li, Y. Li, K. Kristiansen, J. Wang, SOAP: short oligonucleotide alignment program, *Bioinformatics* 24 (2008) 713–714.
- [127] Z. Ning, A. Cox, J. Mullikin, SSAHA: a fast search method for large DNA databases, *Genome Res.* 11 (2001) 1725.
- [128] B. Langmead, C. Trapnell, M. Pop, S.L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, *Genome Biol.* 10 (2009), (R25–2009–10–3–r25. Epub 2009 Mar 4).
- [129] H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform, *Bioinformatics* 26 (2010) 589–595.
- [130] R. Li, C. Yu, Y. Li, T. Lam, S. Yiu, K. Kristiansen, J. Wang, SOAP2: an improved ultrafast tool for short read alignment, *Bioinformatics* 25 (2009) 1966–1967.
- [131] J. Butler, I. MacCallum, M. Kleber, I. Shlyakhter, M. Belmonte, E. Lander, C. Nusbaum, D. Jaffe, ALLPATHS: De novo assembly of whole-genome shotgun microreads, *Genome Res.* 18 (2008) 810.
- [132] A. Sundquist, M. Ronaghi, H. Tang, P. Pevzner, S. Batzoglou, Whole-genome sequencing and assembly with high-throughput, short-read technologies, *PLoS One* 2 (2007) e484.
- [133] R.Z. Daniel, Velvet: algorithms for de novo short read assembly using de Bruijn graphs, *Genome Res.* 18 (2008) 821.
- [134] J.T. Simpson, K. Wong, S.D. Jackman, J.E. Schein, S.J. Jones, I. Birol, ABySS: a parallel assembler for short read sequence data, *Genome Res.* 19 (2009) 1117–1123.
- [135] D. Hernandez, P. Francois, L. Farinelli, M. Osteras, J. Schrenzel, De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer, *Genome Res.* 18 (2008) 802–809.
- [136] www.ensembl.org/info/docs/variation/vep/index.html.
- [137] www.sanger.ac.uk/resources/software/vagrent/.
- [138] www.pupasuite.bioinfo.cipf.es/.
- [139] www.genetics.bwh.harvard.edu/pph2/.
- [140] www.mutationassessor.org.
- [141] www.sequenceontology.org.
- [142] www.seqanswers.com/forums/showthread.php?t=43.
- [143] C. Trapnell, B.A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M.J. van Baren, S.L. Salzberg, B.J. Wold, L. Pachter, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation, *Nat. Biotechnol.* 28 (2010) 511–515.
- [144] M. Guttman, M. Garber, J.Z. Levin, J. Donaghey, J. Robinson, X. Adiconis, L. Fan, M.J. Koziol, A. Gnirke, C. Nusbaum, J.L. Rinn, E.S. Lander, A. Regev, Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs, *Nat. Biotechnol.* 28 (2010) 503–510.
- [145] G. Lunter, M. Goodson, Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads, *Genome Res.* 21 (2011) 936–939.
- [146] K. Wang, D. Singh, Z. Zeng, S.J. Coleman, Y. Huang, G.L. Savich, X. He, P. Mieczkowski, S.A. Grimm, C.M. Perou, J.N. MacLeod, D.Y. Chiang, J.F. Prins, J. Liu, MapSplice: accurate mapping of RNA-seq reads for splice junction discovery, *Nucleic Acids Res.* 38 (2010) e178.
- [147] K.F. Au, H. Jiang, L. Lin, Y. Xing, W.H. Wong, Detection of splice junctions from paired-end RNA-seq data by SpliceMap, *Nucleic Acids Res.* 38 (2010) 4570–4578.
- [148] C. Trapnell, L. Pachter, S.L. Salzberg, TopHat: discovering splice junctions with RNA-Seq, *Bioinformatics* 25 (2009) 1105–1111.
- [149] T.D. Wu, S. Nacu, Fast and SNP-tolerant detection of complex variants and splicing in short reads, *Bioinformatics* 26 (2010) 873–881.
- [150] F. De Bona, S. Ossowski, K. Schneeberger, G. Ratsch, Optimal spliced alignments of short sequence reads, *Bioinformatics* 24 (2008) i174–i180.
- [151] J. Rozowsky, G. Euskirchen, R.K. Auerbach, Z.D. Zhang, T. Gibson, R. Bjornson, N. Carrier, M. Snyder, M.B. Gerstein, PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls, *Nat. Biotechnol.* 27 (2009) 66–75.
- [152] P.V. Kharchenko, M.Y. Tolstourov, P.J. Park, Design and analysis of ChIP-seq experiments for DNA-binding proteins, *Nat. Biotechnol.* 26 (2008) 1351–1359.
- [153] M. Baker, Next-generation sequencing: adjusting to data overload, *Nat. Methods* 7 (2010) 495–499.
- [154] S. Suzuki, N. Ono, C. Furusawa, B. Ying, T. Yomo, Comparison of sequence reads obtained from three next-generation sequencing platforms, *PLoS One* 6 (2011) e19534.
- [155] D. Hedges, T. Guetouche, S. Yang, G. Bademci, A. Diaz, A. Andersen, W. Hulme, S. Linker, A. Mehta, Y. Edwards, G. Beecham, E. Martin, M. Pericak-Vance, S. Zuchner, J. Vance, J. Gilbert, Comparison of three targeted enrichment strategies on the SOLiD sequencing platform, *PLoS One* 6 (2011) e18595.
- [156] W. Ghosh, A. George, A. Agarwal, P. Raj, M. Alam, P. Pyne, S. Das Gupta, Whole-genome shotgun sequencing of the sulfur-oxidizing chemolithotroph *tetrathiodibacter kashmirensis*, *J. Bacteriol.* 193 (2011) 5553–5554.
- [157] M.D. Alter, D.B. Rubin, K. Ramsey, R. Halpern, D.A. Stephan, L. Abbott, R. Hen, Variation in the large-scale organization of gene expression levels in the hippocampus relates to stable epigenetic variability in behavior, *PLoS One* 3 (2008) e3344.
- [158] N.C.T. Schopman, M. Willemsen, Y.P. Liu, T. Bradley, A. van Kampen, F. Baas, B. Berkhout, J. Haasnoot, Deep sequencing of virus-infected cells reveals HIV-encoded small RNAs, *Nucleic Acids Res.* 40 (2012) 414–427.

- [159] M. Zeschnigk, M. Martin, G. Betzl, A. Kalbe, C. Sirsch, K. Buiting, S. Gross, E. Fritzilas, B. Frey, S. Rahmann, B. Horsthemke, Massive parallel bisulfite sequencing of CG-rich DNA fragments reveals that methylation of many X-chromosomal CpG islands in female blood DNA is incomplete, *Hum. Mol. Genet.* 18 (2009) 1439–1448.
- [160] M. Mokry, P. Hatzis, E. de Bruijn, J. Koster, R. Versteeg, J. Schuijers, M. van de Watering, V. Gryev, H. Clevers, E. Cuppen, Efficient double fragmentation ChIP-seq provides nucleotide resolution protein-DNA binding profiles, *PLoS One* 5 (2010) e15092.
- [161] J. Thompson, J. Reifengerger, E. Giladi, K. Kerouac, J. Gill, E. Hansen, A. Kahvejian, P. Kapranov, T. Knope, D. Lipson, K. Steinmann, P. Milos, Single-step capture and sequencing of natural DNA for detection of BRCA1 mutations, *Genome Res.* 22 (2012) 340–345.
- [162] A. Elliott, J. Radecki, B. Moghis, X. Li, A. Kammesheid, Rapid detection of the ACMG/ACOG-recommended 23 CFTR disease-causing mutations using ion torrent semiconductor sequencing, *J. Biomol. Tech.* 23 (2012) 24–30.
- [163] S. Ghosh, F. Krux, V. Binder, M. Gombert, T. Niehues, O. Feyen, H.J. Laws, A. Borkhardt, on behalf of PID-NET: German Network on Primary Immunodeficiency Diseases, Array-based sequence capture and next-generation sequencing for the identification of primary immunodeficiencies, *Scand. J. Immunol.* 75 (2012) 350–354.
- [164] C. Guelly, P. Zhu, L. Leonardi, L. Papić, J. Zidar, M. Schabhtütt, H. Strohmaier, J. Weis, T.M. Strom, J. Baets, J. Willems, P.D. Jonghe, M.M. Reilly, E. Fröhlich, M. Hatz, S. Trajanoski, T.R. Pieber, A.R. Janeczek, C. Blackstone, M. Auer-Grumbach, Targeted high-throughput sequencing identifies mutations in atlastin-1 as a cause of hereditary sensory neuropathy type I, *Am. J. Hum. Genet.* 88 (2011) 99–105.
- [165] Q. Jiang, T. Turner, M. Sosa, A. Rakha, S. Arnold, A. Chakravarti, Rapid and efficient human mutation detection using a bench-top next-generation DNA sequencer, *Hum. Mutat.* 33 (2011) 281–289.
- [166] I.N. Lai, V.S. Yellore, S.A. Rayner, N.D. Silva, C.K. Nguyen, A.J. Aldave, The utility of next-generation sequencing in the evaluation of the posterior polymorphous corneal dystrophy 1 locus, *Mol. Vis.* 16 (2010) 2829–2838.
- [167] J. O'sullivan, C. Bitu, S. Daly, J. Urquhart, M. Barron, S. Bhaskar, H. Martelli-Junior, P. Dos SantosNeto, M. Mansilla, J. Murray, R. Coletta, G. Black, M. Dixon, Whole-exome sequencing identifies FAM20A mutations as a cause of amelogenesis imperfecta and gingival hyperplasia syndrome, *Am. J. Hum. Genet.* 88 (2012) 616–620.
- [168] N. Schlipf, R. Schüle, S. Klimpe, K. Karle, M. Synofzik, J. Schicks, O. Riess, L. Schöls, P. Bauer, Amplicon-based high-throughput pooled sequencing identifies mutations in CYP7B1 and SPG7 in sporadic spastic paraplegia patients, *Clin. Genet.* 80 (2011) 148–160.
- [169] M.M.H. Yang, A. Singhal, S.R. Rassekh, S. Yip, P. Eydoux, C. Dunham, Possible differentiation of cerebral glioblastoma into pleomorphic xanthoastrocytoma: an unusual case in an infant, *J. Neurosurg. Pediatr.* 9 (2012) 517–523.
- [170] C. Lind, D. Ferriola, K. Mackiewicz, S. Heron, M. Rogers, L. Slavich, R. Walker, T. Hsiao, L. McLaughlin, M. D'Arcy, X. Gai, D. Goodridge, D. Sayer, D. Monos, Next-generation sequencing: the solution for high-resolution, unambiguous human leukocyte antigen typing, *Hum. Immunol.* 71 (2010) 1033–1042.
- [171] M. Gerlinger, A.J. Rowan, S. Horswell, J. Larkin, D. Endesfelder, E. Gronroos, P. Martinez, N. Matthews, A. Stewart, P. Tarpey, I. Varela, B. Phillimore, S. Begum, N.Q. McDonald, A. Butler, D. Jones, K. Raine, C. Latimer, C.R. Santos, M. Nohadani, A.C. Eklund, B. Spencer-Dene, G. Clark, L. Pickering, G. Stamp, M. Gore, Z. Szallasi, J. Downward, P.A. Futreal, C. Swanton, Intratumor heterogeneity and branched evolution revealed by multiregion sequencing, *N. Engl. J. Med.* 366 (2012) 883–892.
- [172] W. Lee, Z. Jiang, J. Liu, P.M. Haverty, Y. Guan, J. Stinson, P. Yue, Y. Zhang, K.P. Pant, D. Bhatt, C. Ha, S. Johnson, M.I. Kennemer, S. Mohan, I. Nazarenko, C. Watanabe, A.B. Sparks, D.S. Shames, R. Gentleman, F.J. de Sauvage, H. Stern, A. Pandita, D.G. Ballinger, R. Drmanac, Z. Modrusan, S. Seshagiri, Z. Zhang, The mutation spectrum revealed by paired genome sequences from a lung cancer patient, *Nature* 465 (2010) 473–477.
- [173] B. Timmermann, M. Kerick, C. Roehr, A. Fischer, M. Isau, S. Boerno, A. Wunderlich, C. Barmeyer, P. Seemann, J. Koenig, M. Lappe, A. Kuss, M. Garshasbi, L. Bertram, K. Trappe, M. Werber, B. Herrmann, K. Zatlouk, H. Lehrach, M. Schweiger, Somatic mutation profiles of MSI and MSS colorectal cancer identified by whole exome next generation sequencing and bioinformatics analysis, *PLoS One* 5 (2010) e15661.
- [174] M.J. Walter, D. Shen, L. Ding, J. Shao, D.C. Koboldt, K. Chen, D.E. Larson, M.D. McLellan, D. Dooling, R. Abbott, R. Fulton, V. Magrini, H. Schmidt, J. Kalicki-Veizer, M. O'Loughlin, X. Fan, M. Grillo, S. Witowski, S. Heath, J.L. Frater, W. Eades, M. Tomasson, P. Westervelt, J.F. DiPersio, D.C. Link, E.R. Mardis, T.J. Ley, R.K. Wilson, T.A. Graubert, Clonal architecture of secondary acute myeloid leukemia, *N. Engl. J. Med.* 366 (2012) 1090–1098.
- [175] P.J. Campbell, E.D. Pleasance, P.J. Stephens, E. Dicks, R. Rance, I. Goodhead, G.A. Follows, A.R. Green, P.A. Futreal, M.R. Stratton, Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing, *Proc. Natl. Acad. Sci. U.S.A.* 105 (2008) 13081–13086.
- [176] G.J. Adcock, E.S. Dennis, S. Easteal, G.A. Huttley, L.S. Jermini, W.J. Peacock, A. Thorne, Mitochondrial DNA sequences in ancient Australians: implications for modern human origins, *Proc. Natl. Acad. Sci. U.S.A.* 98 (2001) 537–542.
- [177] R. Green, A. Malaspina, J. Krause, A. Briggs, P. Johnson, C. Uhler, M. Meyer, J. Good, T. Maricic, U. Stenzel, K. Prüfer, M. Siebauer, H. Burbano, M. Ronan, J. Rothberg, M. Egholm, P. Rudan, D. Brajkovic, Z. Kucan, I. Gusic, M. Wikström, L. Laakkonen, J. Kelso, M. Slatkin, S. Pääbo, A complete neandertal mitochondrial genome sequence determined by high-throughput sequencing, *Cell* 134 (2008) 416–426.
- [178] J. Korbel, A. Urban, J. Affourtit, B. Godwin, F. Grubert, J. Simons, P. Kim, D. Palejev, N. Carriero, L. Du, B. Taillon, Z. Chen, A. Tanzer, A. Saunders, J. Chi, F. Yang, N. Carter, M. Hurles, S. Weissman, T. Harkins, M. Gerstein, M. Egholm, M. Snyder, Paired-end mapping reveals extensive structural variation in the human genome, *Science* 318 (2007) 420–426.
- [179] H.C. Fan, Y.J. Blumenfeld, U. Chitkara, L. Hudgins, S.R. Quake, Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood, *Proc. Natl. Acad. Sci. U.S.A.* 105 (2008) 16266–16271.
- [180] H.L. Tyler, L.F. Roesch, S. Gowda, W.O. Dawson, E.W. Triplett, Confirmation of the sequence of '*Candidatus* *liberibacter asiaticus*' and assessment of microbial diversity in Huanglongbing-INFECTED CITRUS PHLOEM USING A METAGENOMIC APPROACH, *Mol. Plant Microbe Interact.* 22 (2009) 1624–1643.
- [181] M. Holland, M. McQuillan, K.O. Hanlon, Second generation sequencing allows for mtDNA mixture deconvolution and high resolution detection of heteroplasmy, *Croat. Med. J.* 52 (2011) 299–313.
- [182] Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project, *Nature* 447 (2007) 799–816.
- [183] An integrated encyclopedia of DNA elements in the human genome, *Nature* 489 (2012) 57–74.
- [184] S.B. Ng, K.J. Buckingham, C. Lee, A.W. Bigham, H.K. Tabor, K.M. Dent, C.D. Huff, P.T. Shannan, E.W. Jabs, D.A. Nickerson, J. Shendure, M.J. Bamshad, Exome sequencing identifies the cause of a mendelian disorder, *Nature Publishing Group* 42 (2010) 30–35.
- [185] C. Rodelsperger, P. Krawitz, S. Bauer, J. Hecht, A.W. Bigham, M. Bamshad, B.J. de Condor, M.R. Schweiger, P.N. Robinson, Identity-by-descent filtering of exome sequence data for disease-gene identification in autosomal recessive disorders, *Bioinformatics* 27 (2011) 829–836.
- [186] B. Rabbani, N. Mahdiah, K. Hosomichi, H. Nakaoka, I. Inoue, Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders, *J. Hum. Genet.* 57 (2012) 621–632.
- [187] H. Kilpinen, J.C. Barrett, How next-generation sequencing is transforming complex disease genetics, *Trends Genet.* 29 (2013) 23–30.
- [188] J. Roach, G. Glusman, R. Hubley, S. Montsaroff, A. Holloway, D. Mauldin, D. Srivastava, V. Garg, K. Pollard, D. Galas, L. Hood, A.A. Smit, Chromosomal haplotypes by genetic phasing of human families, *Am. J. Hum. Genet.* 89 (2011) 382–397.
- [189] B.A. Peters, B.G. Kermani, A.B. Sparks, O. Alferov, P. Hong, A. Alexeev, Y. Jiang, F. Dahl, Y.T. Tang, J. Haas, K. Robasky, A.W. Zaranek, J. Lee, M.P. Ball, J.E. Peterson, H. Perazich, G. Yeung, J. Liu, L. Chen, M.I. Kennemer, K. Pothuraju, K. Konvicka, M. Tsoupo-Sitnikov, K.P. Pant, J.C. Ebert, G.B. Nilsen, J. Baccash, A.L. Halpern, G.M. Church, R. Drmanac, Accurate whole-genome sequencing and haplotyping from 10 to 20 human cells, *Nature* 487 (2012) 190–195.
- [190] Y. Hou, L. Song, P. Zhu, B. Zhang, Y. Tao, X. Xu, F. Li, K. Wu, J. Liang, D. Shao, H. Wu, X. Ye, C. Ye, R. Wu, M. Jian, Y. Chen, W. Xie, R. Zhang, L. Chen, X. Liu, X. Yao, H. Zheng, C. Yu, Q. Li, Z. Gong, M. Mao, X. Yang, L. Yang, J. Li, W. Wang, Z. Lu, N. Gu, G. Laurie, L. Bolund, K. Kristiansen, J. Wang, H. Yang, Y. Li, X. Zhang, J. Wang, Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm, *Cell* 148 (2012) 873–885.
- [191] X. Xu, Y. Hou, X. Yin, L. Bao, A. Tang, L. Song, F. Li, S. Tsang, K. Wu, H. Wu, W. He, L. Zeng, M. Xing, R. Wu, H. Jiang, X. Liu, D. Cao, G. Guo, X. Hu, Y. Gui, Z. Li, W. Xie, X. Sun, M. Shi, Z. Cai, B. Wang, M. Zhong, J. Li, Z. Lu, N. Gu, X. Zhang, L. Goodman, L. Bolund, J. Wang, H. Yang, K. Kristiansen, M. Dean, Y. Li, J. Wang, Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor, *Cell* 148 (2012) 886–895.
- [192] R.D. Morin, M.D. O'Connor, M. Griffith, F. Kuchenbauer, A. Delaney, A.L. Prabhu, Y. Zhao, H. McDonald, T. Zeng, M. Hirst, C.J. Eaves, M.A. Marra, Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells, *Genome Res.* 18 (2008) 610–621.
- [193] M.R. Capobianchi, E. Giombini, G. Rozera, Next-generation sequencing technology in clinical virology, *Clin. Microbiol. Infect.* 19 (2013) 15–22.
- [194] S.E. Calvo, A.G. Compton, S.G. Hersham, S.C. Lim, D.S. Lieber, E.J. Tucker, A. Laskowski, C. Garone, S. Liu, D.B. Jaffe, J. Christodoulou, J.M. Fletcher, D.L. Bruno, J. Goldblatt, S. Dimauro, D.R. Thorburn, V.K. Mootha, Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing, *Sci. Transl. Med.* 4 (2012) 118.
- [195] S.M. Johnson, F.J. Tan, H.L. McCullough, D.P. Riordan, A.Z. Fire, Flexibility and constraint in the nucleosome core landscape of *Caenorhabditis elegans* chromatin, *Genome Res.* 16 (2006) 1505–1516.
- [196] D.S. Johnson, A. Mortazavi, R.M. Myers, B. Wold, Genome-wide mapping of in vivo protein-DNA interactions, *Science* 316 (2007) 1497–1502.
- [197] G. Robertson, M. Hirst, M. Bainbridge, M. Bilenky, Y. Zhao, T. Zeng, G. Euskirchen, B. Bernier, R. Varhol, A. Delaney, N. Thiessen, O.L. Griffith, A. He, M. Marra, M. Snyder, S. Jones, Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing, *Nat. Methods* 4 (2007) 651–657.
- [198] T.S. Mikkelsen, M. Ku, D.B. Jaffe, B. Issac, E. Lieberman, G. Giannoukos, P. Alvarez, W. Brockman, T.K. Kim, R.P. Koche, W. Lee, E. Mendenhall, A. O'Donovan, A. Presser, C. Russ, X. Xie, A. Meissner, M. Wernig, R. Jaenisch, C. Nusbaum, E.S. Lander, B.E. Bernstein, Genome-wide maps of chromatin state in pluripotent and lineage-committed cells, *Nature* 448 (2007) 553–560.
- [199] M.J. Levene, J. Korch, S.W. Turner, M. Foquet, H.G. Craighead, W.W. Webb, Zero-mode waveguides for single-molecule analysis at high concentrations, *Science* 299 (2003) 682–686.
- [200] E. Schadt, S. Turner, A. Kasarskis, A window into third-generation sequencing, *Hum. Mol. Genet.* 19 (2010) R227–R240.
- [201] www.pacificbiosciences.com.
- [202] J. Korch, A. Bibillo, J. Wegener, P. Peluso, T.T. Pham, I. Park, S. Clark, G.A. Otto, S.W. Turner, Long, processive enzymatic DNA synthesis using 100% dye-labeled terminal

- phosphate-linked nucleotides, *Nucleosides Nucleotides Nucleic Acids* 27 (2008) 1072–1083.
- [203] M. Foquet, K.T. Samiee, X. Kong, B.P. Chaudhuri, P.M. Lundquist, S.W. Turner, J. Freudenthal, D.B. Roitman, Improved fabrication of zero-mode waveguides for single-molecule detection, *J. Appl. Phys.* 103 (2008) 034301.
- [204] J. Korlach, P.J. Marks, R.L. Cicero, J.J. Gray, D.L. Murphy, D.B. Roitman, T.T. Pham, G.A. Otto, M. Foquet, S.W. Turner, Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures, *Proc. Natl. Acad. Sci. USA* 105 (2008) 1176–1181.
- [205] J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, A. Bibillo, K. Bjornson, B. Chaudhuri, F. Christians, R. Cicero, S. Clark, R. Dalal, A. deWinter, J. Dixon, M. Foquet, A. Gaertner, P. Hardenbol, C. Heiner, K. Hester, D. Holden, G. Kearns, X. Kong, R. Kuse, Y. Lacroix, S. Lin, P. Lundquist, C. Ma, P. Marks, M. Maxham, D. Murphy, I. Park, T. Pham, M. Phillips, J. Roy, R. Sebra, G. Shen, J. Sorenson, A. Tomaney, K. Travers, M. Trulson, J. Vieceli, J. Wegener, D. Wu, A. Yang, D. Zaccarin, P. Zhao, F. Zhong, J. Korlach, S. Turner, Real-time DNA sequencing from single polymerase molecules, *Science* 323 (2009) 133–138.
- [206] K.J. Travers, C. Chin, D.R. Rank, J.S. Eid, S.W. Turner, A flexible and efficient template format for circular consensus sequencing and SNP detection, *Nucleic Acids Res.* 38 (2010) e159.
- [207] U. Sotaro, E. Colin, J.K. Aitken, A.F. Benjamin, W.T. Stephen, D.P. Joseph, Real-time tRNA transit on single translating ribosomes at codon resolution, *Nature* 464 (2010) 1012–1017.
- [208] B.A. Mulder, S. Anaya, P. Yu, K.W. Lee, A. Nguyen, J. Murphy, R. Willson, J.M. Briggs, X. Gao, S.H. Hardin, Nucleotide modification at the γ -phosphate leads to the improved fidelity of HIV-1 reverse transcriptase, *Nucleic Acids Res.* 33 (2005) 4865–4873.
- [209] www.halcyonmolecular.com.
- [210] www.zsgenetics.com.
- [211] P.M. Lundquist, C.F. Zhong, P. Zhao, A.B. Tomaney, P.S. Peluso, J. Dixon, B. Bettman, Y. Lacroix, D.P. Kwo, E. McCullough, M. Maxham, K. Hester, P. McNitt, D.M. Grey, C. Henriquez, M. Foquet, S.W. Turner, D. Zaccarin, Parallel confocal detection of single molecules in real time, *Opt. Lett.* 33 (2008) 1026–1028.
- [212] O.L. Krivanek, M.F. Chisholm, V. Nicolosi, T.J. Pennycook, G.J. Corbin, N. Dellby, M.F. Murfitt, C.S. Own, Z.S. Szilagy, M.P. Oxley, S.T. Pantelides, S.J. Pennycook, Atom-by-atom structural and chemical analysis by annular dark-field electron microscopy, *Nature* 464 (2010) 571–574.
- [213] www.reveo.com.
- [214] Hall, et al., Hybrid pore formation by directed insertion of [alpha]-haemolysin into solid-state nanopores, *Nat. Nanotechnol.* 5 (2010) 874–877.
- [215] C. James, W. Hai-Chen, J. Lakmal, P. Alpesh, R. Stuart, B. Hagan, Continuous base identification for single-molecule nanopore DNA sequencing, *Nat. Nanotechnol.* 4 (2009) 265–270.
- [216] S. David, H. Andrew, K. Jochen, M. Ellina, M. Giovanni, B. Hagan, Nucleobase recognition in ssDNA at the central constriction of the α -hemolysin pore, *Nano Lett.* 10 (2010) 3633–3637.
- [217] B. McNally, A. Singer, Z. Yu, Y. Sun, Z. Weng, A. Meller, Optical recognition of converted DNA nucleotides for single-molecule DNA sequencing using nanopore arrays, *Nano Lett.* 10 (2010) 2237–2244.
- [218] D. Branton, D. Deamer, A.B. Marziali Hagan, S. Benner, T. Butler, M. Di Ventra, S. Garaj, A. Hibbs, X. Huang, S. Jovanovich, P. Krstic, S. Lindsay, X. Ling, C. Mastrangelo, A. Meller, J. Oliver, Y. Pershin, J. Ramsey, R. Riehn, G. Soni, V. Tabard-Cossa, M. Wanunu, M. Wigginton, J. Schloss, The potential and challenges of nanopore sequencing, *Nat. Biotechnol.* 26 (2008) 1146–1153.
- [219] www.ceesdekkerlab.tudelft.nl/research/projects/nanopores/.
- [220] G.F. Schneider, S.W. Kowalczyk, V.E. Calado, G. Pandraud, H.W. Zandbergen, L.M.K. Vandersypen, C. Dekker, DNA translocation through graphene nanopores, *Nano Lett.* 10 (2010) 3163–3167.
- [221] Q. Zhao, Y. Wang, J. Dong, L. Zhao, X.F. Rui, D. Yu, Nanopore-based DNA analysis via graphene electrodes, *J. Nanomater.* 2012 (2012) 1–5.
- [222] S. Kumar, C. Tao, M. Chien, B. Hellner, A. Balijepalli, J.W.F. Robertson, Z. Li, J.J. Russo, J.E. Reiner, J.J. Kasianowicz, J. Ju, PEG-labeled nucleotides and nanopore detection for single molecule DNA sequencing by synthesis, *Sci. Rep.* 2 (2012).
- [223] www.nanoporetech.com.
- [224] M. Krems, M. Zwolak, Y.V. Pershin, M. Di Ventra, Effect of noise on DNA sequencing via transverse electronic transport, *Biophys. J.* 97 (2009) 1990–1996.
- [225] B. Luan, H. Peng, S. Polonsky, S. Rossmagel, G. Stolovitzky, G. Martyna, Base-by-base ratcheting of single stranded DNA through a solid-state nanopore, *Phys. Rev. Lett.* 104 (2010) 238103.
- [226] L.M. Baudhuin, A new era of genetic testing and its impact on research and clinical care, *Clin. Chem.* 58 (2012) 1070–1071.

Update

Molecular Genetics and Metabolism

Volume 114, Issue 3, March 2015, Page 484

DOI: <https://doi.org/10.1016/j.ymgme.2015.01.009>



Contents lists available at ScienceDirect

Molecular Genetics and Metabolism

journal homepage: www.elsevier.com/locate/ymgme



Corrigendum

Corrigendum to “A glimpse into past, present, and future DNA sequencing” [Mol. Genet. Metab. 110 (2013) 3–24]



Marcos Morey^a, Ana Fernández-Marmiesse^a, Daisy Castiñeiras^a, José M. Fraga^b, María L. Couce^{c,*}, José A. Cocho^a

^a Unidad de Diagnóstico y Tratamiento de Enfermedades Metabólicas Congénitas, Hospital Clínico Universitario de Santiago, A Choupana s/n, 15706 Santiago de Compostela, Spain

^b Unidad de Diagnóstico y Tratamiento de Enfermedades Metabólicas Congénitas, Departamento de Pediatría, Hospital Clínico Universitario de Santiago, Universidad de Santiago de Compostela, A Choupana s/n, 15706 Santiago de Compostela, Spain

^c Unit of Diagnosis and Treatment of Congenital Metabolic Diseases, Department of Pediatrics, Hospital Clínico Universitario de Santiago de Compostela, CIBERER, Health Research Institute of Santiago de Compostela (IDIS), A Choupana, s/n, 15706 Santiago de Compostela, A Coruña, Spain

The authors regret that the affiliation of María L. Couce is incorrect in the published version. The corrected affiliation is shown above.

DOI of original article: <http://dx.doi.org/10.1016/j.ymgme.2013.04.024>.

* Corresponding author.

E-mail address: maria.luz.couce.pico@sergas.es (M.L. Couce).