

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Application of Next-Generation Sequencing in the Era of Precision Medicine

Michele Araújo Pereira,
Frederico Scott Varella Malta,
Maíra Cristina Menezes Freire and
Patrícia Gonçalves Pereira Couto

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.69337>

Abstract

Next-generation sequencing (NGS) technologies represented the next step in the evolution of DNA sequencing, through the generation of thousands to millions of DNA sequences in a short time. The relatively fast emergence and success of NGS in research revolutionized the field of genomics and medical diagnosis. The traditional medicine model of diagnosis has changed to one precision medicine model, leading to a more accurate diagnosis of human diseases and allowing the selection of molecular target drugs for individual treatment. This chapter attempts to review the main features of NGS technique (concepts, data analysis, applications, advances and challenges), starting with a brief history of DNA sequencing followed by a comprehensive description of most used NGS platforms. Further topics will highlight the application of NGS towards routine practice, including variant detection, whole-exome sequencing (WES), whole-genome sequencing (WGS), custom panels (multi-gene), RNA-seq and epigenetic. The potential use of NGS in precision medicine is vast and a better knowledge of this technique is necessary for an efficacious implementation in the clinical workplace. A centralized chapter describing the main NGS aspects in the clinic could help beginners, scientists, researchers and health care professionals, as they will be responsible for translating genomic data into genomic medicine.

Keywords: NGS, precision medicine, diagnostic, exome, panels, diseases, welfare

1. Introduction

Precision medicine is a new way of practising medicine, which has been gaining strength in recent years, is based on the individual characteristics of each patient (genetic, environmental, behavioural) to optimize and customize strategies for prevention, detection and therapy [1, 2]. The molecular knowledge has contributed strongly to the advancement of precision medicine, providing specific strategies for target therapies and diagnosis of patients with cancer, Mendelian diseases and others. Statistics indicated that traditional clinical practices sometimes lead to poor health outcomes and also a waste of medical resources. It is estimated that about 75 billion US dollars per year (30% of health care expenditure) are destined for unnecessary or ineffective treatments in the USA [3].

As a result of the genome project, many molecular tools have been developed and allow medical and scientific groups to improve patient management based on a better understanding of disease biology, providing a more specific and accurate prevention and treatment of diseases [4]. Precision medicine redefines the way traditional medicine is practised. There is a great deal of investment nowadays in prevention using these new technologies, as opposed to old medicine based on treatment since the disease was already evident or irreversible [2].

In recent times, Sanger sequencing, referred to as a 'first-generation' sequencing method, has partly been replaced by 'next-generation' sequencing (NGS) methods [4, 5]. NGS allows identifying biomarkers for early diagnosis as well as for personalized treatments. The emergence of NGS has changed the way clinical research, basic and applied science are done. The NGS allows producing millions of data with a smaller investment [4, 6]. Among the available NGS applications, one of them will be the resequencing of the human genome and the better genetic understanding of various human diseases. A great challenge will be the interpretation of this great number of data and its translation for the medical application [6]. One of the major near-term medical impact of the NGS revolution will be the elucidation of mechanisms of human pathogenesis, leading to improvements in the diagnosis and the selection of treatment and prevention. Thanks to second-generation sequencing technologies, it has become easier to sequence the expressed genes ('transcriptomes'), known exons ('exomes') and complete genomes of patient's samples [7].

This chapter encompasses revised concepts, applications, advances, limitations and the history of technological advances until the emergence of NGS technique in the era of precision medicine, starting with a brief history of DNA sequencing followed by a comprehensive description of most used NGS platforms, sequencing chemistries methodology and general workflows. Further topics will highlight the application of NGS towards routine practice, including variant detection, whole-genome sequencing (WGS), whole-exome sequencing (WES) and multi-gene panels. A centralized chapter describing the main NGS features in the clinic could help beginners, scientists, researchers and health care professionals, as they will be responsible for translating genomic data into genomic medicine.

2. From Sanger to NGS sequencing

In 1908, Garrod introduced his concept 'the inborn error of metabolism' that changed the areas of biochemistry, genetics and medicine [8]. His principal contribution was the understanding about the relationship between gene-enzyme, the molecular basis of genetic diseases. Although today this concept is considered outdated because of discoveries like RNA splicing, RNAi and others, its development allowed the researchers to understand how changes in DNA sequence could cause genetic disease. This finding increased the interest of scientists to know about human DNA sequence and mutations.

The search to know the nucleotide sequence of DNA began in the 1960s with several studies that demonstrated new methods with different strategies [9–13], but it was in 1977 that Sanger developed the method called 'Chain-termination' that became the most used method (first generation) to sequencing DNA (**Figure 1**). The method consisted of the use of dideoxynucleotides (ddNTPs), which are deoxynucleotide analogs (dNTPs) that disrupt DNA synthesis, and the separation of the different DNA fragments in a gel. These special nucleotides were radiolabeled and therefore the sequence could be inferred after the disclosure of gel autoradiography [14]. Numerous modifications have been made in this technique to make the method more efficient, robust and sensitive. Among them are the substitution of nucleotide radiolabeled to fluorescence that allowed the sequencing reaction to occur in one tube [15], the development of the polymerase chain reaction [16], the separation of DNA fragments by capillary electrophoresis [17] and later the development of equipment that allowed the sequencing of

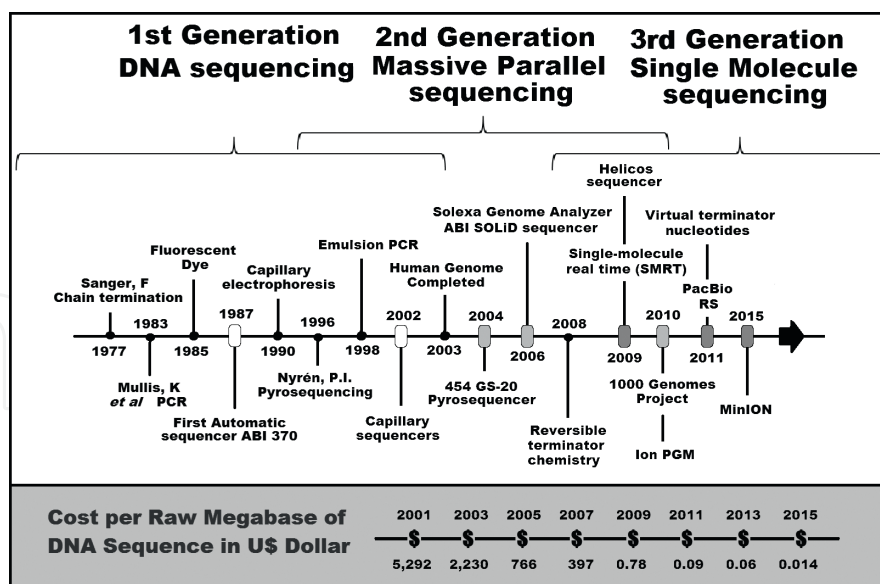


Figure 1. Timeline of DNA sequencing evolution from Sanger to NGS and the cost per raw megabase of DNA sequenced [17]. Equipment of all generations is still being improved and released commercially. Dot: milestones; rectangle: equipments; White: first-generation sequencing; Light gray: second-generation sequencing; Dark gray: third-generation sequencing.

more complex genomes. The most famous sequencing project, the Human Genome Project, produced in 13 years 3 billion of sequenced bases with the estimated cost around \$2.7 billion [18]. To date, Sanger is still the gold-standard method in diagnostic tests and although the most recent methods have a much higher processing capacity, confirmation of some findings is made using this method.

The second generation of DNA sequencing can be defined as the era of the parallel massive sequencing on a micro scale. The Pyrosequencing method developed by Nyrén and colleagues in 1996 was the starting point for this generation. This technique differed substantially from previous ones because it did not use radio or fluorescence-labelled nucleotides and there was no need of electrophoretic run. The method is based on the action of two enzymes: ATP sulfurylase and luciferase. ATP sulfurylase converts pyrophosphate released in nucleotide incorporation into an ATP molecule that is used by luciferase substrate. This process releases light signal in proportion to the amount of nucleotides incorporated, and the sequence can be determined according to the serial addition of nucleotides [19]. Later on, this technology was improved and licensed generating the first 'second-generation' equipment, known as 454 (Roche). Among the improvements made, there are the DNA binding in beads through an adapter and the amplification of this DNA in water-in-oil microreactors (emulsion PCR). These changes and the use of microplates that compartmentalized the process and high-definition detection systems dramatically increased the amount of DNA sequenced and defined the second generation [20]. The disadvantage of this technology is related to homopolymer regions because of difficulty in interpreting the signal strength when five or more nucleotides are incorporated in a single wash cycle. Other technologies were then developed, such as that used by Illumina which consists of binding the DNA in a flow-cell through adapters, and the parallel massive amplification occurs in clusters for each DNA strand that was originally bound in the flow-cell, called bridge-amplification. This process generates paired-ends sequences that are an advantage over other methodologies, since they improve the accuracy of mapping, mainly in repetitive regions or where DNA rearrangements or gene fusions occur. The method uses 'reversible terminator chemistry' which is a modified fluorescent dNTP that reversibly blocks DNA synthesis, so the addition of each nucleotide can be synchronized and monitored by a charge-coupled device (CCD) sensor [21]. This is one of the most accurate and with lowest error rate of sequencing methodologies used currently; however, it generally requires higher DNA concentration. Another methodology is based on oligonucleotide ligation sequencing known as SOLiD and developed by Applied Biosystems (now Thermo Fisher Scientific). The method does not do sequencing by synthesis but by ligation of oligonucleotides fluorescence-labelled. Each probe is an octamer, which contains two known nucleotides in the 3' end followed by six degenerated nucleotides with one of four fluorescent labels linked to the 5' end. After probe annealing and ligation, fluorescent dye is cleavage and a new probe is ligated. Multiple cycles are performed according to the read length. The template from primer (n) is removed and the second round of sequencing is performed with a primer complementary to the (n-1) position [22]. This method shows good results; however, it is considered slow compared to the others and therefore was replaced by Ion Torrent (Thermo Fisher Scientific) technology. Like 454, the DNA bound in a bead is massively amplified by emulsion PCR and detection occurs in picotiter wells using complementary metal-oxide-semiconductor (CMOS) due to the pH difference caused by the

release of H⁺ ions in the nucleotide incorporation. This methodology is the first to use a detection method that does not work with light signal [23]. The advantage of this technology is the speed of the process and the low cost of the equipment; however, it has the same problem about the detection of homopolymers. The second generation of the sequencing was marked by the high capacity of the sequencers in the generation of data in a single run and consequently the computational development-like bioinformatics tools to analyse them. The cost of sequencing decreased dramatically at this stage. At the beginning of the first-generation sequencing (2001), the approximate cost per megabase sequenced was \$5292.39 and at the end of this phase (2007) was \$397.09, while in the second generation the sequencing cost was \$102.13 (2008) and at the end (2015) only \$0.014 [18], showing a more pronounced decline in this phase (**Figure 1**).

There are some discussions about which technology marked the beginning of the third generation [24–27]. In this review, we will consider the technology of single-molecule sequencing (SMS), which has no need to amplify the DNA. The first technology to use SMS was ‘virtual terminators’ based on a method very similar to Illumina, but a single DNA molecule is fixed in a flow-cell with 25 channels. The process occurs in cycles where the dNTPs are incorporated and the corresponding fluorescence is captured by a CCD camera. This process generates short readings (25 bp) and it is considered slow and there is a lot of noise in the signal [28]. Despite being the first third-generation sequencing technology, its history was brief because the company Helicos Biosciences filed for Chap. 11 bankruptcy. Another technology developed is the ‘single molecule real time’ (SMRT) that is commercialized by Pacific Biosciences. The SMRT consists of the immobilization of a single molecule in a chamber called ‘zero-mode waveguide (ZMW)’ where the incorporation of the fluorescent nucleotides occurs. ZMW allows the incorporation of each nucleotide to be monitored in real time and without interference from other light signals. The reads are very long (40 kb) and allow detecting modified bases [29, 30]. Finally, the technology of ‘nanopores’ consists of conducting a molecule of DNA or RNA through a biological or not nanopore. The detection occurs due to differences in the current of ions generated by each nucleotide. The reads are incredibly long (500 kb), and the process is extremely fast without the need for special nucleotides. The company Oxford Nanopore Technologies (ONT) is the first company to commercialize sequencers using this technology, including a portable version (MinION) that was used to sequence a mixture of bacteriophage, *Escherichia coli* and *Mus musculus* DNA at the international space station (ISS) [31]. In common, these technologies still have high error rates that are improving with the development of technology. Its main use today is to aid in the assembly of complex regions of the genome where gene fusions, large deletions and insertions and repetitive regions occur. The third generation will further revolutionize precision medicine, enabling sequencing at lower cost and enabling this to occur virtually anywhere.

3. Clinical applications

In recent times, NGS has made possible a better understanding of genetic diseases and became a significant technological advance in the practice of diagnostic and clinical medicine

[32]. NGS allows the analysis of multiple regions of the genome in one single reaction and has been shown to be a cost-effective and an efficient tool in investigating patients with genetic diseases. Genetic data produced via NGS provides significant benefits to medical practice including accurate identification of biomarkers of disease, detecting inherited disorders and identifying genetic factors that can help predict responses to therapies [32, 33]. However, recommendations on clinical implementation of NGS that are still in discussion and that hamper its use in the genetic clinic. A variety of molecular diagnostic test use sequencing technology, such as single- and multi-gene panel tests, cell-free DNA for non-invasive prenatal testing, whole-exome sequencing (WES), whole-genome sequencing (WGS). Considering that the use of NGS as a diagnostic tool is recent, there are challenges including when to order, on whom to order and how to interpret and communicate the results to the patient and family [32]. Therefore, it is necessary to understand the application, strength and limitations of the different approaches to recognize which one is the most suitable for your case. In the following topics, we will emphasize common applications of this technology into clinical practice.

3.1. Multi-gene panels

The traditional approach still holds great value for many disorders. Single-gene testing is indicated when the clinical features for a patient are typical for a particular disorder and the association between the disorder and the specific gene is well established and has the minimal locus heterogeneity [34]. However, many genetic conditions are intractable to diagnostic evaluation, mainly because of the clinical variability and genetic locus heterogeneity, such as cardiomyopathies, epilepsy, congenital muscular dystrophy, X-linked intellectual disability and cancer susceptibility in families with atypical phenotypes [35]. The diagnostic process is exhausted, with clinical assessment followed by sequential laboratory testing, in most cases tests being negative. In cases with unidentified genetic conditions (e.g., developmental delay/cognitive disability and autism spectrum disorders), the diagnosis rate can vary greatly [36] and a multi-gene panel is more appropriate. In diagnostic of cancer, for example, Tothill and colleagues [37] illustrate the application of these multi-gene panel by analysing samples of patients with cancers of unknown primary (CUP). The clinical management of patients with CUP is hampered by the absence of a definitive site of origin and this kind of NGS analysis could help to define new therapeutic options.

In multi-gene panel tests, many genes associated with a specific phenotype are sequenced and analysed concomitantly, decreasing cost and improving efficiency of genetic diagnostic [37]. The number and which genes will be evaluated for the same or similar indications may vary significantly among different clinical laboratories and several considerations need to be taken for gene inclusion. The majority of authors believe that only genes with a strong disease association should be included since the ability to interpret their findings is much better due to clinical evidence [38]. However, some authors consider including associated genes that have overlapping phenotypes for the purpose of differential diagnosis, or all possible genes that are remotely associated with the phenotype of interest with the objective of a better and faster diagnostic [34]. For cancer diagnostic, multi-gene panel may include high-penetrance genes as well as associated genes with a moderate increase in risk [35].

The transition from single-gene to multi-gene testing should not compromise the sensitivity of the test to identify variants, mainly at genes that are responsible for a significant proportion of the defects (core genes). The sensitivity of NGS does not depend only on horizontal coverage but the vertical coverage is important as well [39]. Additional genes will increase the chance of the diagnostic, but this should not be at cost of missing mutations that would previously have been detected by single-gene testing [38]. Sanger sequencing or other available techniques can help to solve this problem for filling in low-coverage and no-coverage regions.

3.2. Whole-genome and whole-exome sequencing

Whole-genome sequencing (also known as WGS, full-genome sequencing, complete genome sequencing or entire genome sequencing) is the process of determining the complete DNA sequence of an organism's genome at a single time. The major benefit of WGS is completed coverage of the genome, including promoters and regulatory regions. In whole-exome sequencing (WES), all coding regions are sequenced with a relatively deeper depth. Compared to WGS, the major advantage of WES is a significant cost reduction [40].

Human genome comprises $\sim 3 \times 10^9$ bp having coding and non-coding sequences. About 3×10^7 bp (1%) (30 Mb) of the genome are the coding sequences [33]. It is estimated that 85% of the disease-causing mutations are located in coding and functional regions of the genome [41, 42]. For this reason, sequencing the complete coding regions (exome) has the power to uncover the causes of large number of rare, mostly monogenic, genetic disorders as well as predisposing variants in common diseases and cancers [33]. In 2009, Choi and colleagues first showed the value of WES in the medical practice by making genetic diagnoses of congenital chloride diarrhoea in patients suspected of Bartter syndrome, a renal salt-wasting disease. WES was conducted on six patients who do not show any mutations in classic genes for Bartter syndrome. Results revealed homozygous deletion in *SLC26A3* gene for all patients, which provided a molecular diagnosis of congenital chloride diarrhoea that was later confirmed on clinical evaluation. This result was the first to show the value of WES in making a clinical diagnosis and several similar studies have followed [43].

There are certain considerations to order WES instead of other NGS tools [32]. Although exomes are supposed to cover all the protein-coding regions of the genome, the average coverage in many platforms tends to be between 85 and 95% [32, 44]. This means that a particular gene of interest that is closely linked to patient's phenotype may not be covered, completely or partially. There are many reasons that include poorly performing capture probes due to high GC content, sequence homology or repetitive sequences. A targeted approach, such as NGS single- or multi-gene panels, on the other hand, has higher or even complete coverage of all the specific genes by filling in the gaps with complementary technologies such as Sanger sequencing or long-range PCR. Besides offering a more comprehensive coverage of the 'known' phenotype-specific gene panels, this targeted approach also allows for deeper coverage of these genes compared to WES, which provides greater confidence in the variants detected. However, all NGS tools are still prone to sequencing artefacts, and Sanger sequencing is recommended to confirm the variants detected before returning the results to the patient [44]. In addition, the patient and their family need to be aware of all the nuances

related to WES and WGS [45]. It is important to let them know that the test may not yield positive results, and it is crucial to clarify that even positive results can offer diagnoses but do not improve prognosis and treatment.

To request an exam that uses the WES technique, one must start collecting as much information as possible about the patient. It is important to have a detailed family history, phenotype condition, symptoms and also, if possible, the inheritance pattern of the suspected disease [46]. With the phenotype and pedigree information, a systematic review of literature and databases should be performed to guide the clinician on which gene(s) are crucial and must be analysed. In cases of genetic heterogeneity, targeted NGS may be the preferred approach. On the other hand, if the disease mechanism is unknown, WES may be the best choice [47].

WES can result in approximately 60,000–100,000 genetic variants that can be classified into pathogenic, benign or with uncertain significance (VUS) [48]. With WES, a single pathogenic variant that is probably the cause of the patient phenotype can be detected in about 20–36%. For the other cases, it is possible to find multiple candidate variants or even no one. If no candidate variants are found, there are many reasons for it that include poor coverage or the mutation residing outside the protein-coding region of the gene, clinical summary with insufficient information or the defect is not due to a simple nucleotide change in a single gene [49–53].

The outcome of an exome should be evaluated by a multidisciplinary team that is involved with each patient's case. A discussion is necessary between physicians, geneticists, and other health professionals about all the clinical and laboratory findings to make a link with phenotype, family history and symptoms. It is necessary to review the WES results, scientific literature and medical information [32]. If more than one candidate variant is detected, this multidisciplinary team must perform further evaluation(s) to determine which of the variant is causing the phenotype. Finally, if the test results are negative, reasons for this should be discussed in the report. As the use of this tool is becoming more frequent and more accessible, it is possible that in the near future new pathogenic variants and genetic syndromes will be described and characterized, which causes these negative results to be reanalysed within a few years [32].

In cases of suspicion of Mendelian disease, the exome sequencing is usually indicated for the detection of rare variants and samples from the patient and his/her parents could be needed. This is usually the standard setting in cases where the Sanger sequencing of the candidate gene gave negative result or so there are multiple genes that must be tested for the condition that would be costly and time consuming. In most cases, the results obtained from WES reach a molecular diagnosis but do not alter the management, treatment or prognosis [32, 54].

Targeted exome sequencing is becoming increasingly popular in oncology for assessing the full sequence of cancer-related genes. Targeted exome sequencing also facilitates sequencing at a greater depth, and thus the identification of subclonal mutations. Alternately, rather than sequencing the full exome sequence, it is possible to look at all the genes reported to be related to cancer in general. Although hotspot mutation testing facilitates large-scale sequencing of many samples, it does limit the knowledge that is acquired through sequencing because it

limits the evaluation to small regions in selected genes. Consequently, small, targeted NGS panels increase the possibility of omitting relevant mutations for which evaluation is not being conducted, thus limiting the clinical knowledge that is gained through WES. WES could highlight novel insights into cancer mechanisms; identification of the DNA sequence of cancer cells in comparison with that of normal cells could help to reach an in-depth understanding of cancer. Using WES, it is also feasible to check germline and somatic mutations in human cancers [33].

Approximately 5–10% of cancers are hereditary. WES allows testing of multiple genes at once and greatly improves the variation detection rate. Many patients with hereditary cancer have tested negative for one specific genetic variation, but with WES, it is easier to find causative mutations. In a study of 300 high-risk breast cancer families, it was found previously undetected mutations in 52 probands and the reduced sequencing costs and turnaround time made the approach even more practical in clinics [55].

To detect familial germline mutations, WGS might be advantageous for WES-negative cases in families with a great chance of carrying a genetic variant [56]. The major technical advantage of WGS is that the specificity is theoretically 100% (average 95–98% in practice, practically without gaps) with a uniform coverage in the regions of interest (ROIs) throughout the input material. Thus, the chance of losing disease-causing variants due to technical errors is much lower with WGS [57–59]. The major challenge in applying this tool on a medical routine is the great costs, the complex pipeline for data analysis and data interpretation. However, in the near future, the costs of NGS should be lowered, studies on genetics over non-coding regions should be improved and more approach will be implemented. With that, WGS should be performed regularly for diagnostic in order to find the causative genetic variants [56].

Under gene panel analysis, about 70–92% of all cases remain negative, depending on the disease. It is expected that important genes will not be contemplated with these tools, making WES and WGS analysis more appropriate to identify genetic variants in cases of familial syndromes. These tools (WES and WGS) have already been reported in identifying several risk genes for various types of cancer such as the *PALB2* and *ATM* genes in pancreatic cancer, the hereditary pheochromocytoma susceptibility gene *MAX* [60] or the hereditary colorectal cancer moderate-risk genes *POLD1* and *POLE* [61].

Nowadays, the clinical utility of WES and WGS as a generic test for mutation discovery for every genetic diagnostic question is not yet appropriate [62] and should be directed to specific patient groups [63]. This limitation is due to the high cost, the need of complex bioinformatics pipelines, large storage capacity and the expected high number of VUS detected.

3.3. RNA-sequencing

A transcriptome represents the complete set of RNA molecules from any genome at any time or condition and RNA plays essential role in several biological processes, including untranslated RNA species such as microRNAs (miRNAs). RNA-sequencing (RNA-seq) consists of an in-depth RNA analysis through NGS technologies and became the state-of-art technique for transcriptomic [64]. A typical RNA-seq experiment consists of a good experimental design,

sample preparation, library construction, sequencing and data analysis. However, due to several experimental options available, a careful planning and cost estimation is necessary before starting. These include number and type of replicates (technical vs. biological), sequencing platform (e.g. Illumina, Ion Torrent), library preparation method (e.g. rRNA depletion or mRNA enrichment; strand-specific or not; single or paired end), throughput, read length, sequencing depth and coverage. RNA-seq best practices can be found in Chap. *RNA-seq: Applications and Best Practices* from this book.

RNA-seq enables detection of novel genes and isoforms, gene fusions, splice and chimeric variants, genomic alterations and gene expression quantification. Although RNA-seq outperforms microarray in transcriptomic analysis [65], its clinical application is still in its infancy and, for instance, will not replace current approaches. RNA-seq is considered a complementary method depending on the needs and resources available, assisting clinicians in making decisions. In clinical practice, RNA measurement has applications across different areas in human health such as therapeutic selection, disease diagnostic and treatment [66].

Clinical diagnosis of infectious disease through RNA-seq is still rare, since quantitative PCR (RT-qPCR) assays are still the most common technique used for viral detection and genotyping. Applications of NGS in virology diagnostic can be used for analysis of patients with unexplained illness, especially during outbreaks and epidemics [67–70]. It also includes the identification of novel pathogens [71–74], viral community characterization [75–77], whole viral genome reconstruction [73, 78, 79], antiviral drug resistance [80–83], epidemiology [84–87] and transcriptomic [88–90]. The use of NGS in virology is increasing the knowledge of viral infection dynamics and their correlation with human health and treatment.

For oncology, RNA-based cancer diagnostics is being used by clinical oncologist to define tumour transcriptome due to its potential to guide treatment and drug therapy [91]. Its application are especially related to gene expression profile and variants, and gene fusions detection. The pathogenicity of gene fusions in cancer is well known. Most gene fusions are correlated with specific tumour subtypes, representing diagnostic biomarkers and leading to novel therapeutic opportunities and benefits [92–94]. Some pharmacological treatments are already in clinical use [94]. Key somatic DNA mutations can also represent cancer biomarkers and can be identified by transcriptomic mapping [95–98].

Gene expression in cancer is still quantified by non-sequencing methods (e.g. RT-qPCR and microarrays) [91]. RNA-seq can measure expression of tumour antigens or immune checkpoint receptors and ligands after a given treatment, giving some answers about patient drug response [91, 99, 100]. Gene expression signatures can also be used for cancer types' classification that directly impact prognosis and treatment definition and response [100].

NGS can also be applied for circulating tumour RNA (ctRNA) discovery. The analysis of ctRNA in plasma is still in its beginning and presents specific challenges. ctRNA degrades faster than circulating tumour DNA (ctDNA) and needs to be purified rapidly or added in preservative solutions (e.g. TRIzol) and frozen at -80°C , not always an accessible technique to many clinical sites [101]. Despite these challenges, ctRNAs represent good biomarkers of early detection of multiple tumour types, such as breast, lung, prostate and colorectal cancers

[101–109]. NGS is a more powerful tool for ctRNA detection; however, RT-qPCR remains more usable for clinical diagnostic applications [110].

3.4. Epigenetics

An emerging field that has a huge impact on medicine and clinical diagnostic is epigenetics. The term was coined by Conrad Waddington in the 1940s and refers to the study of heritable changes in gene activity and expression that do not involve the DNA sequence itself, that is, a change in phenotype without a change in genotype [111, 112]. Additional information about epigenetics history can be found in Ref. [113]. Epigenetics mechanisms represent another layer of gene regulation and NGS allowed to understand the epigenetics status on a large scale and at a single base-resolution, including mainly DNA methylation, histone modification and non-coding RNA (ncRNA)-associated silencing [111, 112].

DNA methylation was the first epigenetic mechanism identified and is the best known and the most frequent in human cancer. It involves covalent modification of cytosine through the addition of a methyl group to cytosines of CpG (cytosine/guanine) islands [111, 112]. This methylation is maintained by DNA methyltransferase (DNMTs) and plays roles for gene transcriptional repression, transposable elements silencing and viral defence [111]. Unmethylated DNA is found in active regions of chromatin, and methylated DNA is found in inactive regions [112].

Post-translational histone modifications are markers for chromatin activity through acetylation and methylation of conserved lysine residues on the amino-terminal tail domains [112]: acetylation is found in active regions of chromatin, whereas hypoacetylation is found in inactive euchromatic or heterochromatic regions [111, 112]. Enzymes involved in this process include histone deacetylases (HDACs), histone acetylases and histone methyltransferases [112]. These and other post-translational histone modification processes (e.g. phosphorylation) result in distinct histone modification patterns that form a ‘histone code’ [114].

Since epigenetic mechanisms regulate DNA accessibility, perturbations of the cell epigenetic pattern affect gene expression and can give rise to human diseases, that can be inherited or somatically acquired [111, 112]. Prader-Willi, Angelman and Beckwith-Wiedemann syndromes, for example, are the best characterized congenital imprinting disorders [111, 115, 116].

4. Data analysis

Data analysis is a critical step of NGS tests. This analysis consist of a primary analysis, in which the base pairs are called and quality score are generated; a secondary analysis, numerous reads are aligned to the human reference sequence; and a tertiary analysis which consists of variant calling and annotation [117]. Many databases are useful for helping the variant annotation, such as the 1000 Genome Project [118], dbSNP database [119], Clinvar—NCBI [120], LOVD—Leiden Open Variation Database [121], The Cancer Genome Atlas (TCGA) [122] and others. However, information from these sources can contain ambiguous and insufficient information. Variants detected should be reported according to Human Genome Variation

Society (HGVS) recommendations, with information of the human reference genome version and transcript information used to variant description [117]. The reference coding sequence should be preferably from the RefSeq database [123].

All pathogenic, likely pathogenic and VUS variants have to be reported. Secondary or incidental finding (IF) is one significant matter, especially for WES, WGS and multi-gene panels, and its report will depend on local practice [38].

An in-house database containing all relevant variants identified in the laboratory provides an important tool in order to allow for further annotations, which greatly streamline the diagnostic process. Furthermore, an in-house database, linking patients and variants can help when a variant is re-classified. In this case, the laboratory is responsible for re-contacting the clinicians of the patients that are possibly affected by the new status of the variant [38].

4.1. Sanger sequencing validation

Concerning the limitations of technology, the false positive rate for NGS, a second method, as Sanger sequencing, is required to confirm any findings with possible clinical significance. The laboratory must be able to guarantee that report variants are true variants; therefore, it is essential to mention that the variant reports were confirmed by Sanger method. An NGS technology will likely evolve, and within a few years confirmation might prove to be unnecessary [34, 39].

In some cases, mainly in large panels, complementing NGS testing with Sanger sequencing is inevitable. This limitation of NGS is dependent on the platform and on the enrichment methods, once that there are a number of strategies available with advantages and disadvantages. Sanger sequencing can also be used to fill regions that fail to amplify for having sequence complexities, such as sequence homology with pseudo genes, highly repetitive regions, GC-rich content, allelic dropout, or regions that are supported by an insufficient number of reads to call variants confidently [34]. However, in practice, the laboratories can opt to apply different settings for NGS tests. Three kinds of tests of multi-genes panel are identified: (A) the lab informs that more than 99% of interest region are covered, and all the gaps are filled with Sanger sequencing; (B) the lab describes which regions are sequenced and fills some specific gaps (core genes) with Sanger sequencing; and (C) no additional Sanger sequencing is offered [38]. It is essential to mention the horizontal coverage acquired in the test and the limitations of these tests in a disclaimer [39].

5. Challenges

The diversity and rapid evolution of NGS technology causes many challenges associated with data generation, data manipulation and data storage [124]. Some of the major issues with analysis, interpretation, reproducibility and accessibility of NGS data includes: (A) NGS is still too expensive to be accessible by small labs or an individual; (B) data analysis is time-consuming and needs sufficient knowledge of bioinformatics; (C) the short sequencing read lengths supported by NGS is one of the major shortcomings which limit its application,

especially in de novo and highly repetitive regions sequencing; (D) data processing steps or bioinformatics is one major bottleneck for the implementation of NGS; (E) routine analysis of NGS data requires multidisciplinary teams; (F) it is critical to standardize the quality metrics for the NGS data generated. These include validation and comparison among platforms, data reliability, robustness and reproducibility, and quality of assemblers; (G) it is crucial to have a complete knowledge of family and personal history of the patient to help define the ideal analysis method, the analysis of the results obtained, and the post-test counselling and management [124–127].

Despite some challenges, it is hard not to be optimistic about the future of personalized genome sequencing and its potential impact on patient care and the advancement of knowledge of human biology and disease.

5.1. Regulation on NGS tests

With the advancement of gene-sequencing technologies, numerous opportunities have arisen in the genetic diagnostic, preventive medicine and other areas of human health. As a result, several life science companies and clinical laboratories started their activities in this field offering equipment and supplies as well as molecular tests using the new-generation (parallel massive) sequencing methodology. However, most manufacturers do not market IVD products (in vitro diagnostic), but, in general, these products are classified as RUO (research use only). In practice, this difference in the classification of products and reagents represents serious implications on health. Products classified as IVD are regulated and therefore follow technical standards in their production and use, and consequently the efficiency must be guaranteed by the manufacturer. The ISO 13485 [128] is often used to ensure the quality of medical products, but other regulatory agencies such as the US Food and Drug Administration (FDA) may require other tests to prove this product is safe and effective, which is necessary for the product be classified as IVD and be commercialized on the American market. The same applies to the CE-IVD Marking in the European Economic Area (EEA). These requirements are part of an effort to ensure that users of these services and devices do not seek unnecessary treatment, delay their treatment or are exposed to inappropriate therapies. In the case of RUO products, none of these situations can be guaranteed, so the manufacturer will only be obliged to replace the product or its cost if it is performing improperly. In fact, some manufacturers may use standards of good manufacturing practice in the production of RUO equipment and supplies, but rarely perform tests to prove their efficiency in a particular case of diagnostic.

In some cases due to the need to respond quickly to the market, especially in areas where the technological advance exceeds the regulatory capacity, some agencies allow the use of tests developed by clinical laboratories. The regulation in these cases is very simpler and favours the development of new technologies as the case of new-generation sequencing (NGS). However, these tests should also be used with caution, and the laboratories must prove its accuracy, or otherwise there may be the same hazards of products classified as RUO. In 2013, the US FDA agency required to genetic testing company 23andME to suspend the marketing of its products until it receives clearance from the agency. In a letter addressed to one of its

founders, the agency states its concern about the use of one of its tests and the implications on the health of the patient in case of false results.

Some of the uses for which PGS (Personal Genome Service) is intended are particularly concerning, such as assessments for BRCA-related genetic risk and drug responses (e.g., warfarin sensitivity, clopidogrel response, and 5-fluorouracil toxicity) because of the potential health consequences that could result from false positive or false negative assessments for high-risk indications such as these. For instance, if the BRCA-related risk assessment for breast or ovarian cancer reports is false positive, it could lead to undergo prophylactic surgery, chemoprevention, intensive screening, or other morbidity-inducing actions, while false negative could result in failure to recognize an existing risk that may exist. [129]

This example illustrates the importance of evaluating the analytical characteristics of diagnostic tests as well as the reagents and equipment used to perform these tests. In 2013, Illumina was the first company to get FDA approval for the commercialization of four NGS products. It was the first approval for a system based on NGS technology that will allow other companies to develop their own tests using this technology. In 2014, it was the time of SOPHiA Genetics and Vela Diagnostics companies that obtained the CE-IVD Marking of the first products based on the NGS technology for clinical use.

Since then, the number of products that have the classification of IVD has been increasing; however, it is important to note that the classification of an IVD product depends on local regulations, and therefore products that are classified as IVD in a market may not have this classification in other markets. This is due to the regulatory differences between the agencies and the different requirements from each market. Anyway, it is usual that classification process of these products for clinical use must be complex and sometimes elaborated, especially in areas such as genomics. Therefore, initiatives are needed to make the approval process for these products simpler and more flexible, to make the products available, but that ensures the accuracy and usefully testing.

In 2016, the US FDA agency issued two draft guidelines: 'Use of Standards in FDA's Regulatory Oversight of Next Generation Sequencing (NGS) Based In Vitro Diagnostics (IVDs) Used for Diagnosing Germ line Diseases' and 'Use of Public Human Genetic Variant Databases to Support Clinical Validity for Next Generation Sequencing (NGS)-Based In Vitro Diagnostics'. Both are part of an initiative that aims to contribute to new testing using the NGS technology to reach the public with more speed and quality required by the market and health system.

5.2. Clinical validation

Almost all NGS approaches are still RUO, and validation is necessary before implementation as a diagnostic test. Prior clinical utility, a test must demonstrate analytical and clinical validity. Sensitivity, specificity, robustness, limits of detection, reproducibility, accuracy, precision and concordance between test results and clinical diagnosis should be analysed and measured. The test needs to evaluate patient outcomes and have positive impact on patient care [66, 130]. To assist the usage and implementation of NGS in clinical laboratories, some standards and best practice

guidelines are already available [38, 39, 44, 131–134]. Several NGS validation studies in clinical laboratories have been published and are rich sources of information [135–138]. Improvements in NGS technologies and data analysis require revalidation before implementation.

5.3. Computational infrastructure

The high volume of NGS data generated requires a complex computational infrastructure for processing, analysing and storing the data, including sophisticated data analysis pipelines. Cloud solutions such as Google, Amazon and Microsoft can be an alternative to an in-house computational infrastructure. More user-friendly bioinformatics software are desirable for non-bioinformaticians, such as Google Genomics [139], SOPHiA Genetics [140], IBM Watson [141], Illumina BaseSpace [142], Ion Reporter [143], Galaxy [144], CLC Genomics [145]. The variability of data formats generated during the analysis (e.g. FASTQ, UBAM, BAM/SAM and VCF files) and the laboratory must decide the appropriate data to be stored since the cost of managing, analysing and storing is high [124, 130, 146–149].

5.4. Genomic education

A multidisciplinary team of bioinformaticians, computational biologists, IT technicians, statisticians, molecular biologists, geneticists, genetic counsellors and clinicians is strongly needed and should be properly trained and educated for a successful implementation of NGS into routine diagnostic. Other related areas, such as lawyers, policy-makers, sales representative and investors, also need to be trained. Due to the constant updates of NGS approaches, an ongoing and continuing education about emerging technologies, software, databases and data analysis pipelines that reflect current practice is necessary. Genomic education also needs to be incorporated into medical school curriculum [148, 150].

Author details

Michele Araújo Pereira^{1*}, Frederico Scott Varella Malta¹, Maíra Cristina Menezes Freire² and Patrícia Gonçalves Pereira Couto¹

*Address all correspondence to: michele.pereira@hermespardini.com.br

1 Hermes Pardini Group/Federal University of Minas Gerais, Vespasiano, Brazil

2 Progenética Laboratory, Hermes Pardini Group, Rio de Janeiro, Brazil

References

- [1] Langreth R, Waldholz M. New era of personalized medicine: Targeting drugs for each unique genetic profile. *Oncologist*. 1999;4(5):426-427
- [2] Ginsburg GS, Willard HF. Genomic and personalized medicine: Foundations and applications. *Translational Research*. 2009;154(6):277-287

- [3] IOM (Institute of Medicine). The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary. Washington, DC: The National Academies Press; 2010
- [4] Rabbani B, Nakaoka H, Akhondzadeh S, Tekin M, Mahdieh N. Next generation sequencing: Implications in personalized medicine and pharmacogenomics. *Molecular BioSystems*. 2016;**12**(6):1818-1830
- [5] Gonzalez-Garay ML. The road from next-generation sequencing to personalized medicine. *Personalized Medicine*. 2014;**11**(5):523-544
- [6] Vogenberg FR, Isaacson Barash C, Pursel M. Personalized medicine: Part 1: Evolution and development into theranostics. *Pharmacy and Therapeutics*. 2010;**35**(10):560-576
- [7] Wong AHH, Deng CX. Precision medicine for personalized cancer therapy. *International Journal of Biological Sciences*. 2015;**11**(12):1410-1412
- [8] Scriver CR. Garrod's Croonian Lectures (1908) and the charter "Inborn Errors of Metabolism": Albinism, alkaptonuria, cystinuria, and pentosuria at age 100 in 2008. *Journal of Inherited Metabolic Disease*. 2008;**31**(5):580-598
- [9] Holley RW, Apgar J, Everett GA, Madison JT, Marquisee M, Merrill SH, et al. Structure of a ribonucleic acid. *Science*. 1965;**147**(3664):1462-1465
- [10] Sanger F, Brownlee GG, Barrell BG. A two-dimensional fractionation procedure for radioactive nucleotides. *Journal of Molecular Biology*. 1965;**13**(2):373-398
- [11] Min Jou W, Haegeman G, Ysebaert M, Fiers W. Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. *Nature*. 1972;**237**(5350):82-88
- [12] Sanger F, Donelson JE, Coulson AR, Kössel H, Fischer D. Use of DNA polymerase I primed by a synthetic oligonucleotide to determine a nucleotide sequence in phage fl DNA. *Proceedings of the National Academy of Sciences of the USA*. 1973;**70**(4):1209-1213
- [13] Padmanabhan R, Jay E, Wu R. Chemical synthesis of a primer and its use in the sequence analysis of the lysozyme gene of bacteriophage T4. *Proceedings of the National Academy of Sciences of the USA*. 1974;**71**(6):2510-2514
- [14] Sanger F, Nicklen S. DNA sequencing with chain-terminating. 1977;**74**(12):5463-5467
- [15] Smith LM, Fung S, Hunkapiller MW, Hunkapiller TJ, Hood LE. 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 Research*. 1985;**13**(7):2399-2412
- [16] Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, et al. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*. 1985;**230**(4732):1350-1354
- [17] Swerdlow H, Gesteland R. Capillary gel electrophoresis for rapid, high resolution DNA sequencing. *Nucleic Acids Research*. 1990;**18**(6):1415-1419

- [18] Wetterstrand KA. DNA Sequencing Costs. NHGRI Genome Sequencing Program (GSP) [Internet]. Available from: www.genome.gov/sequencingcostsdata [Accessed: 18 January 2017]
- [19] Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyren P. Real-time DNA sequencing using detection of pyrophosphate release. *Analytical Biochemistry*. 1996;**242**(1): 84-89
- [20] Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 2005;**437**(7057): 376-380
- [21] Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008;**456**(7218):53-59
- [22] McKernan KJ, Peckham HE, Costa GL, McLaughlin SF, Fu Y, Tsung EF, et al. Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Research*. 2009;**19**(9):1527-1541
- [23] Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. *Nature*. 2011;**475**(7356): 348-352
- [24] Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing. *Human Molecular Genetics*. 2010;**19**(R2):R227–R240
- [25] Niedringhaus TP, Milanova D, Kerby MB, Snyder MP, Barron AE. Landscape of next-generation sequencing technologies. *Analytical Chemistry*. 2011;**83**(12):4327-4341
- [26] Pareek CS, Smoczynski R, Tretyn A. Sequencing technologies and genome sequencing. *Journal of Applied Genetics*. 2011;**52**(4):413-435
- [27] Gut IG. New sequencing technologies. *Clinical and Translational Oncology*. 2013;**15**(11): 879-881
- [28] Bowers J, Mitchell J, Beer E, Buzby PR, Causey M, Efcavitch JW, et al. Virtual terminator nucleotides for next-generation DNA sequencing. *Nature Methods*. 2009;**6**(8):593-595
- [29] Levene MJ, Korlach J, Turner SW, Foquet M, Craighead HG, Webb WW. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science*. 2003;**299**(5607): 682-686
- [30] Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. *Science*. 2009;**323**(5910):133-138
- [31] Castro-Wallace SL, Chiu CY, John KK, Stahl SE, Rubins KH, McIntyre ABR, et al. Nanopore DNA sequencing and genome assembly on the international space station. *bioRxiv*. 2016;077651
- [32] Jamuar SS, Tan EC. Clinical application of next-generation sequencing for Mendelian diseases. *Human Genomics*. 2015;**9**:10

- [33] Rabbani B, Tekin M, Mahdieh N. The promise of whole-exome sequencing in medical genetics. *Journal of Human Genetics*. 2014;**59**(1):5-15
- [34] Xue Y, Ankala A, Wilcox WR, Hegde MR. Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: Single-gene, gene panel, or exome/genome sequencing. *Genetics in Medicine*. 2015;**17**(6):444-451
- [35] LaDuca H, Stuenkel AJ, Dolinsky JS, Keiles S, Tandy S, Pesaran T, et al. Utilization of multigene panels in hereditary cancer predisposition testing: Analysis of more than 2,000 patients. *Genetics in Medicine*. 2014;**16**(11):830-837
- [36] Shashi V, McConkie-Rosell A, Rosell B, Schoch K, Vellore K, McDonald M, et al. The utility of the traditional medical genetics diagnostic evaluation in the context of next-generation sequencing for undiagnosed genetic disorders. *Genetics in Medicine*. 2014;**16**(2):176-182
- [37] Tothill RW, Li J, Mileskin L, Doig K, Siganakis T, Cowin P, et al. Massively-parallel sequencing assists the diagnosis and guided treatment of cancers of unknown primary. *Journal of Pathology*. 2013;**231**(4):413-423
- [38] Matthijs G, Souche E, Alders M, Corveleyn A, Eck S, Feenstra I, et al. Guidelines for diagnostic next-generation sequencing. *European Journal of Human Genetics*. 2016;**24**(10):1515
- [39] Weiss MM, Van der Zwaag B, Jongbloed JDH, Vogel MJ, Brüggewirth HT, Lekanne Deprez RH, et al. Best practice guidelines for the use of next-generation sequencing applications in genome diagnostics: A national collaborative study of Dutch genome diagnostic laboratories. *Human Mutation*. 2013;**34**(10):1313-1321
- [40] Gonzaga-Jauregui C, Lupski JR, Gibbs RA. Human genome sequencing in health and disease. *Annual Review of Medicine*. 2012;**63**:35-61
- [41] Botstein D, Risch N. Discovering genotypes underlying human phenotypes: Past successes for Mendelian disease, future approaches for complex disease. *Nature Genetics*. 2003;**33**(Suppl):228-237
- [42] Majewski J, Schwartzentruber J, Lalonde E, Montpetit A, Jabado N. What can exome sequencing do for you? *Journal of Medical Genetics*. 2011;**48**(9):580-589
- [43] Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences of the USA*. 2009;**106**(45):19096-19101
- [44] Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Daignan JL, et al. ACMG clinical laboratory standards for next-generation sequencing. *Genetics in Medicine*. 2013;**15**(9):733-747
- [45] van El CG, Cornel MC, Borry P, Hastings RJ, Fellmann F, Hodgson SV, et al. Whole-genome sequencing in health care. Recommendations of the European Society of Human Genetics. *European Journal of Human Genetics*. 2013;**21**(Suppl 1):S1-S5

- [46] Chial H. Mendelian genetics: Patterns of inheritance and single-gene disorders. *Nature Education*. 2008;**1**(1):63
- [47] Biesecker LG, Green RC. Diagnostic clinical genome and exome sequencing. *New England Journal of Medicine*. 2014;**371**(12):1170
- [48] MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, et al. Guidelines for investigating causality of sequence variants in human disease. *Nature*. 2014;**508**(7497):469-476
- [49] Need AC, Shashi V, Hitomi Y, Schoch K, Shianna K V, McDonald MT, et al. Clinical application of exome sequencing in undiagnosed genetic conditions. *Journal of Medical Genetics*. 2012;**49**(6):353-361
- [50] Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, et al. Clinical whole-exome sequencing for the diagnosis of Mendelian disorders. *New England Journal of Medicine*. 2013;**369**(16):1502-1511
- [51] Lee H, Deignan JL, Dorrani N, Strom SP, Kantarci S, Quintero-Rivera F, et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. *Journal of the American Medical Association*. 2014;**312**(18):1880-1887
- [52] Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, et al. Molecular findings among patients referred for clinical whole-exome sequencing. *Journal of the American Medical Association*. 2014;**312**(18):1870
- [53] Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. *Nature*. 2015;**519**(7542):223-228
- [54] ACMG Board of Directors. ACMG policy statement: Updated recommendations regarding analysis and reporting of secondary findings in clinical genome-scale sequencing. *Genetics in Medicine*. 2015;**17**(1):68-69
- [55] Walsh T, Casadei S, Coats KH, Swisher E, Stray SM, Higgins J, et al. Spectrum of mutations in BRCA1, BRCA2, CHEK2, and TP53 in families at high risk of breast cancer. *Journal of the American Medical Association*. 2006;**295**(12):1379-1388
- [56] Kamps R, Brandão RD, Bosch BJ van den, Paulussen ADC, Xanthoulea S, Blok MJ, et al. Next-generation sequencing in oncology: Genetic diagnosis, risk prediction and cancer classification. *International Journal of Molecular Sciences*. 2017;**18**(2)
- [57] Leggett RM, Ramirez-Gonzalez RH, Clavijo BJ, Waite D, Davey RP. Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics. *Frontiers in Genetics*. 2013;**4**:1-5
- [58] Ekblom R, Wolf JBW. A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*. 2014;**7**(9):1026-1042
- [59] Chrystoja CC, Diamandis EP. Whole genome sequencing as a diagnostic test: Challenges and opportunities. *Clinical Chemistry*. 2014;**60**(5):724-733

- [60] Comino-Méndez I, Gracia-Aznárez FJ, Schiavi F, Landa I, Leandro-García LJ, Letón R, et al. Exome sequencing identifies MAX mutations as a cause of hereditary pheochromocytoma. *Nature Genetics*. 2011;**43**(7):663-667
- [61] Palles C, Cazier JB, Howarth KM, Domingo E, Jones AM, Broderick P, et al. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nature Genetics*. 2013;**45**(2):136-144
- [62] Snape K, Ruark E, Tarpey P, Renwick A, Turnbull C, Seal S, et al. Predisposition gene identification in common cancers by exome sequencing: Insights from familial breast cancer. *Breast Cancer Research and Treatment*. 2012;**134**(1):429-433
- [63] Fecteau H, Vogel KJ, Hanson K, Morrill-Cornelius S. The evolution of cancer risk assessment in the era of next generation sequencing. *Journal of Genetic Counseling*. 2014;**23**(4):633-639
- [64] Wang Z, Gerstein M, Snyder M. RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*. 2009;**10**(1):57-63
- [65] Zhang W, Yu Y, Hertwig F, Thierry-Mieg J, Zhang W, Thierry-Mieg D, et al. Comparison of RNA-seq and microarray-based models for clinical endpoint prediction. *Genome Biology*. 2015;**16**(1):133
- [66] Byron SA, Van Keuren-Jensen KR, Engelthaler DM, Carpten JD, Craig DW. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nature Reviews Genetics*. 2016;**17**(5):257-271
- [67] Capobianchi MR, Giombini E, Rozera G. Next-generation sequencing technology in clinical virology. *Clinical Microbiology and Infection*. 2013;**19**(1):15-22
- [68] Barzon L, Lavezzo E, Militello V, Toppo S, Palù G. Applications of next-generation sequencing technologies to diagnostic virology. *International Journal of Molecular Sciences*. 2011;**12**(12):7861-7884
- [69] Quiñones-Mateu ME, Avila S, Reyes-Teran G, Martinez MA. Deep sequencing: Becoming a critical tool in clinical virology. *Journal of Clinical Virology*. 2014;**61**(1):9-19
- [70] Radford AD, Chapman D, Dixon L, Chantrey J, Darby AC, Hall N. Application of next-generation sequencing technologies in virology. *Journal of General Virology*. 2012;**93**(Pt 9):1853-1868
- [71] Datta S, Budhauriya R, Das B, Chatterjee S, Vanlalhmua, Veer V. Next-generation sequencing in clinical virology: Discovery of new viruses. *World Journal of Virology*. 2015;**4**(3):265-276
- [72] Tang P, Chiu C. Metagenomics for the discovery of novel human viruses. *Future Microbiology*. 2010;**5**(2):177-189
- [73] Oude Munnink BB, Cotten M, Canuti M, Deijns M, Jebbink MF, van Hemert FJ, et al. A novel astrovirus-like RNA virus detected in human stool. *Virus Evolution*. 2016;**2**(1):vew005

- [74] Naccache SN, Peggs KS, Mattes FM, Phadke R, Garson JA, Grant P, et al. Diagnosis of neuroinvasive astrovirus infection in an immunocompromised adult with encephalitis by unbiased next-generation sequencing. *Clinical Infectious Diseases*. 2015;**60**(6):919-923
- [75] Parker J, Chen J. Application of next generation sequencing for the detection of human viral pathogens in clinical specimens. *Journal of Clinical Virology*. 2017;**86**:20-26
- [76] Strong MJ, Blanchard E, Lin Z, Morris CA, Baddoo M, Taylor CM, et al. A comprehensive next generation sequencing-based virome assessment in brain tissue suggests no major virus - tumor association. *Acta Neuropathologica Communications*. 2016;**4**(1):71
- [77] Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, et al. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature*. 2010;**466**(7304):334-338
- [78] Donald CL, Brennan B, Cumberworth SL, Rezelj VV, Clark JJ, Cordeiro MT, et al. Full genome sequence and sfRNA interferon antagonist activity of Zika virus from Recife, Brazil. Morrison AC, editor. *PLoS Neglected Tropical Diseases*. 2016;**10**(10):e0005048
- [79] Kuroda M, Katano H, Nakajima N, Tobiume M, Aina A, Sekizuka T, et al. Characterization of quasispecies of Pandemic 2009 influenza A virus (A/H1N1/2009) by de novo sequencing using a next-generation DNA sequencer. Jacobson S, editor. *PLoS One*. 2010;**5**(4):e10256
- [80] Dunn DT, Coughlin K, Cane PA. Genotypic resistance testing in routine clinical care. *Current Opinion in HIV and AIDS*. 2011;**6**(4):251-257
- [81] Quer J, Rodríguez-Frias F, Gregori J, Taberner D, Soria ME, García-Cehic D, et al. Deep sequencing in the management of hepatitis virus infections. *Virus Research*. 2016;pii: S0168-1702(16)30456-7
- [82] Chen X, Zou X, He J, Zheng J, Chiarella J, Kozal MJ. HIV drug resistance mutations (DRMs) detected by deep sequencing in virologic failure subjects on therapy from Hunan Province, China. Jin X, editor. *PLoS One*. 2016;**11**(2):e0149215
- [83] Lataillade M, Chiarella J, Yang R, Schnittman S, Wirtz V, Uy J, et al. Prevalence and clinical significance of HIV drug resistance mutations by ultra-deep sequencing in antiretroviral-naïve subjects in the CASTLE study. Ndhlovu LC, editor. *PLoS One*. 2010;**5**(6):e10952
- [84] Gire SK, Goba A, Andersen KG, Sealfon RSG, Park DJ, Kanneh L, et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science*. 2014;**345**(6202):1369-1372
- [85] Fischer W, Ganusov VV, Giorgi EE, Hraber PT, Keele BF, Leitner T, et al. Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing. Nixon DF, editor. *PLoS One*. 2010;**5**(8):e12303
- [86] Bull RA, Luciani F, McElroy K, Gaudieri S, Pham ST, Chopra A, et al. Sequential bottlenecks drive viral evolution in early acute hepatitis C virus infection. Ou JJ, editor. *PLoS Pathogens*. 2011;**7**(9):e1002243

- [87] Howard S, Qiu W. Viral small RNAs reveal the genomic variations of three grapevine vein clearing virus quasispecies populations. *Virus Research*. 2017;**229**:24-27
- [88] Zhang Q, Lai MM, Lou YY, Guo BH, Wang HY, Zheng XQ. Transcriptome altered by latent human cytomegalovirus infection on THP-1 cells using RNA-seq. *Gene*. 2016;**594**(1):144-150
- [89] Sijmons S, Van Ranst M, Maes P. Genomic and functional characteristics of human cytomegalovirus revealed by next-generation sequencing. *Viruses*. 2014;**6**(3):1049-1072
- [90] Chen SJ, Chen GH, Chen YH, Liu CY, Chang KP, Chang YS, et al. Characterization of Epstein-Barr virus miRNAome in nasopharyngeal carcinoma by deep sequencing. Jin DY, editor. *PLoS One*. 2010;**5**(9):e12745
- [91] Pedersen G, Kanigan T. Clinical RNA sequencing in oncology: Where are we? *Personalized Medicine*. 2016;**13**(3):209-213
- [92] Maher CA, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X, et al. Transcriptome sequencing to detect gene fusions in cancer. *Nature*. 2009;**458**(7234):97-101
- [93] Hessels D, Schalken JA. Recurrent gene fusions in prostate cancer: Their clinical implications and uses. *Current Urology Reports*. 2013;**14**(3):214-222
- [94] Mertens F, Antonescu CR, Mitelman F. Gene fusions in soft tissue tumors: Recurrent and overlapping pathogenetic themes. *Genes, Chromosomes and Cancer*. 2016;**55**(4):291-310
- [95] Heravi-Moussavi A, Anglesio MS, Cheng SWG, Senz J, Yang W, Prentice L, et al. Recurrent somatic DICER1 mutations in nonepithelial ovarian cancers. *New England Journal of Medicine*. 2012;**366**(3):234-242
- [96] Wiegand KC, Shah SP, Al-Agha OM, Zhao Y, Tse K, Zeng T, et al. ARID1A mutations in endometriosis-associated ovarian carcinomas. *New England Journal of Medicine*. 2010;**363**(16):1532-1543
- [97] Shah SP, Köbel M, Senz J, Morin RD, Clarke BA, Wiegand KC, et al. Mutation of FOXL2 in granulosa-cell tumors of the ovary. *New England Journal of Medicine*. 2009;**360**(26):2719-2729
- [98] Wartman LD. A case of me: Clinical cancer sequencing and the future of precision medicine. *Molecular Case Studies*. 2015;**1**(1):a000349
- [99] Linsley PS, Chaussabel D, Speake C. The relationship of immune cell signatures to patient survival varies within and between tumor types. Haibe-Kains B, editor. *PLoS One*. 2015;**10**(9):e0138726
- [100] Oberg JA, Glade Bender JL, Sulis ML, Pendrick D, Sireci AN, Hsiao SJ, et al. Implementation of next generation sequencing into pediatric hematology-oncology practice: Moving beyond actionable alterations. *Genome Medicine*. 2016;**8**(1):133
- [101] Molina-Vila MA, Mayo-de-las-Casas C, Giménez-Capitán A, Jordana-Ariza N, Garzón M, Balada A, et al. Liquid biopsy in non-small cell lung cancer. *Frontiers of Medicine*. 2016;**3**

- [102] Batth IS, Mitra A, Manier S, Ghobrial IM, Menter D, Kopetz S, et al. Circulating tumor markers: Harmonizing the yin and yang of CTCs and ctDNA for precision medicine. *Annals of Oncology*. 2017;1;28(3):468-477
- [103] Chakraborty C, Das S. Profiling cell-free and circulating miRNA: A clinical diagnostic tool for different cancers. *Tumor Biology*. 2016;37(5):5705-5714
- [104] Zhao Y, Song Y, Yao L, Song G, Teng C. Circulating microRNAs: Promising biomarkers involved in several cancers and other diseases. *DNA and Cell Biology*. 2017;36(2):77-94
- [105] Wang WT, Chen YQ. Circulating miRNAs in cancer: From detection to therapy. *Journal of Hematology and Oncology*. 2014;7(1):86
- [106] Zhang L, Xu Y, Jin X, Wang Z, Wu Y, Zhao D, et al. A circulating miRNA signature as a diagnostic biomarker for non-invasive early detection of breast cancer. *Breast Cancer Research and Treatment*. 2015;154(2):423-434
- [107] Brase JC, Johannes M, Schlomm T, Fälth M, Haese A, Steuber T, et al. Circulating miRNAs are correlated with tumor progression in prostate cancer. *International Journal of Cancer*. 2011;128(3):608-616
- [108] Clancy C, Joyce MR, Kerin MJ. The use of circulating microRNAs as diagnostic biomarkers in colorectal cancer. *Cancer Biomarkers*. 2015;15(2):103-113
- [109] Nandagopal L, Sonpavde G. Circulating biomarkers in bladder cancer. *Bladder Cancer*. 2016;2(4):369-379
- [110] Pimentel F, Bonilla P, Ravishankar YG, Contag A, Gopal N, LaCour S, et al. Technology in MicroRNA profiling. *Journal of Laboratory Automation*. 2015;20(5):574-588
- [111] Egger G, Liang G, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. *Nature*. 2004;429(6990):457-463
- [112] Rodenhiser D, Mann M. Epigenetics and human disease: Translating basic biology into clinical applications. *Canadian Medical Association Journal*. 2006;174(3):341-348
- [113] Skinner MK, Manikkam M, Guerrero-Bosagna C. Epigenetic transgenerational actions of environmental factors in disease etiology. *Trends in Endocrinology and Metabolism*. 2010;21(4):214-222
- [114] Strahl BD, Allis CD. The language of covalent histone modifications. *Nature*. 2000;403(6765):41-45
- [115] Nicholls RD, Saitoh S, Horsthemke B. Imprinting in Prader-Willi and Angelman syndromes. *Trends in Genetics*. 1998;14(5):194-200
- [116] Maher ER, Reik W. Beckwith-Wiedemann syndrome: Imprinting in clusters revisited. *Journal of Clinical Investigation*. 2000;105(3):247-252
- [117] Chang F, Li MM. Clinical application of amplicon-based next-generation sequencing in cancer. *Cancer Genetics*. 2013;206(12):413-419

- [118] 1000 Genome Project [Internet]. Available from: <http://www.internationalgenome.org> [Accessed: 6 March 2017]
- [119] dbSNP database [Internet]. Available from: <https://www.ncbi.nlm.nih.gov/SNP> [Accessed: 6 March 2017]
- [120] Clinvar – NCBI [Internet]. Available from: <https://www.ncbi.nlm.nih.gov/clinvar> [Accessed: 6 March 2017]
- [121] LOVD – Leiden Open Variation Database [Internet]. Available from: <http://www.lovd.nl/3.0/home> [Accessed: 6 March 2017]
- [122] The Cancer Genome Atlas (TCGA) Research Network [Internet]. Available from: <https://cancergenome.nih.gov> [Accessed: 18 April 2017]
- [123] RefSeq database [Internet]. Available from: <https://www.ncbi.nlm.nih.gov/refseq> [Accessed: 6 March 2017]
- [124] Gullapalli RR, Desai KV, Santana-Santos L, Kant JA, Becich MJ. Next generation sequencing in clinical medicine: Challenges and lessons for pathology and biomedical informatics. *Journal of Pathology Informatics*. 2012;**3**:40
- [125] Shendure J, Ji H. Next-generation DNA sequencing. *Nature Biotechnology*. 2008;**26**(10): 1135-1145
- [126] Raza K, Ahmad S. Principle, analysis, application and challenges of next-generation sequencing: A review. 2016. arXiv:1606.05254 [q-bio.GN]
- [127] Stanislaw C, Xue Y, Wilcox WR. Genetic evaluation and testing for hereditary forms of cancer in the era of next-generation sequencing. *Cancer Biology and Medicine*. 2016;**13**(1):55-67
- [128] ISO, E. (2012). 13485: 2012. Medical Devices. Quality management systems. Requirements for regulatory purposes (ISO 13485: 2003). Suomen Standardisoimisliitto SFS ry, (s 57)
- [129] Public Health Service Food and Drug Administration. Inspections, Compliance, Enforcement, and Criminal Investigations [Internet]. 2014. Available from: <http://www.fda.gov/ICECI/EnforcementActions/WarningLetters/2013/ucm376296.htm> [Accessed: 29 January 2017]
- [130] Singh RR, Luthra R, Routbort MJ, Patel KP, Medeiros LJ. Implementation of next generation sequencing in clinical molecular diagnostic laboratories: Advantages, challenges and potential. *Expert Review of Precision Medicine and Drug Development*. 2016;**1**(1):109-120
- [131] Strom SP. Current practices and guidelines for clinical next-generation sequencing oncology testing. *Cancer Biology and Medicine*. 2016;**13**(1):3-11
- [132] Gargis AS, Kalman L, Berry MW, Bick DP, Dimmock DP, Hambuch T, et al. Assuring the quality of next-generation sequencing in clinical laboratory practice. *Nature Biotechnology*. 2012;**30**(11):1033-1036

- [133] Aziz N, Zhao Q, Bry L, Driscoll DK, Funke B, Gibson JS, et al. College of American pathologists' laboratory standards for next-generation sequencing clinical tests. *Archives of Pathology and Laboratory Medicine*. 2015;**139**(4):481-493
- [134] Jennings LJ, Arcila ME, Corless C, Kamel-Reid S, Lubin IM, Pfeifer J, et al. Guidelines for validation of next-generation sequencing-based oncology panels: A joint consensus recommendation of the association for molecular pathology and college of American pathologists. *Journal of Molecular Diagnostics*. 2017;**19**(3):341-365
- [135] Lin MT, Mosier SL, Thiess M, Beierl KF, Debeljak M, Tseng LH, et al. Clinical validation of KRAS, BRAF, and EGFR mutation detection using next-generation sequencing. *American Journal of Clinical Pathology*. 2014;**141**(6):856-866
- [136] Singh RR, Patel KP, Routbort MJ, Reddy NG, Barkoh BA, Handal B, et al. Clinical validation of a next-generation sequencing screen for mutational hotspots in 46 cancer-related genes. *Journal of Molecular Diagnostics*. 2013;**15**(5):607-622
- [137] Rathi V, Wright G, Constantin D, Chang S, Pham H, Jones K, et al. Clinical validation of the 50 gene AmpliSeq™ Cancer Panel V2 for use on a next generation sequencing platform using formalin fixed, paraffin embedded and fine needle aspiration tumour specimens. *Pathology*. 2017;**49**(1):75-82
- [138] Wallace AJ. New challenges for BRCA testing: A view from the diagnostic laboratory. *European Journal of Human Genetics*. 2016;**24**(S1):10-18
- [139] Google Genomics [Internet]. Available from: <https://cloud.google.com/genomics> [Accessed: 16 April 2017]
- [140] SOPHiA Genetics [Internet]. Available from: <http://www.sophiagenetics.com> [Accessed: 16 April 2017]
- [141] IBM Watson [Internet]. Available from: <https://www.ibm.com/watson> [Accessed: 16 April 2017]
- [142] Illumina BaseSpace [Internet]. Available from: <https://basespace.illumina.com> [Accessed: 16 April 2017]
- [143] Ion Reporter [Internet]. Available from: <https://ionreporter.thermofisher.com/ir> [Accessed: 16 April 2017]
- [144] Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, et al. The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Research*. 2016;**44**(W1):W3–W10
- [145] CLC Genomics [Internet]. Available from: <https://www.qiagenbioinformatics.com/products/clc-genomics-workbench> [Accessed: 16 April 2017]
- [146] Desai A, Jere A. Next-generation sequencing: Ready for the clinics? *Clinical Genetics*. 2012;**81**(6):503-510
- [147] Sboner A, Mu X, Greenbaum D, Auerbach RK, Gerstein MB. The real cost of sequencing: Higher than you think! *Genome Biology*. 2011;**12**(8):125

- [148] Rizzo JM, Buck MJ. Key principles and clinical applications of "next-generation" DNA sequencing. *Cancer Prevention Research (Philadelphia)*. 2012; **5**(7):887-900
- [149] Xuan J, Yu Y, Qing T, Guo L, Shi L. Next-generation sequencing in the clinic: Promises and challenges. *Cancer Letters*. 2013;**340**(2):284-295
- [150] Schrijver I, Aziz N, Farkas DH, Furtado M, Gonzalez AF, Greiner TC, et al. Opportunities and challenges associated with clinical diagnostic genome sequencing. *Journal of Molecular Diagnostic*. 2012;**14**(6):525-540

IntechOpen