The applications of massive parallel sequencing (next-generation sequencing) in research and molecular diagnosis of human genetic diseases

Hieu T. Nguyen[1*], Huong T.T. Le[1], Liem T. Nguyen[1], Hua Lou[2], Thomas LaFramboise[2]

[1]Vinmec Research Institute of Stem Cell and Gene Technology, Vinmec International Hospitals, Hanoi, Vietnam
[2]Department of Genetics and Genome Sciences, School of Medicine, Case Western Reserve University, Cleveland, Ohio, USA

Received 2 February 2018; accepted 22 May 2018

Abstract:

Next-generation sequencing (NGS) is a high throughput sequencing technology, which has revolutionized both basic and clinical research of the human genetic disorders. This technology is also called massively parallel sequencing (MPS) due to its ability to generate a huge amount of output data in a cost- and time-effective manner. NGS is widely utilized for different sequencing applications such as targeted sequencing (a group of candidate genes), exome sequencing (all coding regions), and whole genome sequencing (the entire human genome). With NGS, a variety of genomic aberrations can be screened simultaneously such as common and rare variants, structural variations (amplifications and deletion), copy-number variation, and fusion transcripts. NGS technologies combined with advanced bioinformatic analysis have tremendously expanded our knowledge. On the one hand, the basic research area involves direct use of NGS to identify novel variations and determine human disease mechanisms. On the other hand, clinical research is being advanced by high-throughput genetic tests with high resolution and clinically relevant genetic information for molecular diagnoses of human disorders. In this communication, we introduce NGS technologies and review a few key areas where NGS has made a significant impact, with an emphasis on the application of NGS to the identification of the molecular bases of human genetic diseases.

Keywords: ***human genetic diseases, massively parallel sequencing, molecular diagnosis, next-generation sequencing.***

Classification numbers: ***3.2, 3.5***

Introduction

For more than four decades, Sanger sequencing based on the dideoxy chain termination principle has been considered the gold standard method for determining a DNA sequence and the identification of genomic variations to support the diagnosis of genetic diseases [1]. For monogenic diseases with clear clinical and biochemical presentations, and well characterized mutation landscapes, sequencing the target regions by the Sanger method is an accurate and cost-effective way to obtain a conclusive molecular diagnosis. Nevertheless, as most inherited diseases are often genetically and clinically heterogeneous, the selection of candidate gene(s) and/or gene region(s) for sequence analysis is costly, laborious, and time consuming, which often delays diagnosis and treatment, causing anxiety for patients and their families. Many neurological disorders such as ataxias, epilepsy, and migraines are caused by mutations in one of many genes. For example, 65 genes were shown to be responsible for Retinitis Pigmentosa, one common form of hereditary retinal degeneration, demonstrating its high heterogeneity and diversity of inheritance patterns. The diagnosis of mitochondrial diseases is another demonstration for an extreme situation, for which clinical phenotypes significantly overlap and heterogeneous mutations span more than 1,300 genes [2-5].

The number of recognized polygenic conditions has greatly increased due to the rapid discovery of new genes, genetic conditions, and phenotypic ranges. Thus, the traditional step-wise molecular diagnostic approach of single genes or candidate genes is no longer adequate to identify the molecular etiologies of diseases. Additionally, amplicon-based Sanger sequencing is notorious for allele

*Corresponding author: Email:htn13@case.edu

dropout due to either Single Nucleotide Polymorphism (SNPs) at the PCR primer sites and large deletions including one or both of the primer sites [6]. Moreover, the complexity of genomic research and application including diagnosis of genetic diseases demands a depth of information beyond the capacity of traditional DNA sequencing technologies. The need to address these drawbacks has spurred the birth of a new NGS approach, which is more comprehensive, accurate, and effective. Massive parallel sequencing technologies (MPS or NGS) have enabled sequencing of many, usually short fragments of nucleic acid at the same time to provide deep sequencing coverage of individual samples or indexing of multiple samples.

During the past decade, NGS has revolutionized nearly every area of biological sciences by generating the enormous genetic information for the identification of genomic variations, disease mechanisms, and disease-associated markers, which has led to the development of better diagnostics tools and treatment therapies. High-throughput sequencing including (1) targeted sequencing (genes of interest), (2) whole exome sequencing (protein-coding portions), and (3) whole genome sequencing (the entire human genome) allows the detection of mutations in multiple genes in a cost-time effective fashion [7-13].

A number of research studies have successfully utilized the NGS technology to identify genes related to diseases [14], causative mutations [15] and epigenetic modulations correlated with particular disorders [16, 17]. NGS approaches have been also applied to the molecular diagnosis of genetic diseases, particularly complex disorders with heterogeneous clinical phenotypes and various underlying genetic causes [18-21]. Generally, sequencing more than 3 billion base pairs of the whole human genome is economically unfeasible, computationally challenging and technically demanding. Thus, in clinical setting, it is frequently desirable to capture or enrich genes demonstrated to be important for a particular clinical phenotype, followed by NGS. Here, we review NGS technologies and summarize the recent applications of NGS in both basic and clinical research with a focus on the molecular diagnosis of human genetic diseases.

Overview of NGS technologies

To date, there have been several generations of sequencing technologies, which are different regarding sequencing principles, sequencing chemistries, and instrumentation (Fig. 1).

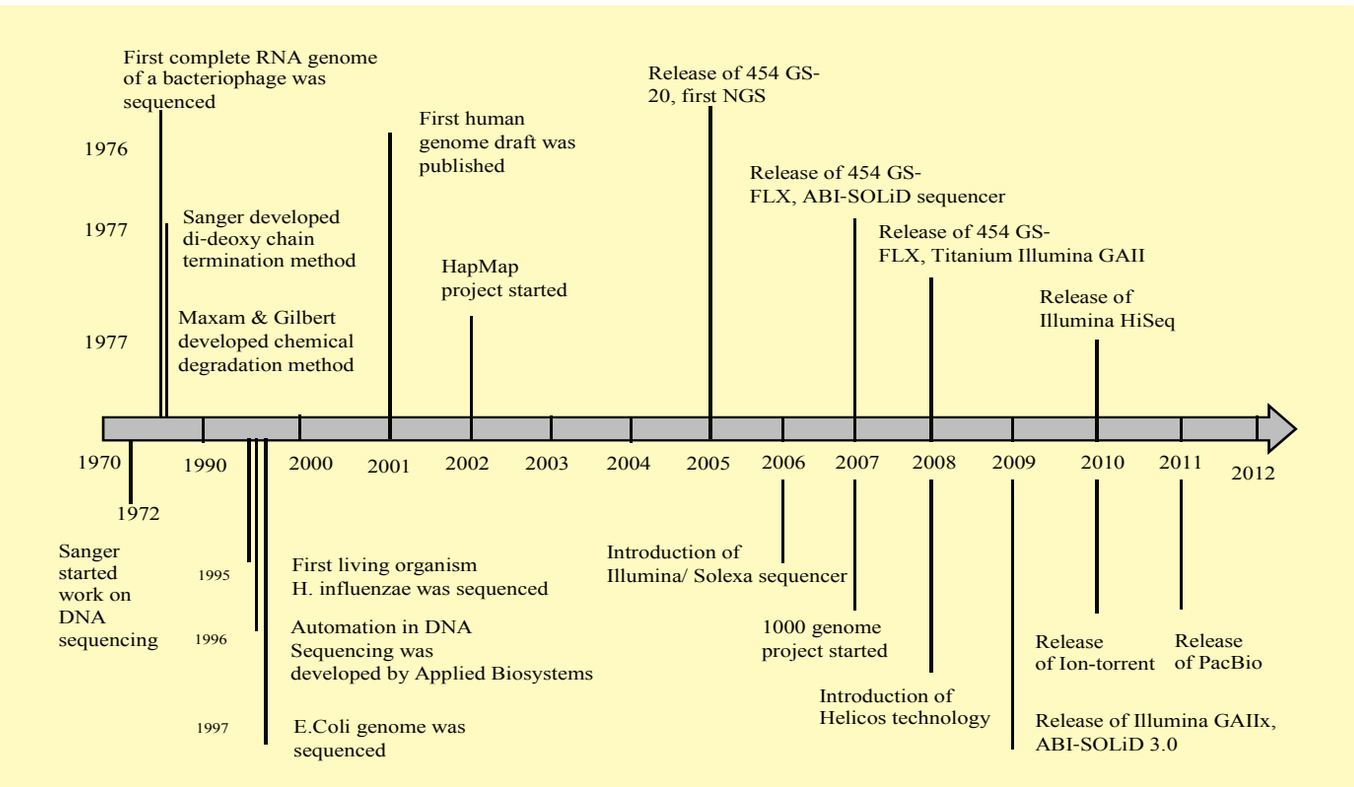

Fig. 1. Timeline of introduction of DNA sequencing technologies and platforms [22].

The first generation of sequencing was defined by the Sanger and Maxam-Gilbert techniques, which are capable of sequencing a few hundred base pairs at a time, and could be used for single gene sequencing [1].

NGS technologies, also called second-generation sequencers, was first introduced to the scientific community in 2005, over 30 years after Sanger sequencing was introduced. The major advancement of second-generation sequencing is its capability to produce sequencing data in a massively parallel manner, thus generating huge amounts of data in a cost- and time-effective fashion. Next generation sequencers are featured by several platforms that produce the large scale of sequencing data (output data size up to gigabases), including Roche 454, Illumina Solexa, and ABI-SOLiD technologies. These technologies differ in their sequencing principles; specifically, Illumina's sequencing by DNA synthesis (Sequencing By Synthesize - SBS), Roche 454's sequencing by pyrosequencing, and ABI SOLiD's sequencing by oligonucleotide ligation (Sequencing by ligation - SBL) [23-26]. Notably, although Roche's 454 was the first commercial NGS platform appeared on the market, it is no longer available, which is indicative of rapid advancement of the field.

Third generation sequencing (or next-NGS) was developed with the purpose of making sequencing cheaper than second-generation sequencing. Third generation sequencers utilize technologies that interrogate single molecules of DNA without amplifying them through PCR, thereby overcoming problems of PCR amplification biases and de-phasing. These sequencers include Helicos Helioscope (Helicos) based on single molecule sequencing [27], which went bankrupt in 2012, and Pacific Bioscience (PacBio), a single molecular real-time (SMRT) instrument [28, 29].

The Ion Torrent, a smaller scale sequencer, could be placed between the second and third generation as it is not a single-molecule sequencing technique and the sequencing detection is not based on fluorescence signal. Semiconductor is the sequencing basis for Ion Torrent, allowing the detection of protons (H^+) generated by enzymatic reactions [30].

Complete Genomics, Oxford Nanopore, and Plonator use different sequencing principles and chemistries, which do not belong to second or third generation and could be placed under fourth generation. No matter what the sequencing chemistries and different company platforms are, these technologies share the same principle, which is to simultaneously sequence an enormous amount of separated genomic regions [7, 29, 31] (Table 1).

Table 1. Comparison of important NGS platforms.

	Roche 454 GS FLX Plus	Illumina Solexa HiSeq200	ABI SOLiD 5500xl	Ion Torrent	Pacific Bio	Helicos
Sequencing methods	Pyrosequencing	Reversible Dye Terminators	Sequencing by ligation	H^+ Detection	ZMW-Single molecule	Heliscope-Single molecule
Read lengths	700-1,000 bp	2 x100 bp	75 bp	200 bp	Average 1,200, up to 3 kb	25-55 (average 35 bp)
Sequencing runtime	23 hrs	11-14 days	6-7 days	2 hrs	Less than a day	Less than a day
Data generated	700 MB/run	600 GB/run	10-15 GB/day	10-1,000 MB (depends upon chip used)	70-140 MB/ cell	21-35 GB/run
Advantages	- Longer read length - Small data files	Huge data	High-quality data	- Low cost - Very fast	- Longer read than 454 - Fast	- Big data among single molecule synthesis
Concerns	- Less data - Homopolymer	- Short reads - Dephasing - Long run time	- Short reads - Long run time	- Less data - Small read	- Random indel errors	- Small reads - Higher raw error rate

Regardless of sequencing platforms and principles, the application of NGS to research and clinical diagnosis comes in several different scales, which are based on coverage depth. In most NGS experiments, the genome (either the whole genome or targeted "panel" of genes) is fragmented into short fragments of a few hundred base pairs. These fragments are individually read and aligned to generate longer contiguous sequences computationally. In order to get significant redundancy, each individual nucleotide needs to be read several times. The number of times that a given nucleotide in the genome has been read in an experiment is indicated as sequencing depth (also known as read depth). Regarding coverage, there are two concepts that need to be clarified. First, the "breath of coverage" concept is often understood as a measure of what proportion of the total intended genome is represented in the data set. Second, the "depth of coverage" concept can be used to describe the average raw or aligned read depth. The coverage depth varies, depending on the size of the targeted region and the application goal. Shifting the focus from a single large gene to a group of genes, to the whole exome (~20,000 genes), and ultimately, to the whole genome, increases the complexity but decrease the read depth coverage and the ability to call copy number variations (CNVs). When designing an NGS experiment to investigate a clinical question or questions, understanding of depth and coverage concepts can help in tailoring the experimental design and bioinformatics tools to obtain the most meaningful data.

Currently, Illumina NGS platforms are the most commonly used tools for NGS-based basic and clinical research of genetic diseases. Therefore, in this review, we will focus on Illumina NGS system for our discussion on the applications NGS in research and the molecular diagnosis of human genetic diseases. The typical Illumina sequencing workflow from sample collection to NGS analysis contains several steps (Fig. 2): DNA extraction, DNA fragmentation, target sequence enrichment, library construction and sample indexing, loading onto the sequencer for cluster generation and sequencing. The sequence images are subsequently converted to base calls followed by filtering for high-quality base calls, sequence alignment, data analysis and variant calling, and finally interpretation and reporting.

In clinical NGS, quality control procedures must be incorporated to monitor the performance of each step and to ensure that the final results are accurately and appropriately interpreted according to each patient's clinical presentation. The sequence analyses consist of three major steps. The

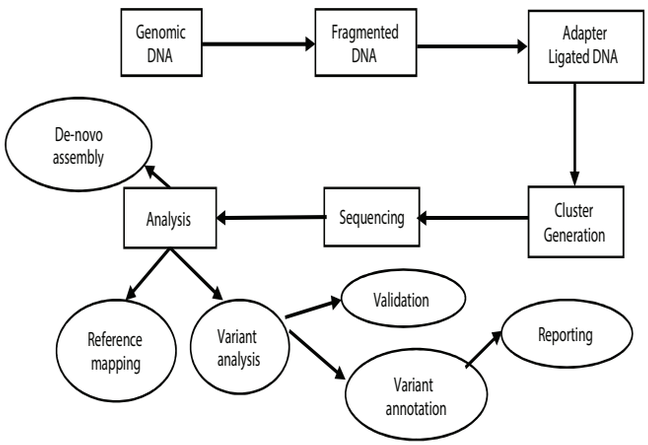

Fig. 2. Basic scheme of a next generation experiment using an Illumina sequencing platform.

primary analysis involves the image capture, the conversion of the image to base calls, and the assignment of quality scores to base calls. The secondary analysis is the filtering of reads based on quality followed by alignment and/or assembly of the reads. Finally, the tertiary analysis involves variant calls based on a reference sequence, variant annotation, data interpretation, and result reporting. Quality control at each step is required because an NGS experiment often involves a large number of samples, a complex workflow and bioinformatic pipeline, and a high reagent cost.

NGS is being applied to identify (causal) genetic variants associated with a genetic disease or phenotype under many different methods such as whole-genome sequencing (WGS), whole-exome sequencing (WES), methylome sequencing, transcriptome sequencing, and targeted sequencing (Fig. 3). While WGS allows sequencing of the entire patients' genomes, WES focuses on the coding regions (exons) of a genome, which take up about 2% of the human genome. However, WES is not suitable to identify most CNVs and other structural modifications. Besides DNA, NGS can also be applied to determine levels of gene expression (transcriptome sequencing or RNA-Seq), splice variants, gene fusions, genomic rearrangements, allele-specific expression, posttranscriptional modifications, microRNAs, small and long noncoding RNAs. Methylome sequencing focuses on DNA methylation. Finally, targeted sequencing, focusing on a selection of genes of interest for a specific disease, is a great choice regarding time and cost for clinical applications of NGS.

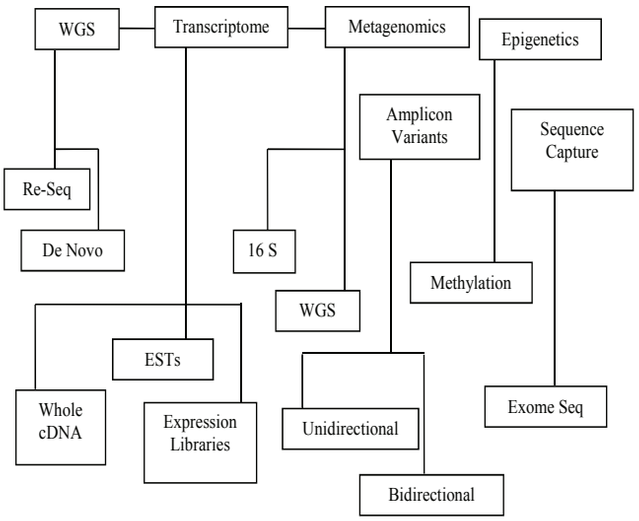

Fig. 3. Schematic diagram of different NGS applications and sequencing methods [22].

Advantages of NGS technologies

Comparing with the first generation sequencing method, NGS technology has the most apparent advantage, which is the capability to massively parallel sequence the genome to obtain high throughput output (up to millions of DNA fragments) on each run. The capability of MPS of NGS has constantly been improved by the development of both sequencing technologies and wet bench. Specifically, the revolutionary development of clonal DNA fragment amplification techniques and the sequence reading technologies together with the improvements in the wet bench portion such as target capture methods enable NGS to sequence the whole genome or particular areas of interest with deep coverage (Table 2).

In the context of clinical applications, NGS technologies offer a number of advantages over the traditional sequencing methods as shown in Table 2. Specifically,

(1) MPS allows sequencing of a group of biomarkers from multiple samples in each run. It is often desirable to simultaneously process many samples to minimize waiting time for results for patients.

(2) Each patient can be simultaneously screened for various genomic aberrations such as single nucleotide and multi-nucleotide polymorphisms (SNPs), insertions and deletions (indels), copy number variations (CNVs), and gene/transcript fusions. Simultaneous screening enables consolidating multiple tests into one MPS run; therefore, lowering the overall healthcare costs and patient sample requirement as compared to the low- and medium-throughput tests.

(3) NGS provides high sequencing depth and coverage for DNA fragments of interest (over 100X), offering sensitive detection (limit-of-detection) and a high level of confidence.

(4) In particular, it is possible to relatively quantitate the allelic fraction of a mutation by estimating the number of DNA strands harboring the genetic alterations and abnormalities among the total sequencing reads, leading to better understanding of the pathogenicity of the tested sample.

Application of NGS

Cancer

As cancer is a genetic disease caused by heritable or somatic mutations, application of NGS has revolutionized

Table 2. Comparison of clinical sequencing technologies [32].

Sequencing technology	DNA input	Throughput of Sequencing output	Multiplexing ability	Types of detected mutations	Workflow and interpretation complexity	Quantitative
Sanger	High	Low	None	Point mutations, Insertions and deletions	Low/ Intermediate	No
Pyrosequencing	Intermediate	Low/ Intermediate	None	Point mutations	Intermediate	Yes
NGS	Low	High	Yes (high)	Point mutations, insertions, deletions, gene expression, fusion, and copy number variations	High	Yes

the field. Cancer research has traditionally been complicated by the fact that there is no clear-cut mechanism for all types of cancers; therefore, there is an urgent need to analyze a large number of genetic variations in the human genome that could be responsible for cancer phenotypes. A large number of cancer cases need to be compared to healthy individuals regarding their genetic make-up, particularly focusing on several genetic targets. This area has been substantially aided by the application of NGS, as many genomes can be sequenced in a short amount of time. In the context of bench-to-bedside applications, NGS has contributed greatly to commercially available gene panels for cancer screening, diagnosis, prognosis and pharmacogenesis. For instances, Extended RAS Panel, an FDA-approved NGS kit, helps clinicians identify colorectal cancer patients eligible for Vectibix treatment [33]. Vectibix is the first FDA-approved monoclonal anti-epidermal growth factor receptor antibody for first-line treatment for patients with wild-type RAS metastatic colorectal cancer (mCRC). NGS targeted panel approach enables simultaneous interrogation of 56 variants across the K-RAS and N-RAS genes to determine the mutant status of RAS genes in a single test. Data generated by the NGS RAS Panel help identify mCRC patients with wild type RAS genes who will be treated with Vectibix. The Extended RAS Panel highlights the importance of NGS-based biomarker screening in therapeutic decision-making in cancer treatment planning.

NGS is empowering the worldwide collaborations for cataloging the mutations and genomic landscapes in multiple cancer types such as The Cancer Genome Atlas (TCGA) [34] and the International Cancer Genome Consortium (ICGC) [35]. These large-scale projects aimed at generating high-quality genomic sequences for a large number of tumors from various types and subtypes of cancer. The massive amount of data generated by TCGA and the ICGC will help us refine cancer classification systems and interrogate the interplay between DNA alternations, RNA expression changes, and epigenomic landscapes in order to gain a comprehensive overall picture of cancer genomics, thereby assisting in discoveries related to diagnostics, prognostics, and therapeutics. For instance, WES and WGS studies have identified new high- and moderate-risk genes in different types of cancers, such as the pancreatic cancer susceptibility genes PALB2 and ATM [36], and the hereditary colorectal cancer moderate-risk genes POLD1 and POLE [37].

In addition to DNA sequencing, RNA sequencing is also used to sequence non-coding RNAs like microRNAs (miRNAs) and long non-coding RNAs (lncRNAs), which have significant functions in cancer pathogenesis and have been demonstrated to be ubiquitously dysregulated in tumorigenesis [38]. Also, epigenetic modifications, particularly DNA methylation, are well-documented and well-studied in some cancers [31]. For example, using DNA methylation and miRNA profiles, a recent study reports that DNA methylation contributes to deregulation of 12 cancer-associated miRNAs and breast cancer progression [39]. The authors also found a strong association between hypermethylation of MIR-127 and MIR-125b-1 and breast cancer progression, particularly metastasis, and concluded that MIR-127 and MIR-125b-1 hypermethylation could be potential biomarkers of breast cancer metastasis [39].

Many NGS-based studies have been conducted to identify novel genetic alterations leading to oncogenesis, metastasis and cancer progression and to survey tumor complexity and heterogeneity [34]. These efforts have provided significant achievements for many diseases such as melanoma, acute myeloid leukemia, breast, lung, liver, kidney, ovarian, colorectal, head and neck cancers [34].

In the past few years, NGS technology has been applied to provide a comprehensive molecular diagnosis of cancers [40]. NGS technology enables the simultaneous sequencing of a large number of target genes and provides early detection and diagnostic markers to develop NGS-based cancer molecular diagnosis [41-43].

WES is currently the most commonly used in clinical diagnostics because it covers more than 95% of the exons, which contain 85% of disease-causing mutations [44]. Moreover, WES has also been applied for determining somatic mutations in tumors [44]. WGS can be utilized to monitor cancer progression, treatment efficacy, and the molecular mechanisms underlying resistance development. However, WGS is expensive and computationally burdensome because of the enormous amount of output data. Indeed, targeted cancer panels are currently most commonly used as diagnostic and prognostic tools in clinical oncology due to their advantages such as low cost and relatively simple interpretation [45, 46].

Breast cancer (BC) is a good example of the application of NGS as an effective method to increase the detection rate of high-risk cases [47]. Previous studies have shown that BRCA1 and BRCA2 mutations cause about 30% of BC cases. A genetic test using BRCA1 and BRCA2 mutations has been recommended; however, mutations in other genes

such as ATM, CHEK2, PALB2, and TP53, have also been shown to confer high BC risk [43]. Therefore, a multiple-gene sequencing panel was developed using NGS, which contained 68 genes including BRCA1, BRCA2, ATM, and TP53. The genes in this panel had cancer risk association for patients with early-onset or familial breast cancer. Currently, the approach of targeted sequencing holds great potential for the rapid diagnosis of not only breast cancer but also other kinds of cancers.

Mendelian and rare diseases

The mendelian or monogenic disease is caused by a mutation at a single gene locus. The location of a single gene could be on an autosome or a sex chromosome, and its inheritance could be either in a dominant or a recessive or an X-linked fashion. There are a number of reports for the use of NGS in identifying the causal variants and in the diagnosis of genetic disorders.

Miller syndrome is the first rare Mendelian disorder whose causal mutations were identified by WES. This syndrome mainly affects the development of the face and limbs. The authors described dihydroorotate dehydrogenase (*DHODH*) mutations in 3 affected families following filtering against public single nucleotide polymorphism (SNP) databases and eight haplotype map (HapMap) exomes [48] U. To date, WES and WGS have identified over 100 genes responsible for various Mendelian diseases; some examples are listed in Table 3.

In the last few years, technological advancements in NGS, especially target enrichment methods, have led to the identification of genetic variations responsible for more than 40 rare disorders. NGS facilitates researchers with the required capacity to analyze large panels of genes for suspected genetic diseases. These diseases vary from single gene disorders such as Neurofibromatosis Type 1 (NF1), Marfan syndrome (MFS), and spastic paraplegia [50, 53, 54] to diseases caused by a group of related genes such as hypertrophic cardiomyopathy and congenital disorders of glycosylation (CDG) [19, 20, 51]. NGS has also been applied to multi-gene disorders including X-linked intellectual disability (XLID) [18] and retinitis pigmentosa [52], as well as defined disorders without identified genetic causes [55-57] (Table 4).

Cystic fibrosis (CF) was the first disease for which the FDA approved an NGS assay for in vitro diagnostic use [62]. It is a Mendelian autosomal recessive disorder that affects the lungs and digestive system of about 70,000 people worldwide. There is no way to prevent CF; therefore, the best defense against this disease is early diagnosis. NGS offers a complete, accurate, and comprehensive interrogation into the whole cystic fibrosis gene for increased clarity in molecular CF testing. NGS-based CF molecular tests enable earlier detection in affected individuals and selection of optimized therapies. Besides diagnosis, NGS-based CF molecular tests can be applied for population screening to determine CF carrier status, newborn screening for CF, and

Table 3. Several publications on the application of WES and WGS to clinical practice [49].

Gene/disease	Enrichment method	Sequencing platform/ chemistry	Sample	Average coverage	References
Miller syndrome (WES)	Agilent array-based capture	Genomic Analyzer (GAII)/76 base read, Single-End (SBS)	Three kindreds	40X	[50]
Kabuki syndrome (WES)	Agilent array-based capture	GAII/Single End or Pair End (SBS)	Ten unknown	40X	[51]
Inflammatory Bowel disease (WES)	NimbleGen exome array-based capture	GS-FLX (SBS, pyrosequencing)	1 patient	34X	[52]
Charcot-Marie-Tooth (WGS)	Direct genomic DNA	SOLiD (SBL)	Family members	30X	[9]
Dopa-responsive dystonia (WGS)	Direct genomic DNA	SOLiD4 (SBL)	Twins and family members	30X	[11]

Table 4. Examples of publications on the application of NGS targeted sequencing to clinical practice [49].

Disease	Gene	Enrichment method	Sequencing platform/ chemistry	Sample	Average coverage	References
Neurofibromatosis Type I (autosomal dominant)	NF1 is a large gene with many exons	NimbleGen oligo array capture	GS-FLX (SBS, pyrosequencing)	2 known	>30X	[53]
Marfan Syndrome (autosomal dominant)	FBN1 is a large gene with many exons	Multiplex PCR	GS-FLX (SBS, pyrosequencing)	5 known 87 unknown	~174X	[54]
Hereditary Spastic Paralegias (HSP: A group of inherited neurodegenerative disorders)	SPG5 and SPG7 genes are involved in the autosomal recessive form of HSP	Fluidigm	GS-FLX (SBS, pyrosequencing)	187 patients	72X for run 1 25X for run 2	[50]
Dilated Cardiomyopathy (DCM) (a group of genetically heterogeneous disorders)	Panel of 19 genes known to cause DCM	Pooled PCR amplicons	GAII (SBS)	5 known	~50X	[19]
Congenital Disorder of Glycosylation (CDG) (a group of diseases caused by over 30 genes involved in the N-linked glycosylation)	Panel of 24 genes known to cause CDG	Fluidigm Raindance	SOLiD version 3/50 base read, SE (SBL)	12 known	616X (FD) 455X (RD)	[36]
Retinitis Pigmentosa (RP) (a group of diseases caused by over 40 known genes)	Panel of 45 genes known to cause RP	NimbleGen oligo array capture	GAII/32 base read, SE (SBS)	2 known 3 unknown	486X (1 sample per lane) 98X (4 samples per lane)	[58]
X-Linked Intellectual Disability (XLMR) (a group of genetically heterogeneous disorders)	Panel of 86 genes known to cause XLMR	Raindance	GAII (SBS)	3 known 21 unknown	Coverage per base ranging from 92X to 445X	[18]
Mitochondrial diseases	Mitochondrial DNA (mtDNA)	2 overlapping PCR fragments	GAII (SBS)	2 known	~1,785x	[59]
Mitochondrial diseases	Panel of 362 nuclear genes are known to involve in mitochondrial diseases.	Agilent array based capture	GAII/36 base read, SE (SBS)	2 patients 1 normal	37X-51X for nuclear genes, 3,000-5,000X for mtDNA	[21]
Human Leukocyte Antigen (HLA) genotyping	HLA genes	HLA gene amplification	MiSeq/250 base read, PE (SBS)	211 known 79 unknown	>67X	[60]
Ataxias	A panel of 58 genes known to cause human ataxia	Agilent SureSelect targeted capture	Illumina/51 base read, PE (SBS)	50 patients	94% of regions of interest > 5X	[61]

genetic counseling regarding couples' reproductive risks and family planning options.

Pre-natal diagnosis

Traditionally, molecular prenatal diagnosis requires invasive methods to draw an amniocentesis or chorionic villus sample and detect chromosomal abnormalities. Besides cost, these procedures pose a miscarriage risk at an approximate rate of 0.5%. Therefore, it is highly desirable to develop a non-invasive method for prenatal diagnosis to avoid the risk of fetal loss.

One of the most valuable applications of NGS technology is molecular genetic testing in pre-diagnostics. The pioneering work of Denis Lo and his coworkers at The Chinese University of Hong Kong [63] demonstrated that more than 10% of a mother's cell-free DNA is from the fetal genome at the end of the first trimester. Recently, there has been rapid progress in applying NGS technology to the detection of fetal chromosomal abnormality in fetal DNA from cell-free DNA fragments in maternal plasma.

In 2011, three large-scale studies involving multiple centers established non-invasive prenatal tests (NIPTs) that have had a significant impact on prenatal care [64-66]. These studies showed that the detection of fetal trisomy 21 could be performed at nearly 100% sensitivity and 98% specificity by multiplexed MPS of maternal plasma DNA. Since its introduction in 2011, NIPT has been standardized as a recommended test for high-risk pregnancies [1]. NIPT has also evolved from exclusively trisomy 21 testing to include trisomy 18, trisomy 13, sex chromosome aneuploidies, and microdeletions. In 2016, one clinical validation study demonstrated that genome-wide NIPT could provide high resolution, sensitive, and specific detection of a wide range of fetal subchromosomal and whole chromosomal abnormalities that were previously only pinpointed by invasive karyotyping testing [67]. In some cases, this NIPT also provided further information about the origin of genetic material that had not been identified by the invasive karyotype method. Therefore, the implementation of the NGS-based prenatal screening of fetal chromosome abnormalities using circulating cell-free nucleic acid in the maternal blood is one of the great advancements in providing effective and safe prenatal diagnostics.

Besides screening for chromosomal abnormalities, NGS technology can also be applied to the prenatal mutation detection of genetic disorders. A proof-of-concept NGS-based study to detect fetal β-thalassemia mutations using maternal blood demonstrated the possibility of investigating specific genetic disease loci [68]. In this study, NGS enables sequencing of fetal DNA fragments that could subsequently be assembled into a complete fetal genomic map with the parental genomes as guides. Then, the fetal genome could then be screened for mutations prenatally and noninvasively. This approach was applied to identify whether the fetus carries β-thalassemia mutations in the case study of a family where the pregnant mother carried one gene mutation, and the father carried a different mutation for the blood disease β-thalassemia. From the maternal plasma DNA sequencing data, they found that the fetus inherited the paternal mutation. Then, they used relative haplotype dosage analysis to test if the fetus had inherited the genomic region that contained the maternal mutation. The authors found that the fetus had not inherited the maternal mutation; therefore, the fetus was a heterozygous carrier for β-thalassemia. This is one of the pioneering studies showing that sequencing of maternal plasma cell-free DNA provides noninvasive prenatal genome-wide scanning for genetic disorders [68].

The current global status of using NGS in disease management

The rapid development of MPS has opened up the opportunities to turn scientific discoveries about DNA and the way it works into a promising life-saving reality for patients worldwide. A clear example of this global influence of NGS in disease management is the launching of several massively sequencing projects for precision medicine in developed countries. These projects are 100K Genome project in the UK, Precision Medicine Initiative in USA, Japan, India, and Middle East, all aiming at bringing the benefits of genomics to patients. Precision medicine is an emerging approach for disease treatment and prevention, in which health care and medical decisions, practices, and products should be individually tailored to each patient's variability in genes, environment, and lifestyle.

Currently, cancer and rare diseases, which are strongly linked to changes in the genome, are the primary focus for precision medicine. In case of cancer, DNA from the tumor and DNA from the patient's healthy cells, thanks to NGS, can be sequenced and compared; the precise gene changes

are detected. Understanding these genomic alterations is crucial to predict how well a person will respond to a particular treatment such as radiotherapy, or indicate which treatment will be the best for individual patients. An excellent example in use already is to prescribe the medicine tailored to a woman's breast cancer genotyping. Herceptin will be effective for a woman with HER2 positive but not for the one who is HER2 negative. Additionally, genomics coupled with NGS can also be used to track infectious disease, precisely pinpointing the origin and nature of the outbreak through examining the whole genomes of infectious agents.

Future perspectives

The advent of NGS has opened up many new frontiers and it, in the future, will continue to play a crucial role in the research and molecular diagnostics of genetic diseases. NGS will keep providing novel insights into disease mechanisms, metabolics, and signaling pathways at a resolution never previously possible. Information obtained from NGS is being used to improve diagnostics, and to develop more effective and more personalized treatments for disease and patient care. Furthermore, targeted NGS will still hold great potential for speed and cost-effectiveness of sequencing by focusing on the portions of the genome that are relevant to the question of the study. It is also beneficial in identifying and developing panels for biomarkers associated with a particular type of health condition.

NGS technologies are capable of helping scientists and clinicians study genomes of individuals faster than ever before, opening up the new era of Personalized Medicine. Every individual is different in their genetic make-up and is susceptible to different diseases, infections, and disorders. Therefore, knowing one's genomic sequence will help determine accurate and proper care and will elucidate increased risks for hereditary diseases. With decreasing costs and rapid developments of NGS technologies, it is easy to envision that all patients will soon have their genomes sequenced when they visit their doctors. The information generated by NGS can provide information on the different types of disease-causing alterations in individuals in the short turn-around time required to screen patients either for clinical trials, or for diagnostics in clinical settings.

In the future, sequencing of individual genomes of interest under different living, nutritional, or treatment conditions will benefit the medical communities by guiding disease control, progression, and prevention, and rational usage of molecularly guided treatments. These discoveries will ultimately bring a better understanding of disease pathogenesis, contributing to a new era of molecular pathology and personalized medicine. With the knowledge of precision medicine, we can increase treatment efficacy, reduce toxicity, and therefore decrease disease burden for both patients and society.

Finally, to better understand the genetic etiology of diseases, to improve effective molecular diagnosis, and to apply genetic information in precision medicine and personalized medicine, it will be critical, in the long run, to combine the NGS results with genome-wide association studies (GWASs) as well as gene-gene and gene-environment interactions.

Challenges

In pursuing NGS-based research and implementing NGS technologies in clinical applications, many hurdles may be encountered that need to be resolved.

First of all, enormous amounts of data that must be properly managed, stored, and analyzed are the obvious challenge posed by NGS. As the reagent costs of sequencing decrease with the development of better reagents and improved protocols, the number of sequencing projects is continuously increasing, the complexity of data analysis and management appear to be the primary limiting factor among researchers. Specifically, computing skills, hardware, storage, and network capabilities are necessary and critical to managing the massive data sets generated by large-scale NGS studies. Also, rapidly developing technology constantly require upgrades of analytic software and bioinformatics pipelines, which is costly and warrants revalidation before implementation.

Second, the complexity of NGS caused by the large size of the genome tested in multiple barcoded samples leads to the challenge of data validation. Thorough validation of the tests must be performed to implement NGS as a routine diagnostic test, as the majority of the NGS assays is intended for research only. NGS is an iterative process, which is the major problem in validating the performance characteristics of a clinical test for accuracy and reproducibility. Validation of a NGS tests involves optimizing simultaneously the

sequencing platform, the specific test/panel of genes, and the bioinformatics pipeline. Therefore, after altering one small part of the workflow such as changing the input DNA quantity, adding or changing even a single primer, or adjusting a parameter in the bioinformatics pipeline, the entire assay must be revalidated. Given that NGS assays require days to weeks to complete, validating these tests is time-consuming and costly. Moreover, the addition of a new marker or a new gene in the implemented sequencing panels will lead to the revalidation of existing ones or complete redesign and validation.

Finally, the clinical laboratories also have to face the challenges of clinical reporting when performing clinical NGS tests. It is necessary to establish standardized mutation annotation and classification such as germline, somatic, variant of unknown significance, that interface with literature and publicly available mutation databases (COSMIC, dbSNP) to generate a comprehensive report for a patient with disease information. In addition, a large number of mutations can be incidental findings and of unknown significance, posing a challenge for the laboratories to report due to the medicolegal implications. Therefore, laboratories should communicate their policy regarding the reporting of incidental findings and follow them through reporting.

Drawbacks

Although the potential of genomics is enormous, certain disease scenarios will be helped by NGS more than others. As aforementioned above, cancer and rare diseases are the areas that NGS has been widely applied to detect *de novo* mutations or mosaicism in sporadic patients without a prior hypothesis about the causal gene. NGS has also been used to elucidate the genetic causes for Mendelian and heterogeneous disorders. However, there are many diseases that do not very much benefit from NGS such as noninherited metabolic syndrome and nutritional deficiency. Indeed, a phenotype is the holistic outcome of many different layers of molecular regulation and is also influenced by environmental interactions. DNA is only one regulatory layer; many other complex molecular processes occurred at RNA and protein levels also contribute to the disease phenotypes. Therefore, it is an unrealistic hope of trying to relate all human traits with DNA level information, and it is not justified to say that NGS is helpful to all diseases.

Conclusions

NGS is a powerful technology generating enormous amounts of data, which is revolutionizing almost all aspects of research in human health and diseases. Several key areas reviewed here demonstrate that NGS is a practical, attractive, and economically feasible technology for clinical applications. The implementation of NGS approaches to clinical diagnostics has still been an on-going pursuit and requires a thorough understanding of the principles, utilities, challenges, limitations, and interpretation of the results as well as the improvement of the technologies. Therefore, continuous improvement of the accuracy of sequencing chemistries, computational algorithms, bioinformatics analytical tools, and the improved methods for sample processing and variant interpretation will make NGS a new "gold standard" method for clinical laboratories.

ACKNOWLEDGEMENTS

We thank Vinmec Research Institute of Stem Cell and Gene Technology, Vinmec International Hospital, and particularly, the Laboratory of Gene Technology for their support and careful review of the manuscript.

REFERENCES

[1] F. Sanger, S. Nicklen, A.R. Coulson (1977), "DNA sequencing with chain-terminating inhibitors", *Proceedings of the National Academy of Sciences of the United States of America*, **74,** pp.5463-5467.

[2] D.J. Pagliarini, S.E. Calvo, B. Chang, S.A. Sheth, S.B. Vafai, S.E. Ong, G.A. Walford, C. Sugiana, A. Boneh, W.K. Chen, D.E. Hill, M. Vidal, J.G. Evans, D.R. Thorburn, S.A. Carr, V.K. Mootha (2008), "A mitochondrial protein compendium elucidates complex I disease biology", *Cell*, **134**, pp.112-123.

[3] S. Calvo, M. Jain, X.H. Xie, S.A. Sheth, B. Chang, O.A. Goldberger, A. Spinazzola, M. Zeviani, S.A. Carr, V.K. Mootha (2006), "Systematic identification of human mitochondrial disease genes through integrative genomics", *Nature Genetics*, **38**, pp.576-582.

[4] D.R. Thorburn (2004), "Mitochondrial disorders: Prevalence, myths and advances", *Journal of Inherited Metabolic Disease*, **27**, pp.349-362.

[5] C. Scharfe, H.H.S. Lu, J.K. Neuenburg, E.A. Allen, G.C. Li, T. Klopstock, T.M. Cowan, G.M. Enns, R.W. Davis (2009), "Mapping Gene Associations in Human Mitochondria using Clinical Disease Phenotypes", *PLOS Computational Biology*, **5**, e1000374.

[6] J. Wang, H. Yu, V.W. Zhang, X. Tian, Y. Feng, G. Wang, E. Gorman, H. Wang, R.E. Lutz, E.S. Schmitt, S. Peacock, L.J. Wong (2015), "Capture-based high-coverage NGS: a powerful tool to uncover a wide spectrum of mutation types", *Genetics in Medicine*, **18**, p.513.

[7] E.R. Mardis (2008), "Next-generation DNA sequencing methods", *Annual Review of Genomics and Human Genetics*, **9**, pp.387-402.

[8] M.L. Metzker (2010), "Applications of NGS sequencing technologies - the next generation", *Nature Reviews Genetics*, **11**, pp.31-46.

[9] J. Shendure, H. Ji (2008), "Next-generation DNA sequencing", *Nature Biotechnology*, **26**, pp.1135-1145.

[10] R.A. Gibbs (2011), "Bringing genomics and genetics back together", *Science*, **331**, pp.548-548.

[11] F.S. Collins (2011), "Faces of the Genome", *Science*, **331**, pp.546-546.

[12] E.D. Green, M.S. Guyer, I. Natl. Human Genome Res. (2011), "Charting a course for genomic medicine from base pairs to bedside", *Nature*, **470**, pp.204-213.

[13] E.S. Lander (2011), "Initial impact of the sequencing of the human genome", *Nature*, **470**, pp.187-197.

[14] S.B. Ng, D.A. Nickerson, M.J. Bamshad, J. Shendure (2010), "MPS and rare disease", *Human Molecular Genetics*, **19**, pp.R119-R124.

[15] 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 (2008), "Application of MPS to microRNA profiling and discovery in human embryonic stem cells", *Genome Research*, **18**, pp.610-621.

[16] R.D. Hawkins, G.C. Hon, B. Ren (2010), "Next-generation genomics: an integrative approach", *Nature Reviews Genetics*, **11**, pp.476-486.

[17] R. Lister, M. Pelizzola, R.H. Dowen, R.D. Hawkins, G. Hon, J. Tonti-Filippini, J.R. Nery, L. Lee, Z. Ye, Q.M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A.H. Millar, J.A. Thomson, B. Ren, J.R. Ecker (2009), "Human DNA methylomes at base resolution show widespread epigenomic differences", *Nature*, **462**, pp.315-322.

[18] H. Hu, K. Wrogemann, V. Kalscheuer, A. Tzschach, H. Richard, S.A. Haas, C. Menzel, M. Bienek, G. Froyen, M. Raynaud, H. Van Bokhoven, J. Chelly, H. Ropers, W. Chen (2009), "Mutation screening in 86 known X-linked mental retardation genes by droplet-based multiplex PCR and massive parallel sequencing", *The HUGO Journal*, **3**, pp.41-49.

[19] S. Gowrisankar, J.P. Lerner-Ellis, S. Cox, E.T. White, M. Manion, K. LeVan, J. Liu, L.M. Farwell, O. Lartchouk, H.L. Rehm, B.H. Funke (2010), "Evaluation of Second-Generation Sequencing of 19 Dilated Cardiomyopathy Genes for Clinical Applications", *Journal of Molecular Diagnostics*, **12**, pp.818-827.

[20] K.V. Voelkerding, S. Dames, J.D. Durtschi (2010), "NGS for Clinical Diagnostics-Principles and Application to Targeted Resequencing for Hypertrophic Cardiomyopathy A Paper from the 2009 William Beaumont Hospital Symposium on Molecular Pathology", *Journal of Molecular Diagnostics*, **12**, pp.539-551.

[21] V. Vasta, S.B. Ng, E.H. Turner, J. Shendure, S.H. Hahn (2009), "Next generation sequence analysis for mitochondrial disorders", *Genome Medicine*, **1**, doi: 10.1186/gm100.

[22] A.K. Gupta, U.D. Gupta (2014), "NGS and its applications A2 - Verma, Ashish S", chapter 19, in A. Singh (Ed.) *Animal Biotechnology*, Academic Press, San Diego, pp.345-367.

[23] M. Margulies, M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka, M.S. Braverman, Y.J. Chen, Z.T. Chen, S.B. Dewell, L. Du, J.M. Fierro, X.V. Gomes, B.C. Godwin, W. He, S. Helgesen, C.H. Ho, G.P. Irzyk, S.C. Jando, M.L.I. Alenquer, T.P. Jarvie, K.B. Jirage, J.B. Kim, J.R. Knight, J.R. Lanza, J.H. Leamon, S.M. Lefkowitz, M. Lei, J. Li, K.L. Lohman, H. Lu, V.B. Makhijani, K.E. McDade, M.P. McKenna, E.W. Myers, E. Nickerson, J.R. Nobile, R. Plant, B.P. Puc, M.T. Ronan, G.T. Roth, G.J. Sarkis, J.F. Simons, J.W. Simpson, M. Srinivasan, K.R. Tartaro, A. Tomasz, K.A. Vogt, G.A. Volkmer, S.H. Wang, Y. Wang, M.P. Weiner, P.G. Yu, R.F. Begley, J.M. Rothberg (2005), "Genome sequencing in microfabricated high-density picolitre reactors", *Nature*, **437**, pp.376-380.

[24] S. Bennett (2004), "Solexa Ltd", *Pharmacogenomics*, **5**, pp.433-438.

[25] S.T. Bennett, C. Barnes, A. Cox, L. Davies, C. Brown (2005), "Toward the $1000 human genome", *Pharmacogenomics*, **6**, pp.373-382.

[26] J. Shendure, G.J. Porreca, N.B. Reppas, X.X. Lin, J.P. McCutcheon, A.M. Rosenbaum, M.D. Wang, K. Zhang, R.D. Mitra, G.M. Church (2005), "Accurate multiplex polony sequencing of an evolved bacterial genome", *Science*, **309**, pp.1728-1732.

[27] T.D. Harris, P.R. Buzby, H. Babcock, E. Beer, J. Bowers, I. Braslavsky, M. Causey, J. Colonell, J. DiMeo, J.W. Efcavitch, E. Giladi, J. Gill, J. Healy, M. Jarosz, D. Lapen, K. Moulton, S.R. Quake, K. Steinmann, E. Thayer, A. Tyurina, R. Ward, H. Weiss, Z. Xie (2008), "Single-molecule DNA sequencing of a viral genome", *Science*, **320**, pp.106-109.

[28] J. Korlach, K.P. Bjornson, B.P. Chaudhuri, R.L. Cicero, B.A. Flusberg, J.J. Gray, D. Holden, R. Saxena, J. Wegener, S.W. Turner (2010), "Real-time DNA sequencing from single polymerase molecules", in: N.G. Walter (Ed.) *Methods in Enzymology*, **Vol.472**, Single Molecule Tools, Pt A: Fluorescence Based Approaches, pp.431-455.

[29] E.H. Turner, S.B. Ng, D.A. Nickerson, J. Shendure (2009), "Methods for Genomic Partitioning", *Annual Review of Genomics and Human Genetics*, **10**, pp.263-284.

[30] J.M. 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.R. Nobile, B.P. 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. Feierstein, 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 (2011), "An integrated semiconductor device enabling non-optical genome sequencing", *Nature*, **475**, pp.348-352.

[31] K.V. Voelkerding, S.A. Dames, J.D. Durtschi (2009), "NGS: From Basic Research to Diagnostics", *Clinical Chemistry*, **55**, pp.641-658.

[32] R.R. Singh, R. Luthra, M.J. Routbort, K.P. Patel, L.J. Medeiros (2016), "Implementation of NGS in clinical molecular diagnostic

laboratories: advantages, challenges and potential", *Expert Review of Precision Medicine and Drug Development*, **1**, pp.109-120.

[33] I. Company (2016), *Extended RAS Panel*.

[34] NIH (2017), *The Cancer genome ATLAS*.

[35] International cancer genome consortium (2017).

[36] S. Jones, R.H. Hruban, M. Kamiyama, M. Borges, X. Zhang, D.W. Parsons, J.C.H. Lin, E. Palmisano, K. Brune, E.M. Jaffee, C.A. Iacobuzio-Donahue, A. Maitra, G. Parmigiani, S.E. Kern, V.E. Velculescu, K.W. Kinzler, B. Vogelstein, J.R. Eshleman, M. Goggins, A.P. Klein (2009), "Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene", *Science*, **324**, p.217.

[37] C. Palles, J.B. Cazier, K.M. Howarth, E. Domingo, A.M. Jones, P. Broderick, Z. Kemp, S.L. Spain, E. Guarino, I. Salguero, A. Sherborne, D. Chubb, L.G. Carvajal-Carmona, Y. Ma, K. Kaur, S. Dobbins, E. Barclay, M. Gorman, L. Martin, M.B. Kovac, S. Humphray, C.C. The, W.G.S.C. The, A. Lucassen, C.C. Holmes, D. Bentley, P. Donnelly, J. Taylor, C. Petridis, R. Roylance, E.J. Sawyer, D.J. Kerr, S. Clark, J. Grimes, S.E. Kearsey, H.J.W. Thomas, G. McVean, R.S. Houlston, I. Tomlinson (2012), "Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas", *Nature Genetics*, **45**, p.136.

[38] M. Esteller (2011), "Non-coding RNAs in human disease", *Nature Reviews Genetics*, **12**, p.861.

[39] I.V. Pronina, V.I. Loginov, A.M. Burdennyy, M.V. Fridman, V.N. Senchenko, T.P. Kazubskaya, N.E. Kushlinskii, A.A. Dmitriev, E.A. Braga (2017), "DNA methylation contributes to deregulation of 12 cancer-associated microRNAs and breast cancer progression", *Gene*, **604**, pp.1-8.

[40] J. Gagan, E.M. Van Allen (2015), "NGS to guide cancer therapy", *Genome Medicine*, **7**, p.80.

[41] R. Luthra, H. Chen, S. Roy-Chowdhuri, R.R. Singh (2015), "NGS in clinical molecular diagnostics of cancer: advantages and challenges", *Cancers*, **7**, pp.2023-2036.

[42] V.G. LeBlanc, M.A. Marra (2015), "NGS Approaches in Cancer: Where have they brought us and where will they take us?", *Cancers*, **7**, pp.1925-1958.

[43] R. Kamps, D.R. Brandão, J.B. Bosch, D.A. Paulussen, S. Xanthoulea, J.M. Blok, A. Romano (2017), "NGS in oncology: genetic diagnosis, risk prediction and cancer classification", *International Journal of Molecular Sciences*, **18**, doi: 10.3390/ijms18020308.

[44] B. Rabbani, M. Tekin, N. Mahdieh (2013), "The promise of whole-exome sequencing in medical genetics", *Journal of Human Genetics*, **59**, p.5.

[45] K.A. Johansen Taber, B.D. Dickinson, M. Wilson (2014), "The promise and challenges of next-generation genome sequencing for clinical care", *JAMA Internal Medicine*, **174**, pp.275-280.

[46] T.A. Manolio, R.L. Chisholm, B. Ozenberger, D.M. Roden, M.S. Williams, R. Wilson, D. Bick, E.P. Bottinger, M.H. Brilliant, C. Eng, K.A. Frazer, B. Korf, D.H. Ledbetter, J.R. Lupski, C. Marsh, D. Mrazek, M.F. Murray, P.H. O'Donnell, D.J. Rader, M.V. Relling, A.R. Shuldiner, D. Valle, R. Weinshilboum, E.D. Green, G.S. Ginsburg (2013), "Implementing genomic medicine in the clinic: the future is here", *Genetics in Medicine*, **15**, p.258.

[47] G.H.J.M. Leyten, D. Hessels, F.P. Smit, S.A. Jannink, H. de Jong, W.J.G. Melchers, E.B. Cornel, T.M. de Reijke, H. Vergunst, P. Kil, B.C. Knipscheer, C.A. Hulsbergen-van de Kaa, P.F.A. Mulders, I.M. van Oort, J.A. Schalken (2015), "Identification of a Candidate Gene Panel for the Early Diagnosis of Prostate Cancer", *Clinical Cancer Research*, **21**, p.3061.

[48] B. Rabbani, N. Mahdieh, K. Hosomichi, H. Nakaoka, I. Inoue (2012), "NGS: impact of exome sequencing in characterizing Mendelian disorders", *Journal of Human Genetics*, **57**, pp.621-632.

[49] M. Seifi, A. Ghasemi, S. Raeisi, S. Heidarzadeh (2017), "Application of NGS in Clinical Molecular Diagnostics", *Brazilian Archives of Biology and Technology*, ***60***, http://dx.doi.org/10.1590/1678-4324-2017160414.

[50] N.A. Schlipf, R. Schuele, S. Klimpe, K.N. Karle, M. Synofzik, J. Schicks, O. Riess, L. Schoels, P. Bauer (2011), "Amplicon-based high-throughput pooled sequencing identifies mutations in CYP7B1 and SPG7 in sporadic spastic paraplegia patients", *Clinical Genetics*, **80**, pp.148-160.

[51] M.A. Jones, S. Bhide, E. Chin, B.G. Ng, D. Rhodenizer, V.W. Zhang, J.J. Sun, A. Tanner, H.H. Freeze, M.R. Hegde (2011), "Targeted polymerase chain reaction-based enrichment and NGS for diagnostic testing of congenital disorders of glycosylation", *Genetics in Medicine*, **13**, pp.921-932.

[52] D.A. Simpson, G.R. Clark, S. Alexander, G. Silvestri, C.E. Willoughby (2011), "Molecular diagnosis for heterogeneous genetic diseases with targeted high-throughput DNA sequencing applied to retinitis pigmentosa", *Journal of Medical Genetics*, **48**, p.145.

[53] L.S. Chou, C.S.J. Liu, B. Boese, X. Zhang, R. Mao (2010), "DNA Sequence Capture and Enrichment by Microarray Followed by NGS for Targeted Resequencing: Neurofibromatosis Type 1 Gene as a Model", *Clinical Chemistry*, **56**, pp.62-72.

[54] M. Baetens, L. Van Laer, K. De Leeneer, J. Hellemans, J. De Schrijver, H. Van De Voorde, M. Renard, H. Dietz, R.V. Lacro, B. Menten, W. Van Criekinge, J. De Backer, A. De Paepe, B. Loeys, P.J. Coucke (2011), "Applying Massive Parallel Sequencing to Molecular Diagnosis of Marfan and Loeys-Dietz Syndromes", *Human Mutation*, **32**, pp.1053-1062.

[55] J.R. Lupski, J.G. Reid, C. Gonzaga-Jauregui, D. Rio Deiros, D.C.Y. Chen, L. Nazareth, M. Bainbridge, H. Dinh, C. Jing, D.A. Wheeler, A.L. McGuire, F. Zhang, P. Stankiewicz, J.J. Halperin, C. Yang, C. Gehman, D. Guo, R.K. Irikat, W. Tom, N.J. Fantin, D.M. Muzny, R.A. Gibbs (2010), "Whole-Genome Sequencing in a Patient with Charcot-Marie-Tooth Neuropathy", *New England Journal of Medicine*, **362**, pp.1181-1191.

[56] S.B. Ng, K.J. Buckingham, C. Lee, A.W. Bigham, H.K. Tabor, K.M. Dent, C.D. Huff, P.T. Shannon, E.W. Jabs, D.A. Nickerson, J. Shendure, M.J. Bamshad (2009), "Exome sequencing identifies the cause of a mendelian disorder", *Nature Genetics*, **42**, p.30.

[57] S.B. Ng, A.W. Bigham, K.J. Buckingham, M.C. Hannibal, M.J. McMillin, H.I. Gildersleeve, A.E. Beck, H.K. Tabor, G.M. Cooper, H.C. Mefford, C. Lee, E.H. Turner, J.D. Smith, M.J. Rieder, K.I. Yoshiura, N. Matsumoto, T. Ohta, N. Niikawa, D.A. Nickerson, M.J. Bamshad, J. Shendure (2010), "Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome", *Nature Genetics*, **42**, p.790.

[58] D.A. Simpson, G.R. Clark, S. Alexander, G. Silvestri, C.E. Willoughby (2011), "Molecular diagnosis for heterogeneous genetic diseases with targeted high-throughput DNA sequencing applied to retinitis pigmentosa", *Journal of Medical Genetics*, **48**, pp.145-151.

[59] S. Tang, T. Huang (2010), "Characterization of mitochondrial DNA heteroplasmy using a parallel sequencing system", *Biotechniques*, **48**, pp.287-292.

[60] E.T. Weimer, M. Montgomery, R. Petraroia, J. Crawford, J.L. Schmitz "Performance Characteristics and Validation of NGS for Human Leucocyte Antigen Typing", *The Journal of Molecular Diagnostics*, **18**, pp.668-675.

[61] A.H. Németh, A.C. Kwasniewska, S. Lise, R. Parolin Schnekenberg, E.B.E. Becker, K.D. Bera, M.E. Shanks, L. Gregory, D. Buck, M. Zameel Cader, K. Talbot, R. de Silva, N. Fletcher, R. Hastings, S. Jayawant, P.J. Morrison, P. Worth, M. Taylor, J. Tolmie, M. O'Regan, R. Valentine, E. Packham, J. Evans, A. Seller, J. Ragoussis (2013), "NGS for molecular diagnosis of neurological disorders using ataxias as a model", *Brain*, **136**, pp.3106-3118.

[62] I. Company (2012), *Cystic Fibrosis*.

[63] Y.M.D. Lo, N. Corbetta, P.F. Chamberlain, V. Rai, I.L. Sargent, C.W.G. Redman, J.S. Wainscoat (1997), "Presence of fetal DNA in maternal plasma and serum", *Lancet*, **350**, pp.485-487.

[64] R.W.K. Chiu, H. Sun, R. Akolekar, C. Clouser, C. Lee, K. McKernan, D. Zhou, K.H. Nicolaides, Y.M.D. Lo (2010), "Maternal Plasma DNA Analysis with MPS by Ligation for Noninvasive Prenatal Diagnosis of Trisomy 21", *Clinical Chemistry*, **56**, pp.459-463.

[65] M. Ehrich, C. Deciu, T. Zwiefelhofer, J.A. Tynan, L. Cagasan, R. Tim, V. Lu, R. McCullough, E. McCarthy, A.O.H. Nygren, J. Dean, L. Tang, D. Hutchison, T. Lu, H. Wang, V. Angkachatchai, P. Oeth, C.R. Cantor, A. Bombard, D. van den Boom (2011), "Noninvasive detection of fetal trisomy 21 by sequencing of DNA in maternal blood: a study in a clinical setting", *American Journal of Obstetrics and Gynecology*, **204**, pp.205e1-205e11.

[66] G.E. Palomaki, E.M. Kloza, G.M. Lambert-Messerlian, J.E. Haddow, L.M. Neveux, M. Ehrich, D. van den Boom, A.T. Bombard, C. Deciu, W.W. Grody, S.F. Nelson, J.A. Canick (2011), "DNA sequencing of maternal plasma to detect Down syndrome: An international clinical validation study", *Genetics in Medicine*, **13**, pp.913-920.

[67] R.B. Lefkowitz, J.A. Tynan, T. Liu, Y. Wu, A.R. Mazloom, E. Almasri, G. Hogg, V. Angkachatchai, C. Zhao, D.S. Grosu, G. McLennan, M. Ehrich (2016), "Clinical validation of a noninvasive prenatal test for genomewide detection of fetal copy number variants", *American Journal of Obstetrics & Gynecology*, **215**, pp.227.e1-227.e16.

[68] Y.M.D. Lo, K.C.A. Chan, H. Sun, E.Z. Chen, P. Jiang, F.M.F. Lun, Y.W. Zheng, T.Y. Leung, T.K. Lau, C.R. Cantor, R.W.K. Chiu (2010), "Maternal Plasma DNA Sequencing Reveals the Genome-Wide Genetic and Mutational Profile of the Fetus", *Science Translational Medicine*, **2**, doi: 10.1126/scitranslmed.3001720.