Hindawi Publishing Corporation
BioMed Research International
Volume 2015, Article ID 161648, 15 pages
http://dx.doi.org/10.1155/2015/161648

Review Article

Cracking the Code of Human Diseases Using Next-Generation Sequencing: Applications, Challenges, and Perspectives

Vincenza Precone,[1,2] Valentina Del Monaco,[1,2] Maria Valeria Esposito,[1] Fatima Domenica Elisa De Palma,[1,2] Anna Ruocco,[1,2] Francesco Salvatore,[1,2,3] and Valeria D'Argenio[1,2]

[1]*CEINGE-Biotecnologie Avanzate, Via G Salvatore 486, 80145 Naples, Italy*
[2]*Department of Molecular Medicine and Medical Biotechnologies, University of Naples Federico II, Ed. 19, Via Sergio Pansini 5, 80131 Naples, Italy*
[3]*IRCCS-Fondazione SDN, 80143 Naples, Italy*

Correspondence should be addressed to Francesco Salvatore; salvator@unina.it
and Valeria D'Argenio; dargenio@ceinge.unina.it

Received 6 June 2015; Revised 30 September 2015; Accepted 18 October 2015

Academic Editor: Paul Baird

Next-generation sequencing (NGS) technologies have greatly impacted on every field of molecular research mainly because they reduce costs and increase throughput of DNA sequencing. These features, together with the technology's flexibility, have opened the way to a variety of applications including the study of the molecular basis of human diseases. Several analytical approaches have been developed to selectively enrich regions of interest from the whole genome in order to identify germinal and/or somatic sequence variants and to study DNA methylation. These approaches are now widely used in research, and they are already being used in routine molecular diagnostics. However, some issues are still controversial, namely, standardization of methods, data analysis and storage, and ethical aspects. Besides providing an overview of the NGS-based approaches most frequently used to study the molecular basis of human diseases at DNA level, we discuss the principal challenges and applications of NGS in the field of human genomics.

1. Introduction

DNA sequencing is the process of determining the exact order of the nucleotides in a DNA segment, corresponding to single gene(s) or to a variety of molecules in the case of the whole genome, or a large part of it. Therefore, techniques able to do this have radically changed the course of molecular research in all its fields of application. In the last 30 years, the so-called Sanger sequencing has been the most widely used sequencing technology worldwide [1]. Its use peaked with the human genome project, which, in 2001, elucidated the entire human genome [2, 3]. Although Sanger sequencing is now completely automated, it is a method based on one or more amplicons that sequence relatively small strings of DNA rather than a complete gene or a set of genes in the human genome. Consequently, it is an expensive and time-consuming procedure if used to determine the entire chromosomal asset of a single organ tissue, or even cell.

In the last ten years, novel technologies, collectively called "next-generation sequencing" (NGS), have become available and have dramatically increased the throughput of DNA sequencing, thereby simultaneously reducing its costs [4]. Just to give an idea of what this means, it took more than 10 years to elucidate the first human genome sequence and it cost USD 3 billion. Using NGS instruments, the entire genome sequence of an individual can now be elucidated in less than 1 year and at a much lower cost [5]. It is expected that the sequencing of the entire genome of an individual will soon cost a few thousand USD (now 2,000 USD/genome without interpretation) [6].

These advances in genomic technologies have accelerated the search for genetic causes of human diseases and have enabled investigators to answer previously unapproachable questions about disease pathogenesis. In recent years, several NGS-based approaches have been devised and validated to facilitate the study of the molecular basis of human diseases [6]. The ultimate aims are (i) to develop novel, sensitive, accurate, and cost- and time-effective pipelines for molecular diagnostics and (ii) to elucidate the mechanisms involved in disease development and so to identify novel diagnostic, prognostic, and therapeutic markers [7]. In fact, NGS technologies can be easily applied to a wide range of research fields; thus far, they have been successfully used to analyze target regions of the human genome ranging in size from the entire exome to a restricted number of genes or a single amplicon [8–10]. In addition to the detection of nucleotide variants on DNA regions, NGS-based strategies can shed light on the DNA methylation status, at both single gene and genome-wide level [11]. Importantly, apart from the human genome, NGS gave immeasurable impetus to metagenomics. In fact, work based on NGS revealed the complexity of the microbiota not only in diverse tissues and organs, but also in relation to a variety of physiological conditions (e.g., sex, age, and circadian rhythm) as well as in pathological conditions [12].

Novel NGS-based strategies are continuously being developed and it is conceivable that these technologies will become even more routine also for diagnostic purposes, particularly in view of the progressive simplification of NGS protocols, the reduction in the operator's "hands on" work, and the launch of the "benchtop" NGS platforms. Notably, the integration of data obtained using several NGS-based strategies could shed light on the mechanisms involved in disease development and, in turn, reveal targets that can be manipulated to obtain better identification, stratification, and treatment of patients.

Here, we review the main NGS-based approaches currently used to study the molecular basis of human diseases at DNA level and discuss the main advantages, principal applications, and possible limitations of each.

2. NGS-Based Analytic Approaches for the Study of Human Diseases

2.1. Identification of DNA Sequence Variants. The study of entire genomes is faster and cheaper with NGS techniques than with conventional Sanger sequencing [13, 14]. However, the entire sequencing of a large number of samples is not yet feasible for routine use due to the cost, time, and infrastructures required. Consequently, various approaches to specifically enrich target genomic regions simultaneously thereby allowing barcoding of samples for sequence multiplexing have been developed and can be divided into polymerase chain reaction- (PCR-) based and non-PCR-based strategies. Both strategies can be used for NGS library preparation; the choice of the most appropriate strategy depends on the size of the target regions, the number of samples to be analyzed, the cost and time required, and the biological questions to be addressed, as reviewed in the following subsections. The specific analytic features and applications to human diseases of each of these approaches are summarized in Tables 1 and 2, respectively.

2.1.1. PCR-Based Strategies. PCR-based strategies have been the most widely used presequencing strategies since they are perfectly compatible with Sanger sequencing and also with all NGS methodologies and instruments: the amplicons resulting from amplification have NGS-platform-specific adapters ligated to their ends. These represent a library that is suitable for downstream sequencing reactions [8]. Since barcode sequence tags can be added during this step, sample multiplexing is possible [15]. Amplicon sequencing refers to ultradeep sequencing of PCR products obtained from one or several specific targets in order to determine, for example, genetic variations contained in precise portions of the genome. However, PCR amplification is too laborious for large-scale NGS downstream applications and risks being a bottleneck in the sample preparation workflow; consequently, more appropriate PCR procedures were developed. Below we briefly review long-range PCR, multiplex PCR, and microdroplet PCR [16–18].

Long-Range PCR. Long-range PCR procedures can be used to analyze up to several hundreds of kilobases (kb), such as the entire sequence of a single gene of interest. This enrichment strategy requires the design of large overlapping amplicons ranging in size from 2 to 12 kb. Each amplicon must be individually amplified, purified, and quantified. Then, all amplicons from the same sample can be pooled in equimolar amounts and used to obtain a library obtained using shotgun protocols. Since barcoded adapters can be added to the mixture in this step, sample multiplexing is allowed.

Long-range PCR has been used to obtain a comprehensive genetic map of a specific genomic locus, 136 kb on chr8q24, which is related to an increased risk of prostate and colon cancer [16]. The same strategy has been successfully used for the complete sequence analysis of large genes such as dystrophin, BRCA1, and BRCA2 [19, 20]. Although long-range PCR can be used to analyze an entire gene of interest including introns, the promoter, and the 3′ untranslated regions (UTR), it has some limitations. Long PCRs require high fidelity Taq polymerase, and amplifications conditions could require a long optimization time. Therefore, this approach is not suitable for large-scale applications and should be limited to the analysis of single genes.

Multiplex PCR. Multiplex PCR is a PCR improvement that enables the simultaneous analysis of multiple targets. In this case, amplicons of different sizes are amplified in one or more multiplex mixtures. Next, a nested PCR is carried out for adapter ligation and amplicon barcoding; all reactions from the same sample are pooled to obtain a multiple amplicon library/sample so that several tagged libraries can be sequenced together.

Multiplex PCR is especially suitable for the simultaneous sequencing of 15 and even more exons, in one or more genes of interest, which results in significant savings in terms of time and costs. Indeed, by targeting multiple genes simultaneously, more comprehensive information can be obtained

Table 1: Currently available NGS-based approaches for the study of the molecular basis of human diseases at DNA level.

Enrichment system	DNA input	Analytical time*	Sensitivity	Specificity	Max target size	Pros/cons
Long PCR	5 ng/amplicon	4-5 h	High	High	Depends on amplicon length	Rapid analysis of large genes/excellent quality of template DNA required; longer read lengths; greater fidelity and higher yields of Taq DNA polymerase; optimization of primer design and amplification conditions.
Multiplex PCR	5 ng/multiplex	3-4 h	High	High	Depends on amplicon length and multiplexing	Simultaneous analysis of several different targets; significant time and cost reductions/primer design and amplification conditions optimization; analysis of specific genes sequences only.
Microdroplet PCR	1.5 μg	48 h†	High	High	Up to 20,000 genomic loci	Analysis of a large number of genes in a single sequencing run; rapid high-quality data production/specific instrument acquisition required; high-quality DNA required; complex data analysis; high costs.
WES	500 ng–2 μg	92 h	High	>60%§	50–75 Mb§	High coverage in targeted regions; reduced costs for large genome analysis/mutations in regulatory regions will be missed; hard to identify structural variants and copy number variations; high sequencing depth required; complex data analysis.
Targeted capture	500 ng–2 μg	92 h	High	>60%§	Up to 50 Mb of custom regions	High resolution; cost reduction by pooling tagged samples; less susceptible to contamination and to mismatches/large amount of high-quality DNA required; influenced by repeating elements and high values of guanine-cytosine content; complex data analysis.
WG bisulfite sequencing	50 ng–1.5 μg	30 h	High	95% of CpGs	WG	High resolution and quantitative data/high DNA and sequencing coverage requirements; influenced by efficiency of bisulfite conversion; high cost per sample.

WES, whole-exome sequencing; WG, whole genome; Mb, megabase.
*In hours (h) and without considering sequencing time.
†Calculated considering 48 samples amplified/run.
§Depending on the design and on the commercial enrichment technology used.

Table 2: Principal clinical applications of NGS technologies to the study of the molecular basis of human diseases.

Clinical application	NGS approaches
Variant detection	Long PCR [15, 18, 19] Multiplex PCR [16, 20–30] Microdroplet PCR [31, 32] WES [47–50, 53–59, 67–75] Targeted capture [86–88, 91–96] WGS [110–112, 115–118]
Prenatal diagnosis	WES [60, 64–70] Targeted capture [101]
Circulating tumor cells analysis	WES [80]
Pharmacogenomics	WES [81–83] Targeted capture [97–99] WGS [113, 114]
Gene expression regulation	Microdroplet PCR [38] WG bisulfite sequencing [132–139] RNA-seq [156–162]

WES, whole-exome sequencing; WG, whole genome.

from a single test run. Notwithstanding these advantages, multiplex PCR can be used only for the analysis of specific known, not very large, carefully selected DNA regions, and the efficiency of the procedure depends strictly on primer design. Hence, time and costs increase in proportion to the number of genes analyzed and their size.

Various commercial kits are available for multiplex PCR-based enrichment. This approach has been used to study the molecular basis of several inherited diseases, namely, familial hypercholesterolemia [17], Alport syndrome [21], hematologic malignancies [22, 23], and cystic fibrosis [24, 25]. It has been also used to identify both somatic and germline mutations in cancer-related genes [26, 27]. Hansen et al., for example, developed an amplicon-based sequencing workflow for the analysis of four genes, predisposing to colorectal cancer [28]. This approach has also been used for the molecular analysis of the BRCA genes [29–33]. Finally, Sie et al. successfully used a multiplex PCR approach for somatic mutation profiling [34].

Taken together, these data indicate that multiplex PCR-based approaches are reliable in detecting mutations in small targets and suggest they could be used in routine diagnostic workflows thanks to their accuracy, speed, and cost-saving features. One advantage of multiplex PCR-based methods is the homogeneous distribution of amplicons, which in turn results in a homogeneous distribution of sequencing reads and a more uniform coverage. This is particularly important when these methods are used in clinical diagnostic procedures. In fact, in such cases, the entire target region must be sequenced without leaving any small sequence gaps, which might harbor the disease-causing mutation. In this context, it is advisable to verify the presence and relative abundance of the different amplicons using presequencing fragment analysis protocols. This verification step not only ensures a high sequence quality and a target coverage close to 100% but also may detect deleterious mutations [35].

Microdroplet PCR. Microdroplet PCR is a target enrichment procedure based on high throughput multiplex amplification. Each amplicon is amplified within a droplet of a water-in-oil emulsion to prevent cross-contamination in the reaction tube. Thus, while classical PCR is the amplification of one amplicon in one reaction tube and multiplex PCR is the amplification of a mixture of amplicons in the same reaction tube, microdroplet PCR is the amplification of up to 20,000 different amplicons each inside its own droplet all of which are collected in the same reaction tube. Droplets are generated by fully automated systems that use microfluidic supports (e.g., the RainDance ThunderStorm System). Primer droplets, each containing just one primer pair, are produced to specifically amplify each target of interest. In the same way, each DNA molecule is eluted in a droplet together with the amplification reaction mixture. Each of these two kinds of droplets flows along their microfluidic channel to the merge area, where they fuse into a single droplet that contains one primer pair and one DNA molecule. All the droplets are collected in a reaction tube and amplified before sequencing. This approach can be used to enrich targets within the human genome [18].

Microdroplet PCR has been applied in cancer [36] and in inherited diseases such as congenital muscular dystrophy [37]. In addition, it has been proposed for large-scale, targeted bisulfite sequencing for methylation profiling [38]. Based on the studies conducted so far, microdroplet technology could be used to process DNA for the massively parallel amplification of specific subsets of the human genome for targeted sequencing. More evidence must be obtained before microdroplet PCR can be applied for routine diagnostic purposes.

2.1.2. Non-PCR-Based Strategies. The so-called "sequence capture" approach, unlike PCR, is an excellent way to isolate large or highly dispersed regions from a pool of DNA molecules [39]. Sequence capture is essentially based on hybrid capture reactions for the selective enrichment of targeted genomic regions. Specific capture probes can be synthesized to enrich the regions of interest from the whole genome, thus obtaining a captured, adapted, and barcoded library for NGS applications [9, 40, 41]. In detail, DNA fragments hybridize to the capture probes synthesized on DNA microarray glass slides in array-based hybridization [9], whereas biotinylated DNA or RNA probes are used in liquid-phase hybridization. The nontargeted DNA fragments are washed away and the enriched DNA is recovered and used for high throughput sequencing. Currently, three major commercial products, namely, Nimblegen's SeqCap (array-based and solution-based), Illumina's TruSeq (solution-based), and Agilent's Sure-Select (array-based and solution-based), are used in combination with NGS platforms (e.g., Illumina, Roche 454, and Solid) to achieve efficient target enrichment. In addition to these, the HaloPlex technology from Agilent (solution-based) is a well-known example of an enrichment system featuring a selective circularization-based method that is a further development of the principle of selector probes used in several diagnostic approaches [42, 43].

Although PCR-based enrichment methods have the benefit of even coverage and high specificity, in some cases DNA sequence capture is to be preferred because hybridization is less sensitive to contamination and the procedure is less vulnerable to mismatches. In addition, PCR specificity depends on reaction optimization and primer design: large rearrangements in genomes, for example, may be undetectable. Unlike PCR, hybridization capture requires relatively large amounts of high-quality DNA and target molecules may be lost during library preparation. Therefore, this approach is more suitable for the study of large genomic regions, either contiguous or not contiguous, including the entire exome, as discussed in greater detail under "*Whole-Exome Sequencing*" and "*Targeted DNA Sequence Capture Sequencing*" below.

Whole-Exome Sequencing. Whole-exome sequencing (WES) serves to selectively sequence the coding regions of a genome in order to discover rare or common variants associated with a disorder or phenotype [44, 45]. In humans, the exome represents approximately 1% (30 megabases (Mb)) of the genome and it accounts for over 85% of disease-causing mutations [10, 46]. Consequently, WES is an attractive and practical approach for the study of coding variants related to rare Mendelian disorders and of many disease-predisposing single nucleotide polymorphisms throughout the exome [47].

Various target enrichment strategies are used in exome sequencing to pull out the coding regions from the whole genome, namely, array-based technologies, multiplex PCR, selector probe (HaloPlex), solution hybridization (Illumina TruSeq, Agilent SureSelect, and NimbleGen SeqCap EZ), and molecular-inversion probes (MIPs) [47–50]. Each approach has its capture efficiency and the choice of one approach over another depends mainly on the researcher's preference and needs [51, 52].

The last few years has seen an exponential growth in WES studies. This technique has been successfully used to identify the causative variations in such heterogeneous conditions as hearing loss [53–57], monogenic types of diabetes [58, 59], nonsyndromic mental retardation [60], and lysosomal disorders [61]. In addition, WES has become a tool in routine clinical practice [62]. One example is the case reported by Choi et al. in 2009 [63] of a patient supposedly affected by Bartter syndrome, a recessive renal tubular disorder characterized by hypokalemia and hypochloremia, who carried a novel homozygous mutation in the SLC26A3 gene that is known to cause congenital chloride-losing diarrhea. Therefore, the patient underwent a clinical reevaluation that led to a final diagnosis of congenital chloride-losing diarrhea.

An exciting new application of WES is in the prenatal diagnosis of aneuploidies. Fetal DNA can be extracted from the plasma fraction of maternal peripheral blood thereby avoiding an invasive approach [64–69]. A very recent paper published in the New England Journal of Medicine showed that prenatal testing on plasma-cell free DNA had significantly lower false positive rates and higher positive predictive values in detecting trisomies 21 and 18 than standard screening [70].

Yet another field of WES application is in cancer genetic research. Human cancer is characterized by the accumulation of genetic alterations and WES is an optimal tool for multiple gene testing. For example, in the field of the molecular genetics of cancer, WES has been applied in gastric cancer [71, 72], lymphomas [73, 74], breast cancer [75], melanomas [76, 77], and prostate cancer [78, 79]. Recently, WES was used to study circulating tumor cells in metastatic prostate cancer and revealed, with high accuracy, single nucleotide variants that can potentially track tumor evolution, guide therapy, and monitor relapses [80]. Cost-effective targeted WES methods have also resulted in a step forward in the field of personalized chemotherapy [81] and are now being used in the fields of pharmacogenetics and personalized medicine in general to identify inherited genetic variants able to predict individual responses to specific treatment. In fact, Daneshjou et al. [82] sequenced the entire exome of about 100 individuals and identified the genetic factors associated with the response to different doses of warfarin. More recently, using WES, Apellániz-Ruiz et al. [83] identified rare *CYP3A4* variants associated with individual susceptibility to toxicity to paclitaxel, which is a frequently used chemotherapy agent.

To sum up, WES has many strengths: it is less expensive than whole-genome sequencing (WGS) (one exome costs USD 1,000 whereas one genome costs around USD 2,000) [6], it is an efficient strategy with which to identify the genetic basis that underlies rare Mendelian disorders, and it provides a small dataset compared to WGS and is thus easier to interpret. On the other hand, exome sequencing obviously analyzes only about one percent of the entire genome and leaves out noncoding regions, such as untranslated regions (5′UTR and 3′UTR), promoters, and other potentially functional regions. This limitation could be overcome by customizing commercial exome capture kits according to the researcher's scientific needs. Other limitations and challenges related to WES are genetic heterogeneity (where several genes are associated with the same disorder), natural duplicated sequences throughout the genome, and pathogenic mutations located in noncoding genes, have corresponding pseudogenes, contain repetitive or high CG-rich regions, or are within the mitochondrial genome which are not detected. In addition, large deletions/duplications/rearrangements and mosaicism may not be detected by WES.

Finally, hybridization methods like WES may show some biases when analyzing repetitive regions versus PCR-based enrichment strategies and may not be able to efficiently enrich them. This uneven sequence coverage should be carefully evaluated in the diagnostic context since it could result in the loss of clinically relevant sequence variations and, thus, in false-negative results. Recently, Patwardhan et al. demonstrated that "augmented exome sequencing," which is designed to increase sequence coverage in medically relevant regions and in difficult-to-sequence regions, was more efficient in the clinical setting than other WES strategies [84].

Targeted DNA Sequence Capture Sequencing. Target enrichment strategies enable the efficient and rapid querying of specific large genomic regions of interest. Using this approach, one can isolate genomic regions in a library and thereby quantitate both germline and somatic variants. As mentioned in

Section 2.1.2, various hybridization-based enrichment methods, either in solution or on array, have been developed [85, 86]. Targeted enrichment-based approaches necessitate the identification of the target regions of interest and appropriately designed probes because repeating elements and high guanine-cytosine content hamper complete target coverage. Repetitive elements and internal duplications that may lead to cross-hybridization are usually removed with repeat-masked methods [87]. Long oligo probes (>50 base pairs (bp)), generally based on information from web-based tools (e.g., UCSC, Ensemble, and RefSeq Database), should be designed so as to increase probe specificity. Enrichment reliability can be evaluated based on several parameters: sensitivity and specificity, that is, the percentage of target bases that are represented by sequence reads and the percentage of sequences that map to the intended targets, respectively, uniformity of enrichment results and reproducibility of sequencing runs. Enrichment is considered good when more than 60% of reads map against the target regions. In addition, high coverage (>30x–40x) improves sequencing accuracy. Many enrichment protocols have been tested to increase target enrichment efficiency [88]. Compared to WGS, targeted resequencing can yield a much higher coverage of genomic regions and reduce the time and cost of the analysis. Currently, Illumina sequencing seems to dominate the sequencing market thanks to its lower cost and the shorter time required to process a large number of samples [51].

In terms of accuracy, Illumina sequencing provides a good data backbone and overcomes the homopolymer errors that occur with other sequencing technologies; consequently, it performs better than other strategies in the coverage of medically relevant DNA sequence variations [89, 90].

In conclusion, given the above and the fact that target enrichment technologies are easy to use, these methodologies lend themselves to the study of the molecular basis of genetic diseases, for both research and diagnostic purposes. Thus far, target enrichment-based strategies have been used to identify nucleotide variants [91, 92] and to validate novel diagnostic tools [93–96] and for drug resistance/sensitivity profiling [97–99]. This method is useful for the study of complex families with different genotype/phenotype correlations and to identify at-risk subjects [100]. Finally, hybridization-based targeted sequencing of single nucleotide polymorphism loci in maternal plasma DNA is a promising noninvasive approach to the prenatal diagnosis of fetal chromosomal anomalies [101]. Targeted enrichment technologies can be used to study not only monogenic and polygenic diseases [102–106], but also mitochondrial diseases [107], somatic mutations in tumors [108], and chromosomal anomalies [109]. Importantly, in all these settings, the analysis can be extended to the noncoding regions. Thus, although WES has some limitations in terms of data analysis, management, and interpretation, target capture enables the study of a more restricted and personalized target genomic region, thereby simplifying data analysis and reducing experimental time and costs.

2.1.3. Whole-Genome Sequencing. In the previous sections, we have reviewed the different analytical presequencing strategies currently available for the enrichment of specific genomic regions of interest (from PCR-based approaches to WES) and able to identify gene sequence variations at DNA level. Although not yet suitable for routine applications due to costs, analytical time required, and the huge amounts of data produced, WGS has some advantages over enrichment strategies and is briefly discussed in this section.

Whole-genome sequencing means the sequencing of the human genome in its totality. Therefore, it allows greater sequencing coverage uniformity and can identify also copy number variations, large insertions/deletions, and gene fusions. In addition, WGS covers all the genomic noncoding regions, including introns, promoters, UTRs, and regulatory elements; thus, it can shed light on the molecular alterations involved in specific diseases [110, 111]. Recently, Belkadi et al. compared the performances of WES and WGS in 6 individuals and found that sequencing quality was better and the detection rate of variant higher with WGS than with WES [112].

Whole-genome sequencing has also been applied in the fields of pharmacogenetics and pharmacogenomics [113]. For example, Mizzi et al. analyzed 482 unrelated individuals and identified several pharmacogene-related variants potentially involved in a given individual's response to treatment [114]. Notably, WGS-based approaches are effective tools for large population-based studies. Nagasaki el al. sequenced about 1,000 Japanese individuals and obtained a large population-specific dataset of DNA variants that is useful for epidemiological evaluations [115].

Finally, WGS has been used also to study the molecular basis of human diseases. A study involving 50 patients with severe intellectual disability showed the potential of this approach for the molecular diagnosis of complex diseases that can be caused by diverse kinds of mutations, including *de novo* mutations and copy number variants [116]. Whole-genome sequencing in a patient affected by familial adenomatous polyposis revealed an *APC* mosaicism, suggesting that WGS could be a powerful tool also in the detection of genetic mosaicism related to disease onset [117]. The WGS of a large cohort of patients with early onset familial Alzheimer disease revealed a disease-specific haplotype and a potential disease-progression modifier [118]. Of course, we expect that the number of studies designed to assess the power of WGS will grow rapidly in the next few years. However, as extensively reviewed elsewhere [110, 111], the applications of WGS for diagnostic purposes are still at an embryonic stage due to its costs and particularly due to problems associated with data analysis, interpretation, and storage. However, it is conceivable that as new bioinformatic pipelines for WGS data handling and interpretation become available and sequencing costs decrease, WGS will become the strategy of choice for studies of DNA sequence alterations.

2.2. Genome-Wide DNA Methylation Analysis. DNA methylation is one of the more stable and heritable epigenetic marks and its dysregulation is associated with many human diseases [119]. The human genome is highly methylated; approximately 80% of cytosines in CpG dinucleotides are

chemically modified [120]. The assays for identifying methylated CpG dinucleotides in a genome vary in terms of resolution and cost. Although PCR-based DNA methylation approaches have several advantages [121], the ongoing revolution in sequencing technology has opened the door to whole-genome DNA methylation analysis at a single-base-pair resolution. There are three main approaches to whole-genome methylation studies: enzyme digestion [11, 122], affinity enrichment [123–126], and bisulfite sequencing [127–131].

It is difficult to compare these three approaches given their complexity and diversity. The choice of the most suitable method depends on the desired coverage, accuracy, and resolution, as well as on the number of samples and the DNA quality and quantity. In general, enzyme digestion and affinity enrichment-based methods are low-resolution and essentially qualitative. Instead, bisulfite sequencing approaches have a higher resolution and provide quantitative estimates of methylation [132, 133]. Whole-genome bisulfite sequencing can assess about 95% of all CpG sites in the genome; however, high coverage (>500 million paired-end reads to achieve ~30x coverage) and high DNA input (1–1.5 versus 0.2 micrograms (g)) are required [131]; here it should be noted that Illumina's new EpiGnome Methyl-Seq kit requires only 50–100 nanograms (μg) of genomic DNA. Incomplete bisulfite conversion and differential PCR efficiency for methylated versus unmethylated sequences are the main limitations of bisulfite sequencing approaches. Recently, sequence capture enrichment methods have been developed also for DNA methylation assessment at a single base resolution (Nimblegen/Roche SeqCap Epi and Agilent MethylSeq). These approaches are based on the same enrichment methodology used to identify DNA sequence variations for the enrichment of bisulfite-converted DNA. Also in this case, these methods enable genome-wide capture of all the annotated CpG islands and customized enrichment to study pathways of interest [134].

Advances in genome-wide DNA methylation technology have also resulted in new strategies for the timely identification of novel diagnostic and prognostic biomarkers. For example, Huang and colleagues identified functional DNA methylation biomarkers predictive of the clinical outcome of ovarian cancer [135]; Jerónimo and colleagues showed that the glutathione-S-transferase P1 gene is methylated in >90% of prostate cancers [136]. Moreover, 70 genes were found to be significantly hypermethylated in gastric cancer tissue compared with those observed in normal tissue [137]. Ghosh and coworkers used whole-genome DNA methylation profiling to explore a potential association between parity and epigenetic changes in breast tissue from women with early parity and nulliparity. They identified six genes that are hypermethylated in the parous group [138]. Furthermore, in a very recent article, Warton and colleagues described a comprehensive technical analysis of free cell DNA (fcDNA) isolation from healthy subjects and enrichment of methylated sequences followed by NGS [139]. Their findings provide further support that whole-genome analysis of even small amounts of fcDNA can provide high-quality, validated genomic data that strengthen the potential of the usefulness of the methylation signature of fcDNA in clinical applications. The epigenomic data and the discovery of a specific pattern of epigenomic marks associated with specific functional regions have also helped to clarify genotype-phenotype association data.

2.3. Metagenomics. Although metagenomics does not include the study of human genes but focuses on the characterization of microbial communities, namely, microbiota, living in specific environments, such as skin or mucosal districts, growing evidence implicates the human microbiome in the development of various diseases. Therefore, it seems appropriate to mention this technique, although for further details we refer readers to more specific reviews on this topic [12, 140–142]. It is now well known that the human microbiome is required for the maintenance of the healthy status [143] and that microbial dysbiosis could play a role in several diseases like diabetes, inflammatory bowel diseases, obesity, and cancer [144–149]. So, the study of microbiome composition (both qualitatively and quantitatively) could clarify diseases pathogenesis and, in turn, pave the way to the development of novel diagnostic, prognostic, and therapeutic targets. This explains the great interest in the field. NGS-based approaches have greatly impacted also metagenomics, thereby providing a comprehensive view of microbial communities.

As we reviewed elsewhere [150], 16S bacterial ribosomal RNA (rRNA) characterization, obtained using NGS-based strategies, is now the technique most widely used to study the microbiome. It is an amplicon-based method that uses bacterial universal primers to amplify the entire microbiome in one PCR reaction [151]. The complexity of microbial communities is resolved after sequencing, specifically assigning each read to a group of bacteria through specific bioinformatics tools. Using this approach, we monitored the gut microbiome of a patient with Crohn disease before and after nutritional therapy and showed that this therapy was effective in restoring gut microbiome dysbiosis [152]. The same strategy has been recently used to characterize the esophageal microbiome in eosinophilic esophagitis [153], the lung microbiome of cystic fibrosis patients [154], and the subgingival microbiota in different periodontal diseases [155], and other studies are appearing almost daily. It is conceivable that future technical advances (especially those related to data analysis tools and to the availability of microbial community databases) will shed light on the functions of the human microbiome and its role in human diseases, and metagenomics could be an integrative means with which to study the molecular basis of human diseases at DNA level.

2.4. RNA Sequencing. The key to the molecular basis of human diseases and the genotype-phenotype relationship lies in gene expression and the mechanisms that control it. Consequently, genome-wide expression analyses are now pivotal in genomics and biomedical research. RNA sequencing (RNA-seq) technologies are elucidating the mechanisms that expand the genome's coding capacity and are revolutionizing the concept of gene expression regulation. RNA-seq is gradually replacing microarrays in high throughput

gene expression studies because it provides more quantitatively accurate measurements and also absolute transcript abundance data [156, 157]. RNA-seq also detects annotated transcripts as well as novel sequences, splice variants, exon junctions, noncoding RNA [158], single nucleotide polymorphisms [159], and fusion genes [160]. Various RNA-seq techniques are available, and the one to use depends on the RNA species being investigated: (i) total RNA sequencing (total RNA-seq) is a process that removes rRNAs and thus captures a broader range of gene expression changes and reveals novel transcripts in both coding and noncoding RNA species; (ii) coding RNA sequencing (mRNA-seq) provides information about poly-A tailed RNAs; and (iii) small RNAs sequencing (small RNA-seq) is used to discover and analyze novel microRNAs (miRNAs) and other small noncoding RNAs.

Changes in the expression of coding genes are controlled at multiple levels, from transcription to RNA processing and translation. Interestingly, the abundance of a transcript is directly modified by polymorphisms in regulatory elements [161]. An important class of variants, called *expression quantitative trait loci* (eQTL), influences the expression level of the gene in two ways (local or distant) [162, 163]. Genome-wide association studies combined with RNA-seq analysis can reveal the eQTL and can shed light on the mechanism whereby gene variability controls gene expression. The development of techniques based on the integration of these data will help to understand putative causal links between DNA variation and expression.

3. NGS Data Analysis and Storage

Data management and analysis pipelines, based on bioinformatics expertise and hardware infrastructures, have been developed to manage the massive sets of data produced by NGS. The analysis of NGS data is commonly based on three main analytic steps [164, 165], which are usually implemented via specific bioinformatic tools: (i) generation of sequences and assignment of base quality scores; (ii) demultiplexing (if necessary), read alignment, and variant calling; and (iii) identification and interpretation of variants according to guidelines [166]. Different scripts and/or pipelines are used in this last step depending on the kind of application, the type of samples sequenced, and the biological question to be addressed. For example, specific tools are available for metagenomics [167–169] and RNA-seq studies [170, 171].

Storage of the huge amount of data generated by NGS is an important issue. The Centers for Disease Control and Prevention (http://www.cdc.gov/clia/Resources/GetRM) requires storage of analytic systems records and data reports for at least 2 years [172]. However, given the rapid growth of knowledge in this field, longer storage could be contemplated for raw data files (e.g., fastq files) so that the primary results can be regenerated and analyzed as more advanced accurate techniques become available to verify the original interpretation. Consequently, sequencing centers should be equipped with powerful dedicated storage equipment: it has been estimated that 3.2 terabytes is required for the backup of 200 exomes [173]. These infrastructures may be beyond some laboratories, especially small laboratories. Publicly accessible clouds represent a possible solution to data storage, although, in a clinical context, data privacy issues should be carefully addressed [174].

4. Next-Generation Sequencing in the Clinical Setting and Ethical Disclosures

As reviewed in the previous sections, several NGS-based approaches are available to study the molecular bases of human diseases at DNA (and also at RNA) level and are now routinely used in clinical diagnostics [7, 110]. In essence, single gene analysis (multiplex PCR or long PCR) should be restricted to cases of low genetic heterogeneity; gene panel screening (microdroplet PCR or targeted sequence capture enrichment) should be preferred in case of highly heterogeneous diseases and/or for the differential diagnosis of very similar diseases, while WES and WGS should be considered in case of a very complex/rare phenotype, when *de novo* mutations are suspected or in case of noninformative results after the analysis of a panel of targeted genes [175]. The main issue concerning the clinical use of NGS-based approaches is the huge amount of data produced and its interpretation: the greater the genomic region analyzed, the greater the number of variants of uncertain significance identified. Another concern regards the so-called incidental findings, that is, mutation(s) with a known pathogenicity but not related to the medical condition for which the test was requested. Notably, the issue as to whether or not incidental findings should be communicated to the patient is hotly debated. Of course, a carefully produced patient-informed consent procedure should be part of pretest genetic counseling to prepare the patient for such kinds of results and ask for their concerns regarding the knowledge of the results.

5. Development of Next-Generation Sequencing Technology

Next-generation technologies were launched on the market about ten years ago and their history is characterized by a continuous release of novel instruments that usually feature an increased throughput/sequencing run [176]. The sample preparation workflow is usually based on three main steps: library preparation (achieved with different strategies depending on the project, as discussed in Section 2), library amplification, and high throughput sequencing. Each instrument uses specific chemistry, which accounts for some differences in sequencing accuracy and quality. The features of specific NGS platforms have been extensively reviewed and compared elsewhere, Illumina technology being the most widely used procedure to date [89]. Here, we provide an overview of the more recently developed technologies that may undergo further improvements in the next few years and possibly replace currently used platforms.

Pacific Biosciences has developed a NGS platform based on the real-time sequencing of single molecules (SMRT) during polymerization reactions. The DNA polymerase is

immobilized on the bottom of microscope chambers: the four phosphate-labelled nucleotides are eluted in the chambers and the sequences are read in real-time since fluorescent specific signals are recorded after incorporation of each nucleotide [177]. This procedure avoids library amplification and consequently reduces the risk of PCR artefacts. The latest version of this instrument has 150,000 chambers each able to sequence 55,000 reads/run with an average length of 20 kb and a maximum throughput of 1 gigabase per chamber in a four-hour run. The absence of clonal amplification, together with the long read length, makes this technology appealing for a variety of applications [178, 179]; however, the error rate is still high [180].

Another promising single molecule sequencing method is the nanopore-based strategy developed by Oxford Nanopore Technologies. It is based on a flow-cell containing hundreds of microwells. Each microwell contains a membrane with a nanopore through which an ionic current flows. When a biological molecule (DNA, RNA, or protein) flows through the pore, each nucleotide/amino acid results in a specific current change that discriminates among them [181]. Oxford Nanopore Technologies has developed three scalable instruments, including the MinION, a USB-device portable sequencer able to generate up to 16,000 reads/run with a maximum length of 60 kb and an average throughput of 90 Mb in an 18-hour run. Also in this case, the error rate is still too high for routine use. Finally, Complete Genomics has launched a fully automated apparatus for large-scale WES and WGS currently available only as a service. The instrument is based on Combinatorial Probe-Anchor Ligation chemistry [182] and is designed to sequence up to 10,000 genomes per year with a 50x coverage.

To overcome the limitations and drawbacks of currently available NGS instruments in terms of productivity, speed, cost, and accuracy, many other sequencers, based on diverse technologies (microfluidics, electron microscopy, nanopore-based strategies, and DNA transistor-based procedures), are currently under development. Therefore, it is conceivable that NGS will become a "routine" procedure in the not too distant future.

6. Conclusions

Next-generation sequencing technologies and the various associated procedures, together with the plethora of data they have generated and continue to generate, prompt several basic concepts. Many diseases, if not all, are characterized by genetic changes and thus are related to DNA sequence variants that must be analyzed for their consequences. However, not all these variants shed light on a given disease, since they are not directly pathogenetic or they contribute only slightly to the pathogenesis of the disease. Therefore, to address the complexity of the link between DNA sequence variants and human diseases, high throughput DNA sequencing should be as rapid and inexpensive as possible. Moreover, the sequence and bioinformatic strategies that are continuously emerging, together with the detection of DNA sequence changes, should also take account of DNA modifying processes, such as methylation. In addition, going from DNA to RNA, NGS has enabled researchers to mine the enormous amount of quantitative and qualitative data buried in the myriad of regulatory sequence elements that have been discovered in recent years, namely, miRNA, long noncoding RNA, small circular RNA, nuclear RNA, and nucleolar RNA. Notwithstanding the enormous potential of these methodologies, the procedures and guidelines have yet to be standardized, which probably reflects the continuous innovations that are being made in this field.

Notably, in the clinical context, besides confirming a disease or correlating a gene alteration to a given disease, NGS has the potential to become the frontline analysis for differential diagnosis among clinically confounding diseases. Thus, gene sequence profiling may be able to discriminate one disease from another, therefore enabling timely effective therapy.

The easier and faster production of highly accurate sequence data will certainly give impetus to what is now called "personalized genome analysis." Besides examining nosographically established diseases that affect many people, personalized genome analysis can help to understand, at single individual level, the minute differences that can affect the health status of each person, which, in turn, can lead to the application of "personalized medicine."

Another important area of research that has benefitted from NGS technology is the development of target drugs, namely, drugs or compounds that act as bullets able to strike a precise target in a DNA sequence, or at a corresponding protein level, in order to nullify or even to reverse the nucleic sequence from the variant to the wild-type status. This newly emerging technique of gene editing requires very accurate DNA and RNA sequences in order to design the most effective tools with which to revert the altered nucleotide into the wild type. This is a very promising avenue of research, provided the related ethical issues are overcome.

Finally, NGS-based approaches have greatly improved our understanding of the molecular basis of human diseases in a variety of ways that were unthinkable just a few years ago. The challenge now is to resolve the outstanding issues of standardization of procedures, the production and storage of personal data, and other ethical aspects, which we suspect will animate scientific and regulatory debates in the next few years.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the work described in this paper.

Acknowledgments

The authors thank Jean Ann Gilder (Scientific Communication srl., Naples, Italy) for editing the text and Vittorio Lucignano, CEINGE-Biotecnologie Avanzate, for technical assistance. This work has been supported by DIAINTECH-Regione Campania (to Francesco Salvatore), by PRIN 2012 (no. 2012WJSX8K), by POR Campania FSE 2007–2013,

Project CREME (to Francesco Salvatore), by Grant PS 35-126/Ind and Grant PON01_02589 (MICROMAP) 2012 from the Ministry of University and Research (to Francesco Salvatore), and by Grant RF-2010-2318372 from the Ministry of Health (to Francesco Salvatore).

References

[1] F. Sanger, S. Nicklen, and A. R. Coulson, "DNA sequencing with chain-terminating inhibitors," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 74, no. 12, pp. 5463–5467, 1977.

[2] J. C. Venter, M. D. Adams, E. W. Myers et al., "The sequence of the human genome," *Science*, vol. 291, no. 5507, pp. 1304–1351, 2001.

[3] International Human Genome Sequencing Consortium, "Finishing the euchromatic sequence of the human genome," *Nature*, vol. 431, no. 7011, pp. 931–945, 2004.

[4] E. R. Mardis, "The impact of next-generation sequencing technology on genetics," *Trends in Genetics*, vol. 24, no. 3, pp. 133–141, 2008.

[5] A. von Bubnoff, "Next-generation sequencing: the race is on," *Cell*, vol. 132, no. 5, pp. 721–723, 2008.

[6] J. Reuter, D. V. Spacek, and M. Snyder, "High-throughput sequencing technologies," *Molecular Cell*, vol. 58, no. 4, pp. 586–597, 2015.

[7] R. R. Gullapalli, K. V. Desai, L. Santana-Santos, J. A. Kant, and M. J. Becich, "Next generation sequencing in clinical medicine: challenges and lessons for pathology and biomedical informatics," *Journal of Pathology Informatics*, vol. 3, article 40, 2012.

[8] R. K. Thomas, E. Nickerson, J. F. Simons et al., "Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing," *Nature Medicine*, vol. 12, no. 7, pp. 852–855, 2006.

[9] T. J. Albert, M. N. Molla, D. M. Muzny et al., "Direct selection of human genomic loci by microarray hybridization," *Nature Methods*, vol. 4, no. 11, pp. 903–905, 2007.

[10] E. H. Turner, C. Lee, S. B. Ng, D. A. Nickerson, and J. Shendure, "Massively parallel exon capture and library-free resequencing across 16 genomes," *Nature Methods*, vol. 6, no. 5, pp. 315–316, 2009.

[11] M. P. Ball, J. Billy Li, Y. Gao et al., "Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells," *Nature Biotechnology*, vol. 27, no. 4, pp. 361–368, 2009.

[12] Human Microbiome Project Consortium, "Structure, function and diversity of the healthy human microbiome," *Nature*, vol. 486, no. 7402, pp. 207–214, 2012.

[13] J. R. ten Bosch and W. W. Grody, "Keeping up with the next generation: massively parallel sequencing in clinical diagnostics," *Journal of Molecular Diagnostics*, vol. 10, no. 6, pp. 484–492, 2008.

[14] J. Zhang, R. Chiodini, A. Badr, and G. Zhang, "The impact of next-generation sequencing on genomics," *Journal of Genetics and Genomics*, vol. 38, no. 3, pp. 95–109, 2011.

[15] A. M. Smith, L. E. Heisler, R. P. St.Onge et al., "Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples," *Nucleic Acids Research*, vol. 38, article e142, Article ID gkq368, 2010.

[16] M. Yeager, N. Xiao, R. B. Hayes et al., "Comprehensive resequence analysis of a 136 kb region of human chromosome 8q24 associated with prostate and colon cancers," *Human Genetics*, vol. 124, no. 2, pp. 161–170, 2008.

[17] S. Hollants, E. J. W. Redeker, and G. Matthijs, "Microfluidic amplification as a tool for massive parallel sequencing of the familial hypercholesterolemia genes," *Clinical Chemistry*, vol. 58, no. 4, pp. 717–724, 2012.

[18] R. Tewhey, J. B. Warner, M. Nakano et al., "Microdroplet-based PCR enrichment for large-scale targeted sequencing," *Nature Biotechnology*, vol. 27, no. 11, pp. 1025–1031, 2009.

[19] R. J. P. Bonnal, M. Severgnini, A. Castaldi et al., "Reliable resequencing of the human dystrophin locus by universal long polymerase chain reaction and massive pyrosequencing," *Analytical Biochemistry*, vol. 406, no. 2, pp. 176–184, 2010.

[20] I. Hernan, E. Borrs, M. de Sousa Dias et al., "Detection of genomic variations in *BRCA1* and *BRCA2* genes by long-range PCR and next-generation sequencing," *Journal of Molecular Diagnostics*, vol. 14, no. 3, pp. 286–293, 2012.

[21] V. Morinière, K. Dahan, P. Hilbert et al., "Improving mutation screening in familial hematuric nephropathies through next generation sequencing," *Journal of the American Society of Nephrology*, vol. 25, no. 12, pp. 2740–2751, 2014.

[22] A. Kohlmann, V. Grossmann, N. Nadarajah, and T. Haferlach, "Next-generation sequencing—feasibility and practicality in haematology," *British Journal of Haematology*, vol. 160, no. 6, pp. 736–753, 2013.

[23] C. Y. Ok, K. P. Patel, G. Garcia-Manero et al., "Mutational profiling of therapy-related myelodysplastic syndromes and acute myeloid leukemia by next generation sequencing, a comparison with de novo diseases," *Leukemia Research*, vol. 39, no. 3, pp. 348–354, 2015.

[24] D. Trujillano, M. D. Ramos, J. González et al., "Next generation diagnostics of cystic fibrosis and CFTR-related disorders by targeted multiplex high-coverage resequencing of CFTR," *Journal of Medical Genetics*, vol. 50, no. 7, pp. 455–462, 2013.

[25] A. N. Abou Tayoun, C. D. Tunkey, T. J. Pugh et al., "A comprehensive assay for CFTR mutational analysis using next-generation sequencing.," *Clinical Chemistry*, vol. 59, no. 10, pp. 1481–1488, 2013.

[26] F. Chang and M. M. Li, "Clinical application of amplicon-based next-generation sequencing in cancer," *Cancer Genetics*, vol. 206, no. 12, pp. 413–419, 2013.

[27] L. Beltrame, M. Di Marino, R. Fruscio et al., "Profiling cancer gene mutations in longitudinal epithelial ovarian cancer biopsies by targeted next-generation sequencing: a retrospective study," *Annals of Oncology*, vol. 26, no. 7, pp. 1363–1371, 2015.

[28] M. F. Hansen, U. Neckmann, L. A. Lavik et al., "A massive parallel sequencing workflow for diagnostic genetic testing of mismatch repair genes," *Molecular Genetics & Genomic Medicine*, vol. 2, no. 2, pp. 186–200, 2014.

[29] C. Desmedt, T. Voet, C. Sotiriou, and P. J. Campbell, "Next-generation sequencing in breast cancer: first take home messages," *Current Opinion in Oncology*, vol. 24, no. 6, pp. 597–604, 2012.

[30] F. Vaca-Paniagua, R. M. Alvarez-Gomez, V. Fragoso-Ontiveros et al., "Full-exon pyrosequencing screening of BRCA germline mutations in Mexican women with inherited breast and ovarian cancer," *PLoS ONE*, vol. 7, no. 5, Article ID e37432, 2012.

[31] F. Pern, N. Bogdanova, P. Schürmann et al., "Mutation analysis of BRCA1, BRCA2, PALB2 and BRD7 in a hospital-based series of German patients with triple-negative breast cancer," *PLoS ONE*, vol. 7, no. 10, Article ID e47993, 2012.

[32] L. Balabanski, G. Antov, I. Dimova et al., "Next-generation sequencing of *BRCA1* and *BRCA2* in breast cancer patients and control subjects," *Molecular and Clinical Oncology*, vol. 2, no. 3, pp. 435–439, 2014.

[33] V. D'Argenio, M. V. Esposito, A. Telese et al., "The molecular analysis of BRCA1 and BRCA2: next-generation sequencing supersedes conventional approaches," *Clinica Chimica Acta*, vol. 446, pp. 221–225, 2015.

[34] D. Sie, P. J. F. Snijders, G. A. Meijer et al., "Performance of amplicon-based next generation DNA sequencing for diagnostic gene mutation profiling in oncopathology," *Cellular Oncology*, vol. 37, no. 5, pp. 353–361, 2014.

[35] P. Concolino, A. Costella, A. Minucci et al., "A preliminary Quality Control (QC) for next generation sequencing (NGS) library evaluation turns out to be a very useful tool for a rapid detection of BRCA1/2 deleterious mutations," *Clinica Chimica Acta*, vol. 437, pp. 72–77, 2014.

[36] A. ElSharawy, J. Warner, J. Olson et al., "Accurate variant detection across non-amplified and whole genome amplified DNA using targeted next generation sequencing," *BMC Genomics*, vol. 13, article 500, 2012.

[37] C. A. Valencia, D. Rhodenizer, S. Bhide et al., "Assessment of target enrichment platforms using massively parallel sequencing for the mutation detection for congenital muscular dystrophy," *Journal of Molecular Diagnostics*, vol. 14, no. 3, pp. 233–246, 2012.

[38] D. S. Paul, P. Guilhamon, A. Karpathakis et al., "Assessment of raindrop BS-seq as a method for large-scale, targeted bisulfite sequencing," *Epigenetics*, vol. 9, no. 5, pp. 678–684, 2014.

[39] C. M. Watson, L. A. Crinnion, J. E. Morgan et al., "Robust diagnostic genetic testing using solution capture enrichment and a novel variant-filtering interface," *Human Mutation*, vol. 35, no. 4, pp. 434–441, 2014.

[40] E. Hodges, Z. Xuan, V. Balija et al., "Genome-wide in situ exon capture for selective resequencing," *Nature Genetics*, vol. 39, no. 12, pp. 1522–1527, 2007.

[41] A. Gnirke, A. Melnikov, J. Maguire et al., "Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing," *Nature Biotechnology*, vol. 27, no. 2, pp. 182–189, 2009.

[42] Z.-J. Liu, H.-F. Li, G.-H. Tan et al., "Identify mutation in amyotrophic lateral sclerosis cases using HaloPlex target enrichment system," *Neurobiology of Aging*, vol. 35, no. 12, pp. 2881.e11–2881.e15, 2014.

[43] E. Schaefer, P. Helms, L. Marcellin et al., "Next-generation sequencing (NGS) as a fast molecular diagnosis tool for left ventricular noncompaction in an infant with compound mutations in the MYBPC3 gene," *European Journal of Medical Genetics*, vol. 57, no. 4, pp. 129–132, 2014.

[44] The 1000 Genomes Project Consortium, G. R. Abecasis, D. Altshuler et al., "A map of human genome variation from population-scale sequencing," *Nature*, vol. 467, no. 7319, pp. 1061–1073, 2010.

[45] W. Fu, T. D. O'Connor, G. Jun et al., "NHLBI Exome Sequencing Project, Akey JM.Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants," *Nature*, vol. 493, no. 7431, pp. 216–220, 2013.

[46] J. Majewski, J. Schwartzentruber, E. Lalonde, A. Montpetit, and N. Jabado, "What can exome sequencing do for you?" *Journal of Medical Genetics*, vol. 48, no. 9, pp. 580–589, 2011.

[47] S. B. Ng, K. J. Buckingham, C. Lee et al., "Exome sequencing identifies the cause of a mendelian disorder," *Nature Genetics*, vol. 42, no. 1, pp. 30–35, 2010.

[48] J. K. Teer and J. C. Mullikin, "Exome sequencing: the sweet spot before whole genomes," *Human Molecular Genetics*, vol. 19, no. 2, pp. R145–R151, 2010.

[49] M. J. Clark, R. Chen, and M. Snyder, "Exome sequencing by targeted enrichment," *Current Protocols in Molecular Biology*, chapter 7:unit 7.12, 2013.

[50] Asan, Y. Xu, H. Jiang et al., "Comprehensive comparison of three commercial human whole-exome capture platforms," *Genome Biology*, vol. 12, article R95, 2011.

[51] K. Bodi, A. G. Perera, P. S. Adams et al., "Comparison of commercially available target enrichment methods for next-generation sequencing," *Journal of Biomolecular Techniques*, vol. 24, no. 2, pp. 73–86, 2013.

[52] A. Warr, C. Robert, D. Hume, A. Archibald, N. Deeb, and M. Watson, "Exome sequencing: current and future perspectives," *G3 (Bethesda)*, vol. 5, pp. 1543–1550, 2015.

[53] T. Walsh, H. Shahin, T. Elkan-Miller et al., "Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein GPSM2 as the cause of nonsyndromic hearing loss DFNB82," *American Journal of Human Genetics*, vol. 87, no. 1, pp. 90–94, 2010.

[54] A. Sirmaci, T. Walsh, H. Akay et al., "MASP1 mutations in patients with facial, umbilical, coccygeal, and auditory findings of carnevale, malpuech, OSA, and michels syndromes," *The American Journal of Human Genetics*, vol. 87, no. 5, pp. 679–686, 2010.

[55] O. Diaz-Horta, D. Duman, J. Foster II et al., "Whole-exome sequencing efficiently detects rare mutations in autosomal recessive nonsyndromic hearing loss," *PLoS ONE*, vol. 7, no. 11, Article ID e50628, 2012.

[56] H.-M. Woo, H.-J. Park, M.-H. Park et al., "Identification of CDH23 mutations in Korean families with hearing loss by whole-exome sequencing," *BMC Medical Genetics*, vol. 15, article 46, 2014.

[57] G. Girotto, K. Abdulhadi, A. Buniello et al., "Linkage study and exome sequencing identify a BDP1 mutation associated with hereditary hearing loss," *PLoS ONE*, vol. 8, no. 12, Article ID e80323, 2013.

[58] A. Bonnefond, E. Durand, O. Sand et al., "Molecular diagnosis of neonatal diabetes mellitus using next-generation sequencing of the whole exome," *PLoS ONE*, vol. 5, no. 10, Article ID e13630, 2010.

[59] S. Hillman, M. Smart, C. Bacchelli, L. Ocaka, and D. J. Williams, "2.4 whole exome sequencing of growth restricted offspring identifies gene variants implicated in maturity onset diabetes of the young," *Archives of Disease in Childhood: Fetal and Neonatal Edition*, vol. 99, article A2, 2014.

[60] M. Çalişkan, J. X. Chong, L. Uricchio et al., "Exome sequencing reveals a novel mutation for autosomal recessive non-syndromic mental retardation in the TECR gene on chromosome 19p13," *Human Molecular Genetics*, vol. 20, no. 7, Article ID ddq569, pp. 1285–1289, 2011.

[61] C. E. Prada, C. Gonzaga-Jauregui, R. Tannenbaum et al., "Clinical utility of whole-exome sequencing in rare diseases: galactosialidosis," *European Journal of Medical Genetics*, vol. 57, no. 7, pp. 339–344, 2014.

[62] A. Iglesias, K. Anyane-Yeboa, J. Wynn et al., "The usefulness of whole-exome sequencing in routine clinical practice," *Genetics in Medicine*, vol. 16, no. 12, pp. 922–931, 2014.

[63] M. Choi, U. I. Scholl, W. Ji et al., "Genetic diagnosis by whole exome capture and massively parallel DNA sequencing," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 45, pp. 19096–19101, 2009.

[64] R. W. K. Chiu, K. C. A. Chan, Y. Gao et al., "Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 51, pp. 20458–20463, 2008.

[65] A. J. Sehnert, B. Rhees, D. Comstock et al., "Optimal detection of fetal chromosomal abnormalities by massively parallel DNA sequencing of cell-free fetal DNA from maternal blood," *Clinical Chemistry*, vol. 57, no. 7, pp. 1042–1049, 2011.

[66] E. Z. Chen, R. W. K. Chiu, H. Sun et al., "Noninvasive prenatal diagnosis of fetal trisomy 18 and trisomy 13 by maternal plasma DNA sequencing," *PLoS ONE*, vol. 6, no. 7, Article ID e21791, 2011.

[67] D. W. Bianchi, R. Lamar Parker, J. Wentworth et al., "DNA sequencing versus standard prenatal aneuploidy screening," *The New England Journal of Medicine*, vol. 370, no. 9, pp. 799–808, 2014.

[68] Y. Song, C. Liu, H. Qi, Y. Zhang, X. Bian, and J. Liu, "Noninvasive prenatal testing of fetal aneuploidies by massively parallel sequencing in a prospective Chinese population," *Prenatal Diagnosis*, vol. 33, no. 7, pp. 700–706, 2013.

[69] M. M. Gil, M. S. Quezada, B. Bregant, M. Ferraro, and K. H. Nicolaides, "Implementation of maternal blood cell-free DNA testing in early screening for aneuploidies," *Ultrasound in Obstetrics and Gynecology*, vol. 42, no. 1, pp. 34–40, 2013.

[70] D. W. Bianchi, L. D. Platt, J. D. Goldberg, A. Z. Abuhamad, A. J. Sehnert, and R. P. Rava, "Genome-wide fetal aneuploidy detection by maternal plasma DNA sequencing," *Obstetrics and Gynecology*, vol. 119, no. 5, pp. 890–901, 2012.

[71] H. Yamamoto, Y. Watanabe, T. Maehata et al., "An updated review of gastric cancer in the next-generation sequencing era: insights from bench to bedside and vice versa," *World Journal of Gastroenterology*, vol. 20, no. 14, pp. 3927–3937, 2014.

[72] K. Wang, J. Kan, S. T. Yuen et al., "Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer," *Nature Genetics*, vol. 43, no. 12, pp. 1219–1223, 2011.

[73] N. F. C. C. de Miranda, K. Georgiou, L. Chen et al., "Exome sequencing reveals novel mutation targets in diffuse large B-cell lymphomas derived from Chinese patients," *Blood*, vol. 124, no. 16, pp. 2544–2553, 2014.

[74] I. Vater, M. Montesinos-Rongen, M. Schlesner et al., "The mutational pattern of primary lymphoma of the central nervous system determined by whole-exome sequencing," *Leukemia*, 2014.

[75] K. Snape, E. Ruark, P. Tarpey et al., "Predisposition gene identification in common cancers by exome sequencing: insights from familial breast cancer," *Breast Cancer Research and Treatment*, vol. 134, no. 1, pp. 429–433, 2012.

[76] H. Shi, G. Moriceau, X. Kong et al., "Melanoma whole-exome sequencing identifies ^{V600E}*B-RAF* amplification-mediated acquired B-RAF inhibitor resistance," *Nature Communications*, vol. 3, article 724, 2012.

[77] M. Kunz, M. Dannemann, and J. Kelso, "High-throughput sequencing of the melanoma genome," *Experimental Dermatology*, vol. 22, no. 1, pp. 10–17, 2013.

[78] A. Kumar, T. A. White, A. P. MacKenzie et al., "Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 41, pp. 17087–17092, 2011.

[79] J. Lindberg, D. Klevebring, W. Liu et al., "Exome sequencing of prostate cancer supports the hypothesis of independent tumour origins," *European Urology*, vol. 63, no. 2, pp. 347–353, 2013.

[80] J. G. Lohr, V. A. Adalsteinsson, K. Cibulskis et al., "Whole-exome sequencing of circulating tumor cells provides a window into metastatic prostate cancer," *Nature Biotechnology*, vol. 32, no. 5, pp. 479–484, 2014.

[81] A. Wesolowska, M. D. Dalgaard, L. Borst et al., "Cost-effective multiplexing before capture allows screening of 25 000 clinically relevant SNPs in childhood acute lymphoblastic leukemia," *Leukemia*, vol. 25, no. 6, pp. 1001–1006, 2011.

[82] R. Daneshjou, E. R. Gamazon, B. Burkley et al., "Genetic variant in folate homeostasis is associated with lower warfarin dose in African Americans," *Blood*, vol. 124, no. 14, pp. 2298–2305, 2014.

[83] M. Apellániz-Ruiz, M.-Y. Lee, L. Sánchez-Barroso et al., "Whole-exome sequencing reveals defective CYP3A4 variants predictive of paclitaxel dose-limiting neuropathy," *Clinical Cancer Research*, vol. 21, no. 2, pp. 322–328, 2015.

[84] A. Patwardhan, J. Harris, N. Leng et al., "Achieving high-sensitivity for clinical applications using augmented exome sequencing," *Genome Medicine*, vol. 7, article 71, 2015.

[85] L. Mamanova, A. J. Coffey, C. E. Scott et al., "Target-enrichment strategies for next-generation sequencing," *Nature Methods*, vol. 7, no. 2, pp. 111–118, 2010.

[86] F. Mertes, A. ElSharawy, S. Sauer et al., "Targeted enrichment of genomic DNA regions for next-generation sequencing," *Briefings in Functional Genomics*, vol. 10, no. 6, pp. 374–386, 2011.

[87] W. Li, J. Freudenberg, and P. Miramontes, "Diminishing return for increased Mappability with longer sequencing reads: implications of the k-mer distributions in the human genome," *BMC Bioinformatics*, vol. 15, article 2, 2014.

[88] L.-S. Chou, C.-S. J. Liu, B. Boese, X. Zhang, and R. Mao, "DNA sequence capture and enrichment by microarray followed by next-generation sequencing for targeted resequencing: neurofibromatosis type 1 gene as a model," *Clinical Chemistry*, vol. 56, no. 1, pp. 62–72, 2010.

[89] S. Dames, J. Durtschi, K. Geiersbach, J. Stephens, and K. V. Voelkerding, "Comparison of the illumina genome analyzer and roche 454 GS FLX for resequencing of hypertrophic cardiomyopathy-associated genes," *Journal of Biomolecular Techniques*, vol. 21, no. 2, pp. 73–80, 2010.

[90] D. Shigemizu, Y. Momozawa, T. Abe et al., "Performance comparison of four commercial human whole-exome capture platforms," *Scientific Reports*, vol. 5, article 12742, 2015.

[91] A. E. Shearer, A. P. DeLuca, M. S. Hildebrand et al., "Comprehensive genetic testing for hereditary hearing loss using massively parallel sequencing," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 49, pp. 21104–21109, 2010.

[92] Z. Brownstein, L. M. Friedman, H. Shahin et al., "Targeted genomic capture and massively parallel sequencing to identify genes for hereditary hearing loss in middle eastern families," *Genome Biology*, vol. 12, no. 9, article R89, 2011.

[93] A. H. Németh, A. C. Kwasniewska, S. Lise et al., "Next generation sequencing for molecular diagnosis of neurological disorders using ataxias as a model," *Brain*, vol. 136, no. 10, pp. 3106–3118, 2013.

[94] V. D'Argenio, G. Frisso, V. Precone et al., "DNA sequence capture and next-generation sequencing for the molecular diagnosis of genetic cardiomyopathies," *Journal of Molecular Diagnostics*, vol. 16, no. 1, pp. 32–44, 2014.

[95] M. E. Shanks, S. M. Downes, R. R. Copley et al., "Next-generation sequencing (NGS) as a diagnostic tool for retinal degeneration reveals a much higher detection rate in early-onset disease," *European Journal of Human Genetics*, vol. 21, no. 3, pp. 274–280, 2013.

[96] I. J. Nijman, J. M. van Montfrans, M. Hoogstraat et al., "Targeted next-generation sequencing: a novel diagnostic tool for primary immunodeficiencies," *Journal of Allergy and Clinical Immunology*, vol. 133, no. 2, pp. 529–534, 2014.

[97] G. D. Hariani, E. T. Lam, T. Havener et al., "Correction: application of next generation sequencing to CEPH cell lines to discover variants associated with FDA approved chemotherapeutics," *BMC Research Notes*, vol. 7, article 652, 2014.

[98] D. B. Johnson, K. H. Dahlman, J. Knol et al., "Enabling a genetically informed approach to cancer medicine: a retrospective evaluation of the impact of comprehensive tumor profiling using a targeted next-generation sequencing panel," *Oncologist*, vol. 19, no. 6, pp. 616–622, 2014.

[99] J. S. Ross, K. Wang, J. V. Rand et al., "Comprehensive genomic profiling of relapsed and metastatic adenoid cystic carcinomas by next-generation sequencing reveals potential new routes to targeted therapies," *American Journal of Surgical Pathology*, vol. 38, no. 2, pp. 235–238, 2014.

[100] H. K. Chong, T. Wang, H.-M. Lu et al., "The validation and clinical implementation of BRCAplus: a comprehensive high-risk breast cancer diagnostic assay," *PLoS ONE*, vol. 9, no. 5, Article ID e97408, 2014.

[101] G. J. W. Liao, K. C. A. Chan, P. Jiang et al., "Noninvasive prenatal diagnosis of fetal trisomy 21 by allelic ratio analysis using targeted massively parallel sequencing of maternal plasma DNA," *PLoS ONE*, vol. 7, no. 5, Article ID e38154, 2012.

[102] B. C. Lim, S. Lee, J.-Y. Shin et al., "Molecular diagnosis of congenital muscular dystrophies with defective glycosylation of alpha-dystroglycan using next-generation sequencing technology," *Neuromuscular Disorders*, vol. 23, no. 4, pp. 337–344, 2013.

[103] M. Hinchcliffe, H. Le, A. Fimmel et al., "Diagnostic validation of a familial hypercholesterolaemia cohort provides a model for using targeted next generation DNA sequencing in the clinical setting," *Pathology*, vol. 46, no. 1, pp. 60–68, 2014.

[104] L. Castéra, S. Krieger, A. Rousselin et al., "Next-generation sequencing for the diagnosis of hereditary breast and ovarian cancer using genomic capture targeting multiple candidate genes," *European Journal of Human Genetics*, vol. 22, no. 11, pp. 1305–1313, 2014.

[105] A. Fernández-Marmiesse, M. Morey, M. Pineda et al., "Assessment of a targeted resequencing assay as a support tool in the diagnosis of lysosomal storage disorders," *Orphanet Journal of Rare Diseases*, vol. 9, article 59, 2014.

[106] W. Wooderchak-Donahue, C. VanSant-Webb, T. Tvrdik et al., "Clinical utility of a next generation sequencing panel assay for Marfan and Marfan-like syndromes featuring aortopathy," *American Journal of Medical Genetics A*, vol. 167, no. 8, pp. 1747–1757, 2015.

[107] S. Dames, L.-S. Chou, Y. Xiao et al., "The development of next-generation sequencing assays for the mitochondrial genome and 108 nuclear genes associated with mitochondrial disorders," *Journal of Molecular Diagnostics*, vol. 15, no. 4, pp. 526–534, 2013.

[108] E. H. Kim, S. Lee, J. Park, K. Lee, J. Bhak, and B. C. Kim, "New lung cancer panel for high-throughput targeted resequencing," *Genomics & Informatics*, vol. 12, no. 2, pp. 50–57, 2014.

[109] A. Cortes, J. Field, E. A. Glazov et al., "Resequencing and fine-mapping of the chromosome 12q13-14 locus associated with multiple sclerosis refines the number of implicated genes," *Human Molecular Genetics*, vol. 22, no. 11, pp. 2283–2292, 2013.

[110] H. Stranneheim and A. Wedell, "Exome and genome sequencing: a revolution for the discovery and diagnosis of monogenic disorders," *Journal of Internal Medicine*, 2015.

[111] B. Royer-Bertrand and C. Rivolta, "Whole genome sequencing as a means to assess pathogenic mutations in medical genetics and cancer," *Cellular and Molecular Life Sciences*, vol. 72, pp. 1463–1471, 2015.

[112] A. Belkadi, A. Bolze, Y. Itan et al., "Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants," *Proceedings of the National Academy of Sciences*, vol. 112, no. 17, pp. 5473–5478, 2015.

[113] T. Katsila and G. P. Patrinos, "Whole genome sequencing in pharmacogenomics," *Frontiers in Pharmacology*, vol. 6, article 61, 2015.

[114] C. Mizzi, B. Peters, C. Mitropoulou et al., "Personalized pharmacogenomics profiling using whole-genome sequencing," *Pharmacogenomics*, vol. 15, no. 9, pp. 1223–1234, 2014.

[115] M. Nagasaki, J. Yasuda, F. Katsuoka et al., "Rare variant discovery by deep whole-genome sequencing of 1,070 Japanese individuals," *Nature Communications*, vol. 6, article 8018, 2015.

[116] C. Gilissen, J. Y. Hehir-Kwa, D. T. Thung et al., "Genome sequencing identifies major causes of severe intellectual disability," *Nature*, vol. 511, no. 7509, pp. 344–347, 2014.

[117] K. Yamaguchi, M. Komura, R. Yamaguchi et al., "Detection of APC mosaicism by next-generation sequencing in an FAP patient," *Journal of Human Genetics*, vol. 60, no. 5, pp. 227–231, 2015.

[118] M. A. Lalli, B. M. Bettcher, M. L. Arcila et al., "Whole-genome sequencing suggests a chemokine gene cluster that modifies age at onset in familial Alzheimer's disease," *Molecular Psychiatry*, vol. 20, pp. 1294–1300, 2015.

[119] T. Mikeska and J. M. Craig, "DNA methylation biomarkers: cancer and beyond," *Genes*, vol. 5, no. 3, pp. 821–864, 2014.

[120] T. De Meyer, E. Mampaey, M. Vlemmix et al., "Quality evaluation of methyl binding domain based kits for enrichment DNA-methylation sequencing," *PLoS ONE*, vol. 8, no. 3, Article ID e59068, 2013.

[121] H. G. Hernández, M. Y. Tse, S. C. Pang, H. Arboleda, and D. A. Forero, "Optimizing methodologies for PCR-based DNA methylation analysis," *BioTechniques*, vol. 55, no. 4, pp. 181–197, 2013.

[122] A. L. Brunner, D. S. Johnson, W. K. Si et al., "Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver," *Genome Research*, vol. 19, no. 6, pp. 1044–1056, 2009.

[123] T. A. Down, V. K. Rakyan, D. J. Turner et al., "A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis," *Nature Biotechnology*, vol. 26, no. 7, pp. 779–785, 2008.

[124] D. Serre, B. H. Lee, and A. H. Ting, "MBD-isolated genome sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome," *Nucleic Acids Research*, vol. 38, no. 2, pp. 391–399, 2010.

[125] O. Bogdanović, A. Fernández-Miñán, J. J. Tena, E. de la Calle-Mustienes, and J. L. Gómez-Skarmeta, "The developmental epigenomics toolbox: ChIP-seq and MethylCap-seq profiling of early zebrafish embryos," *Methods*, vol. 62, no. 3, pp. 207–215, 2013.

[126] B. A. Rodriguez, D. Frankhouser, M. Murphy et al., "Methods for high-throughput MethylCap-Seq data analysis," *BMC Genomics*, vol. 13, supplement 6, article S14, 2012.

[127] A. Meissner, A. Gnirke, G. W. Bell, B. Ramsahoye, E. S. Lander, and R. Jaenisch, "Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis," *Nucleic Acids Research*, vol. 33, no. 18, pp. 5868–5877, 2005.

[128] R. Lister, R. C. O'Malley, J. Tonti-Filippini et al., "Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*," *Cell*, vol. 133, no. 3, pp. 523–536, 2008.

[129] M. J. Booth, M. R. Branco, G. Ficz et al., "Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution," *Science*, vol. 336, no. 6083, pp. 934–937, 2012.

[130] S. K. Stewart, T. J. Morris, P. Guilhamon et al., "oxBS-450K: a method for analysing hydroxymethylation using 450K BeadChips," *Methods*, vol. 72, pp. 9–15, 2015.

[131] C. Stirzaker, P. C. Taberlay, A. L. Statham, and S. J. Clark, "Mining cancer methylomes: prospects and challenges," *Trends in Genetics*, vol. 30, no. 2, pp. 75–84, 2014.

[132] L. Cheng and Y. Zhu, "A classification approach for DNA methylation profiling with bisulfite next-generation sequencing data," *Bioinformatics*, vol. 30, no. 2, pp. 172–179, 2014.

[133] K. Mensaert, S. Denil, G. Trooskens, W. Van Criekinge, O. Thas, and T. De Meyer, "Next-generation technologies and data analytical approaches for epigenomics," *Environmental and Molecular Mutagenesis*, vol. 55, no. 3, pp. 155–170, 2014.

[134] F. Allum, X. Shao, F. Guénard et al., "Characterization of functional methylomes by next-generation capture sequencing identifies novel disease-associated variants," *Nature Communications*, vol. 6, article 7211, 2015.

[135] R.-L. Huang, F. Gu, N. B. Kirma et al., "Comprehensive methylome analysis of ovarian tumors reveals hedgehog signaling pathway regulators as prognostic DNA methylation biomarkers," *Epigenetics*, vol. 8, no. 6, pp. 624–634, 2013.

[136] C. Jerónimo, P. J. Bastian, A. Bjartell et al., "Epigenetics in prostate cancer: biologic and clinical relevance," *European Urology*, vol. 60, no. 4, pp. 753–766, 2011.

[137] J. Nakamura, T. Tanaka, Y. Kitajima, H. Noshiro, and K. Miyazaki, "Methylation-mediated gene silencing as biomarkers of gastric cancer: a review," *World Journal of Gastroenterology*, vol. 20, no. 34, pp. 11991–12006, 2014.

[138] S. Ghosh, F. Gu, C.-M. Wang et al., "Genome-wide DNA methylation profiling reveals parity-associated hypermethylation of FOXA1," *Breast Cancer Research and Treatment*, vol. 147, no. 3, pp. 653–659, 2014.

[139] K. Warton, V. Lin, T. Navin et al., "Methylation-capture and next-generation sequencing of free circulating DNA from human plasma," *BMC Genomics*, vol. 15, article 476, 2014.

[140] A. W. Walker, S. H. Duncan, P. Louis, and H. J. Flint, "Phylogeny, culturing, and metagenomics of the human gut microbiota," *Trends in Microbiology*, vol. 22, no. 5, pp. 267–274, 2014.

[141] A. W. Janssen and S. Kersten, "The role of the gut microbiota in metabolic health," *The FASEB Journal*, vol. 29, no. 8, pp. 3111–3123, 2015.

[142] K. Faust, L. Lahti, D. Gonze, W. M. de Vos, and J. Raes, "Metagenomics meets time series analysis: unraveling microbial community dynamics," *Current Opinion in Microbiology*, vol. 25, pp. 56–66, 2015.

[143] I. Sekirov, S. L. Russell, L. C. M. Antunes, and B. B. Finlay, "Gut microbiota in health and disease," *Physiological Reviews*, vol. 90, no. 3, pp. 859–904, 2010.

[144] A. D. Kostic, R. J. Xavier, and D. Gevers, "The microbiome in inflammatory bowel disease: current status and the future ahead," *Gastroenterology*, vol. 146, no. 6, pp. 1489–1499, 2014.

[145] Y. Goto, Y. Kurashima, and H. Kiyono, "The gut microbiota and inflammatory bowel disease," *Current Opinion in Rheumatology*, vol. 27, no. 4, pp. 388–396, 2015.

[146] A. J. Cox, N. P. West, and A. W. Cripps, "Obesity, inflammation, and the gut microbiota," *The Lancet Diabetes & Endocrinology*, vol. 3, no. 3, pp. 207–215, 2015.

[147] C. Sanmiguel, A. Gupta, and E. A. Mayer, "Gut microbiome and obesity: a plausible explanation for obesity," *Current Obesity Reports*, vol. 4, no. 2, pp. 250–261, 2015.

[148] H. Yu, H. Meng, F. Zhou, X. Ni, S. Shen, and U. N. Das, "Urinary microbiota in patients with prostate cancer and benign prostatic hyperplasia," *Archives of Medical Science*, vol. 11, no. 2, pp. 385–394, 2015.

[149] C. Weber, "Nutrition: diet change alters microbiota and might affect cancer risk," *Nature Reviews Gastroenterology & Hepatology*, vol. 12, article 314, 2015.

[150] V. D'Argenio and F. Salvatore, "The role of the gut microbiome in the healthy adult status," *Clinica Chimica Acta*, 2015.

[151] C. K. Lee, C. W. Herbold, S. W. Polson et al., "Groundtruthing next-gen sequencing for microbial ecology-biases and errors in community structure estimates from PCR amplicon pyrosequencing," *PLoS ONE*, vol. 7, no. 9, Article ID e44224, 2012.

[152] V. D'Argenio, V. Precone, G. Casaburi et al., "An altered gut microbiome profile in a child affected by Crohn's disease normalized after nutritional therapy," *American Journal of Gastroenterology*, vol. 108, no. 5, pp. 851–852, 2013.

[153] J. K. Harris, R. Fang, B. D. Wagner et al., "Esophageal microbiome in eosinophilic esophagitis," *PLoS ONE*, vol. 10, no. 5, Article ID e0128346, 2015.

[154] B. Coburn, P. W. Wang, J. Diaz Caballero et al., "Lung microbiota across age and disease stage in cystic fibrosis," *Scientific Reports*, vol. 5, Article ID 10241, 2015.

[155] O. J. Park, H. Yi, J. H. Jeon et al., "Pyrosequencing analysis of subgingival microbiota in distinct periodontal conditions," *Journal of Dental Research*, vol. 94, no. 7, pp. 921–927, 2015.

[156] J. C. Marioni, C. E. Mason, S. M. Mane, M. Stephens, and Y. Gilad, "RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays," *Genome Research*, vol. 18, no. 9, pp. 1509–1517, 2008.

[157] S. Zhao, W.-P. Fung-Leung, A. Bittner, K. Ngo, and X. Liu, "Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells," *PLoS ONE*, vol. 9, no. 1, Article ID e78644, 2014.

[158] N. E. Ilott and C. P. Ponting, "Predicting long non-coding RNAs using RNA sequencing," *Methods*, vol. 63, no. 1, pp. 50–59, 2013.

[159] M. Yang, L. Xu, Y. Liu, and P. Yang, "RNA-seq uncovers SNPs and alternative splicing events in Asian Lotus (*Nelumbo nucifera*)," *PLoS ONE*, vol. 10, no. 4, Article ID e0125702, 2015.

[160] C. A. Maher, C. Kumar-Sinha, X. Cao et al., "Transcriptome sequencing to detect gene fusions in cancer," *Nature*, vol. 458, no. 7234, pp. 97–101, 2009.

[161] GTEx Consortium, "Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans," *Science*, vol. 348, pp. 648–660, 2015.

[162] M. Melé, P. G. Ferreira, F. Reverter et al., "Human genomics. The human transcriptome across tissues and individuals," *Science*, vol. 348, no. 6235, pp. 660–665, 2015.

[163] F. W. Albert and L. Kruglyak, "The role of regulatory variation in complex traits and disease," *Nature Reviews Genetics*, vol. 16, no. 4, pp. 197–212, 2015.

[164] G. R. Oliver, S. N. Hart, and E. W. Klee, "Bioinformatics for clinical next generation sequencing," *Clinical Chemistry*, vol. 61, no. 1, pp. 124–135, 2015.

[165] A. S. Gargis, L. Kalman, D. P. Bick et al., "Good laboratory practice for clinical next-generation sequencing informatics pipelines," *Nature Biotechnology*, vol. 33, no. 7, pp. 689–693, 2015.

[166] S. Richards, N. Aziz, S. Bale et al., "Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology," *Genetics in Medicine*, vol. 17, no. 5, pp. 405–424, 2015.

[167] J. G. Caporaso, J. Kuczynski, J. Stombaugh et al., "QIIME allows analysis of high-throughput community sequencing data," *Nature Methods*, vol. 7, no. 5, pp. 335–336, 2010.

[168] V. D'Argenio, G. Casaburi, V. Precone, and F. Salvatore, "Comparative metagenomic analysis of human gut microbiome composition using two different bioinformatic pipelines," *BioMed Research International*, vol. 2014, Article ID 325340, 10 pages, 2014.

[169] B. Fosso, M. Santamaria, M. Marzano et al., "BioMaS: a modular pipeline for Bioinformatic analysis of Metagenomic AmpliconS," *BMC Bioinformatics*, vol. 16, article 203, 2015.

[170] M. D'Antonio, P. D'Onorio De Meo, M. Pallocca et al., "RAP: RNA-Seq analysis pipeline, a new cloud-based NGS web application," *BMC Genomics*, vol. 16, supplement 6, article S3, 2015.

[171] R. A. Calogero and F. Zolezzi, "Quantifying entire transcriptomes by aligned RNA-seq data," *Methods in Molecular Biology*, vol. 1269, pp. 163–172, 2015.

[172] H. L. Rehm, S. J. Bale, P. Bayrak-Toydemir et al., "ACMG clinical laboratory standards for next-generation sequencing," *Genetics in Medicine*, vol. 15, no. 9, pp. 733–747, 2013.

[173] R. Bao, L. Huang, J. Andrade et al., "Review of current methods, applications, and data management for the bioinformatics analysis of whole exome sequencing," *Cancer Informatics*, vol. 13, supplement 2, pp. 67–82, 2014.

[174] B. Liu, R. K. Madduri, B. Sotomayor et al., "Cloud-based bioinformatics workflow platform for large-scale next-generation sequencing analyses," *Journal of Biomedical Informatics*, vol. 49, pp. 119–133, 2014.

[175] Y. Xue, A. Ankala, W. R. Wilcox, and M. R. Hegde, "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*, vol. 17, no. 6, pp. 444–451, 2014.

[176] J. A. Reuter, D. V. Spacek, and M. P. Snyder, "High-throughput sequencing technologies," *Molecular Cell*, vol. 58, no. 4, pp. 586–597, 2015.

[177] J. Eid, A. Fehr, J. Gray et al., "Real-time DNA sequencing from single polymerase molecules," *Science*, vol. 323, no. 5910, pp. 133–138, 2009.

[178] E. W. Loomis, J. S. Eid, P. Peluso et al., "Sequencing the unsequenceable: expanded CGG-repeat alleles of the fragile X gene," *Genome Research*, vol. 23, no. 1, pp. 121–128, 2013.

[179] M. J. P. Chaisson, J. Huddleston, M. Y. Dennis et al., "Resolving the complexity of the human genome using single-molecule sequencing," *Nature*, vol. 517, pp. 608–611, 2015.

[180] M. O. Carneiro, C. Russ, M. G. Ross, S. Gabriel, C. Nusbaum, and M. A. DePristo, "Pacific biosciences sequencing technology for genotyping and variation discovery in human data," *BMC Genomics*, vol. 13, article 375, 2012.

[181] M. Jain, I. T. Fiddes, K. H. Miga, H. E. Olsen, B. Paten, and M. Akeson, "Improved data analysis for the MinION nanopore sequencer," *Nature Methods*, vol. 12, pp. 351–356, 2015.

[182] R. Drmanac, A. B. Sparks, M. J. Callow et al., "Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays," *Science*, vol. 327, no. 5961, pp. 78–81, 2010.