Tec Empresarial
P-ISSN: 1659-2395; E-ISSN: 1659-3359

ADVANCEMENTS IN NEXT-GENERATION SEQUENCING TECHNOLOGIES FOR GENETIC ANALYSIS

Saeed Jumaan Alzahrani[1]*, Ahmed Bashir Alrashidi[2], Abdullah Khalaf Alkhathami[3], Nouf Mohammed Alhussain[4], Hadeel Adel Alahmadi[5] and Hassan Abduh Ahmed Hazzazi[6]

1 Corresponding Author, Laboratory Specialist, Saeed_vip@hotmail.com, Al-Kharj Military Industries Corporation Hospital
2 Laboratory Specialist, A55680@hotmail.com, Al-Kharj Military Industries Corporation Hospital
3 Lab technician, Hokhok6996@gmail.com, Al-Kharj Military Industries Corporation Hospital
4 Technician-Laboratory, Alhussain.nouf2@gmail.com, Al Kharj Military Industries Corporation Hospital
5 Technician-Laboratory, doolycute@hotmail.com, Al Kharj Military Industries Corporation Hospital
6 Laboratory Specialist, Hassan_ss999@hotmail.com, Al-Kharj Military Industries Corporation Hospital

Abstract

Next-generation sequencing (NGS) technologies have emerged as powerful tools for analysis of DNA and RNA sequences. NGS platforms can generate millions of sequences in parallel, producing vast amounts of biological data. Recent advances in NGS technologies and applications for genomic analysis are discussed. Next-generation sequencing (NGS) technologies have revolutionized genomic research by enabling rapid, high-throughput sequencing of DNA and RNA at significantly reduced cost. NGS platforms can generate millions of sequences in parallel, producing vast amounts of biological data (Shekhar Pareek et al., 2011). NGS technologies have opened new avenues in basic research and numerous practical applications in agriculture, diagnostics, forensics, biomedicine, and other fields. This chapter reviews recent advances in NGS technologies and discusses different applications of NGS in genomic research, including de novo sequencing, comparative sequencing, resequencing, transcriptome analysis, and epigenomics. Next-generation sequencing (NGS) technologies have emerged as powerful tools for analysis of DNA and RNA sequences (Shekhar Pareek et al., 2011). NGS platforms can generate millions of sequences in parallel, producing vast amounts of biological data. This chapter briefly introduces NGS technologies and principles, reviews recent advances in NGS technologies, and discusses applications of NGS in genomic research.

1.2 Keywords

Keywords: Next-generation sequencing, technologies, applications, data analysis, DNA sequencing, genetic sequencing, electrophoresis, DNA polymerase chain reaction amplification, data interpretation

1.3 1. Introduction

Over the last decade, a variety of next-generation sequencing (NGS) technologies have been developed. The first such technology to receive FDA approval was the 454 Roche GS FLX system, which is based on pyrosequencing. Today's NGS systems utilize widely varying chemistries and approaches, including the massive parallelism (clustering) and optics used in the Illumina Solexa benchtop and HiSeq systems, as well as the SOLiD system, which uses a unique ligation-based approach. Semiconductor sequencing is being pursued by Ion Torrent and is similar to pyrosequencing. Single-molecule sequencing systems are also under development, including Pacific Biosciences (SMRT) and Oxford Nanopore sequencing technologies (Shekhar Pareek et al., 2011).

Each NGS platform has advantages and disadvantages in terms of read length, throughput, accuracy, and cost. Because these technologies are relatively new and because new platforms are being developed and commercialized, throughput and other specifications are constantly changing. In addition to NGS systems, there are still "first-generation" (Sanger) sequencers that can be used alongside NGS systems. Some applications may be best served by using two or more sequencing technologies. It is noteworthy that NGS technologies have been used successfully in many diverse applications since they became commercially available.

1.4 2. Historical Development of Next-Generation Sequencing Technologies

The innovations in next-generation sequencing technologies that have recently emerged are discussed in this review, with a specific focus on their applications for study design, library preparation, data generation, and bioinformatic analysis and interpretation for genetic analysis in plant research. Next-generation sequencing (NGS) technologies, also known as high-throughput sequencing, were first commercialized in 2005 by several biotechnology companies, and rapidly became widely used in many fields of molecular biology research, including environmental and metagenomic studies, whole genome sequencing and analysis of various organisms, transcriptome sequencing, marker discovery and genotyping, genome-assisted breeding, and resequencing studies of mutated lines or variants (Shekhar Pareek et al., 2011).

NGS platforms generate massive amounts of sequence data quickly and cost-effectively. Small to medium-sized genome plants, such as Arabidopsis thaliana, rice, and soybean, have been completely sequenced and analyzed with the assistance of Sanger sequencing or NGS technologies. The NGS technologies have triggered a genomics revolution in plant research, including whole genome sequencing, resequencing, and comparative genomic studies; transcriptome sequencing and analysis; genotyping and haplotyping of DNA polymorphic sites, quantitative trait loci, and fixed single nucleotide polymorphisms; small RNA sequencing and characterization; de novo genome and transcriptome assembly for species without existing genomic information; and methylome sequencing and analysis for DNA methylation studies. These applications are made possible by customized bioinformatics tools used to handle the massive NGS data, which also generate supplementary information for designing NGS experiments (Chen et al., 2023).

1.5 3. Principles of Next-Generation Sequencing

With the completion of the Human Genome Project in 2003, an era of genomics began, enabling an analysis of how the genome relates to biology and disease. Since then, the focus has been on low-cost, high-throughput sequencing technologies to analyze not just the human genome but also other species' genomes, transcriptomes, epigenomes, and microbiomes (Shekhar Pareek et al., 2011). Rapid advancements in sequencing technologies paved the way for next-generation sequencing (NGS) technologies. With the advent of NGS technologies, the cost per base sequenced is dramatically declining, enabling the genotyping of large sample sizes. The current sequencing technology can sequence a human genome in less than one day at a cost less than $1000. The various small genomes can now be sequenced in a matter of hours or days, enabling de novo genome assembly and annotation. Moreover, the same sequencing technology can be used to analyze transcriptomes, epigenomes, and resequencing applications. These competitive advantages have made NGS technologies popular in clinics and laboratories worldwide (R. Gullapalli et al., 2012). NGS technologies are relatively new bench-top technologies that are generally faster, cheaper, and smaller than traditional large-scale DNA sequencers. In NGS systems, millions of DNA molecules are sequenced simultaneously, massively parallelizing the sequencing process. A common feature of NGS technology involves the isolation of DNA of interest from biological samples, followed by the generation of an NGS library, a collection of DNA fragments with specific adaptor sequences ligated to their end. Before sequencing, libraries are amplified, either in vitro or in emulsion, creating millions of copies of each library fragment. Sequencing-by-synthesis cycles detect the incorporation of fluorescently labeled reversible terminator nucleotides, one type of DNA base at a time, generating sequence information for each fragment in the library. After imaging and cleavage of the labels, the cycle repeats for the next base addition. After sequencing, the output data is in the form of image files that need to be converted to text files for further sequence data analysis.

1.6 4. Key Components and Instrumentation in Next-Generation Sequencing

Next-generation sequencing (NGS), often called "high-throughput sequencing," refers to a number of different platforms that can rapidly sequence large numbers of DNA strands in parallel. It has become a mature technology in research laboratories around the world. NGS systems commercialized by Illumina, Ion Torrent, Roche, and others are commonly used in academic laboratories. NGS applications include whole-genome, whole-exome, and targeted sequencing for disease research; transcriptome (RNA-Seq) sequencing for gene expression profiling and discovery of non-coding RNAs; epigenomic assays such as ChIP-Seq, Bisulfite-Seq, and ATAC-Seq; and large-scale metagenomic studies of microbial communities ((A Pavlopoulos et al., 2013)). Sequencing techniques can be chronologically divided into three generations. The key principle of the first generation (Sanger or dideoxy) sequencing techniques was the use of dideoxy nucleotide triphosphates (ddNTPs) as DNA chain terminators. Dye-terminator sequencing utilized labeling in a single reaction. Dye-terminator sequencing combined with capillary electrophoresis succeeded in speeding up performance and became one of the most standardized and widely used techniques. Second-generation high-throughput sequencing techniques generate thousands or millions of short sequences at higher speed and better accuracy. High-throughput second-generation commercial technologies have already been developed by Illumina, Roche 454, and Biosystems/SOLiD. Today, Illumina is the most widely used platform. Recent HiSeq Illumina

systems make it possible for researchers to perform large and complex sequencing studies at a lower cost.

1.7 5. Applications of Next-Generation Sequencing in Genetic Analysis

Next-generation sequencing (NGS) technologies have become fundamental tools in research and clinical laboratories. In recent years, these powerful high-throughput platforms have been widely adopted for both whole-exome sequencing and genome-wide association studies. The rapid decline in the cost of sequencing has made DNA analysis affordable to many academic institutions, creating new opportunities for studying the genetic basis of various traits in model organisms, crops, and livestock (Di Resta et al., 2020). NGS technologies are experiencing rapid developments in research laboratories, and improvements in accuracy and throughput are expected in the next few years. NGS platforms enable massively parallel sequencing, producing millions of short reads simultaneously. This high-throughput capacity allows the determination of an entire genome in a single experiment at a relatively low cost. The introduction of NGS technologies has transformed the approach of studying genetics in non-model species, for which there is no complete genomic information available. Even simple NGS applications, such as transcriptome sequencing, can generate sufficient genomic information to examine genetic diversity and infer population demographic history. Small-scale projects utilizing NGS analysis and bioinformatics tools have successfully identified genomic variations and candidate genes associated with economically important traits.

1.8 6. Advantages and Limitations of Next-Generation Sequencing Technologies

Next-generation sequencing (NGS) platforms have emerged in response to the demand for rapid, low-cost, and large-scale genome sequencing. NGS technology allows simultaneous sequencing of millions of fragments, in contrast to the previous Sanger method's one-at-a-time approach. The fall in cost and time required for sequencing has encouraged large-scale genome sequencing projects at a national and international level.

NGS platforms include 454 pyrosequencing, Illumina, SOLiD, and ABI platforms. Though each of these platforms has its own mechanism for sequencing, they all rely on massively parallel sequencing of short fragments of the genome and the use of high-throughput imaging systems to record the sequences. NGS technologies have advantages, limitations, and challenges, especially in bioinformatics applications (Shekhar Pareek et al., 2011). In recent years, next-generation sequencing (NGS) technologies emerged as a relatively low-cost, comprehensive approach for clinical genetic testing. NGS platforms have been widely implemented in clinical laboratories for the genetic analysis of various disorders, including hereditary cancer, cardiomyopathies, and neurological diseases.

NGS has advanced screening capabilities, allowing for the simultaneous analysis of a large number of genes or entire gene panels. The reduced cost of the NGS approach has favored its adoption as the first-tier genetic test for many disorders. NGS has been implemented in clinical laboratories for the genetic analysis of various disorders, including hereditary cancer, cardiomyopathies, and neurological diseases. However, NGS has limitations that may affect the efficiency of data interpretation and the detection of clinically relevant variants (Di Resta et al., 2020). To reduce costs, clinical laboratories typically use in-house bioinformatics pipelines, which may lack the sensitivity and reliability of off-the-shelf software. The reliability of variant filtering in NGS data

interpretation strictly depends on the set of criteria used to prioritize the detected variants; inappropriate filtering strategies may lead to misclassifying pathogenic mutations as non-pathogenic, thus resulting in false-negative tests.

1.9 7. Recent Innovations and Emerging Trends in Next-Generation Sequencing

The rapidly developing field of next-generation sequencing technologies has drastically transformed research in life sciences and biomedicine, opening new windows of opportunity for understanding transcriptomic, genomic, epigenomic, and metagenomic events in different living systems. The past decade has witnessed remarkable improvements concerning NGS throughput, read length, accuracy, portability, and cost, which has furthered practical applications of NGS in comprehensive genetic analyses (Shekhar Pareek et al., 2011). Recent innovations, upgrades, and emerging trends in NGS technologies are discussed, focusing on introducing novel devices, library preparation methods, and bioinformatics tools that empower and facilitate broader implementation of NGS in genetic research and diagnostics. After the first introduction of commercially available high-throughput sequencing platforms, often referred to as second-generation sequencing systems, in the mid-2000s, life science research entered the NGS era. The second NGS technology wave commenced with early whole-genome shotgun sequencing projects for a variety of model organisms. Driven by Moore's law, NGS systems continuously improved their massively parallel sequencing capabilities. In the course of the past decade, yield per run increased more than one hundredfold while the cost for sequencing one million bases decreased by more than a thousandfold. Therefore, massively parallel NGS systems democratized access to genomic resources by enabling widespread de novo genome projects for various non-model organisms. NGS is no longer focused solely on genomic applications but encompasses different omics fields (A Pavlopoulos et al., 2013). Transcriptional profiling studies triggered the broader implementation of NGS technologies for transcriptomic analyses in both simple model organisms and complex tissues of non-model organisms.

1.10 8. Clinical Utility of Next-Generation Sequencing in Precision Medicine

The rapid adoption of next generation sequencing (NGS) in genomic medicine has been driven by the low cost and high throughput sequencing of the human genome, as well as the rapid advances in understanding the genetic bases of human diseases and traits (Dong et al., 2015). Today, the NGS method dominates the sequencing space in genomic research, and quickly entered clinical practice. At present, NGS is used in clinical laboratories for constitutional analysis, infectious disease typing and pathogen detection, and cancer variant detection, most often as "targeted" panels. Because unique features of NGS perfectly meet the clinical reality, the NGS technology is becoming a driving force to realize the dream of precision medicine. This article describes the strengths of NGS, NGS panels used in precision medicine, and current applications of NGS in cytology. Lastly, it discusses the challenges ahead and future directions of NGS for routine clinical use.

Precision medicine, or personalized medicine, is an emerging approach for disease prevention and treatment that takes individual variability into account (Hassouneh, 2014). With precision medicine, health care decisions, practices, and treatments will be tailored to the individual patient. Cancer is the most intensively studied disease in the context of precision medicine. One of the core ideas in precision medicine is to replace presently used empirical approaches to drug treatment of

diseases with a more rational strategy in which the patient's genomics will determine the treatment. To achieve "individual variability" requires analyzing multiple genes with little amounts of specimen inexpensively, quickly and sensitively. More recently, Sanger sequencing, the first generation sequencing technology, has been supplanted by the next generation sequencing (NGS) technology. NGS, sometimes referred to as massively parallel sequencing, is a group of high-throughput DNA sequencing technologies that were developed in the last decade. NGS has many advantages over traditional Sanger sequencing.

1.11 9. Ethical, Legal, and Social Implications of Next-Generation Sequencing

Ethical, legal, and social implications (ELSI) research considers the impact of new technologies on society. The advent of next-generation sequencing (NGS) technologies presents social, ethical, and legal challenges. DNA sequencing technologies make it possible to determine the order of nucleotides in DNA. High-throughput DNA sequencing enables massively parallel sequencing of millions of DNA fragments. NGS is often referred to as second-generation sequencing or post-Sanger sequencing technologies. The growth of NGS has been exponential. The decline in the cost of DNA sequencing resulted in the one-million-dollar human genome in 2001 to the 600-dollar human genome in 2016. The projected cost per exome and genome is 99 and 329 dollars, respectively, in 2020 (L. Blackburn et al., 2015). It makes NGS accessible for research and clinical purposes. Clinical NGS largely focuses on the analysis of germline mutations underlying genetic disorders. NGS has been transformative for genetic analysis in clinical diagnostics, translational research, and drug discovery. With these advancements come challenges and potential unintended consequences. In addition to the technical and data analysis hurdles, NGS raises social, ethical, and legal issues that may influence the technology adoption by end-users.

1.12 10. Challenges and Future Directions in Next-Generation Sequencing Technologies

Despite the impressive advancements and ubiquitous applications of next-generation sequencing (NGS) technologies, there remain some challenges to tackle and future directions to explore. For instance, Illumina sequencing may allow accurate, easy to use, and cost-effective small-RNA library generation for deep sequencing; but it may generate biased small-RNA libraries and is unsuitable for generating small-RNA libraries from low-quality total RNA. Currently, there are no small-RNA library preparation methods available that can overcome all the drawbacks of the existing methods. Exploring new strategies in RNA ligation and reverse transcription is desirable for developing a quick and easy small-RNA library preparation method that minimizes biases in small-RNA libraries and accepts inputs of various RNA qualities (Shekhar Pareek et al., 2011). NGS has contributed to advancements in neurogenetics research and clinical applications, but technical issues and challenges still need to be solved. NGS relies on polymerase chain reaction amplification, which does not efficiently amplify GC-rich genomic regions, where many disease-associated genes are located. While SNPs, or single nucleotide variants, can be detected using short reads, the identification of copy number variants is more challenging (Di Resta et al., 2020). Other biases have been also reported for both NGS and Sanger sequencing, and new methods should be further developed or optimized to reduce such biases. Third-generation sequencing platforms have been developed that generate long reads and can overcome some of the technical issues associated with second-generation sequencing. However, platforms such as PacBio SMRT and Oxford Nanopore Technologies still present other intrinsic limitations, such as high error rates and costs.

1.13 11. Comparative Analysis of Different Next-Generation Sequencing Platforms

The Sanger method of sequencing by capillary electrophoresis using the ABI 3730xL platform is considered the 'gold standard' in terms of read length and sequencing accuracy (Harismendy et al., 2009). However, it is also the most expensive method. With the ABI 3730xL, a DNA-base sequencer operating at 96 lanes per run, up to 126 million bases can be generated in a 3-hour run at a cost of $0.0027/base. In comparison, several next generation sequencing (NGS) technologies that generate significantly more sequence and are less expensive than the ABI Sanger method are now commercially available. These include Roche 454 (up to 1 million reads of 100 bases, $0.00074/base), Illumina GA (up to 2 billion reads of 36 bases, $0.00009/base), ABI SOLiD (up to 800 million reads of 50 bases, $0.0008/base) platforms. Each of these new NGS technologies has been successfully applied to different applications, including ChIP-sequencing, RNA-sequencing, and whole human genome sequencing. However, there is a limited understanding of NGS technology for population targeted sequencing study applications. This study was motivated by a recent genome-wide association study of coronary artery disease with a Native American population, for which custom 260 kb whole-exome capture design was generated. Because the NGS platforms available for use in this analysis were from three different manufacturers, it was important to determine the issues in generating and analyzing data produced by NGS platforms for population targeted sequencing studies. To this end, 260 kb of targeted sequence in four samples was generated using recommended sample library preparation methods, sequence generation, alignment tools, and base calling algorithms for Roche 454, Illumina GA, and ABI SOLiD platforms. The performance of each platform for these applications, as well as important issues in data analysis, is discussed.

1.14 12. Bioinformatics Tools and Data Analysis Pipelines for Next-Generation Sequencing Data

The advancement of next-generation sequencing (NGS) technologies has enabled rapid sequencing of millions of fragments of DNA. The NGS data consists of millions of short reads (and their quality scores) generated from a reference DNA template. The bioinformatics toolkits are designed to perform critical analyses and interpretation of the sequence data, allowing users to investigate DNA, RNA, or epigenome samples qualitatively and quantitatively. Potential applications include discovery of SNPs, indels, structural variants, and CNVs in DNA samples, transcript expression profiling in RNA samples, and experimental design and peak detection for ChIP-seq, Bisulfite-seq, and other epigenome assays (Cabello-Aguilar et al., 2023).

The basic architecture of the bioinformatics toolkits includes an “analysis” module that implements algorithms for interpreting sequencing data and a “data management” module that controls input/output files and serves metadata in an SQL relational database. Any additional modules are optional. To keep all the components well organized, a single directory is recommended as the home directory of the toolkit containing the above modules and example input/output files along with documentation. The directory can be compressed to prepare a distribution package. The toolkits can be deployed on personal computers, workstations, and service nodes of local computing clusters. They are compatible with Linux systems and have been tested on different flavors of Linux including Red Hat, Ubuntu, and SUSE (Corbett, 2015).

1.15 13. Integration of Next-Generation Sequencing with Other Omics Technologies

The integration of NGS methodologies into existing workflows or the development of new methods presents an opportunity to better utilize sequencing data in everyday applications. Several examples of NGS applications in the laboratory are highlighted here. These include metagenome analysis of microbial communities from biogas plants, targeting the genomes of the ectomycorrhizal fungus Laccaria bicolor, and novel methods for generating transcriptome data from long-established RNA collections. All applications take advantage of the widespread availability of NGS technologies and accessible bioinformatics tools (Vogl et al., 2013).

The 'Omics' suite of high-throughput technologies has the potential to interrogate biological systems at every level of functionality from genomic sequence to phenotypic output. With the maturation of next-generation sequencing (NGS) technology, the 'genomic' level is increasingly affordable and becoming routine. Like any biological system, mammalian cell factories are a product of their genome, transcriptome, proteome, and metabolome. They respond to and control their environment through the coordinated action of all biochemicals, critically affected by time and space. Understanding the relationship of each 'omic' layer to the system's biophysics, biology, and chemistry is the key to controlling it (Wright et al., 2016).

1.16 14. Case Studies and Success Stories in Genetic Analysis Using Next-Generation Sequencing

The advent of next-generation sequencing (NGS) technologies has transformed the landscape of genetic analysis, enabling researchers and clinicians to explore complex genetic information with unprecedented efficiency and accuracy. This section highlights a selection of noteworthy NGS success stories across various fields of research, emphasizing the technology's impact on unraveling genetic mysteries, diagnosing diseases, and advancing personalized medicine.

In recent years, the advent of NGS has opened a new era in neurogenetics. The second-generation sequencing is the cornerstone for molecular diagnosis for patients affected by heterogenous genetic diseases. The causative variant's identification in rare neurological diseases is a challenge for many clinical centres because, despite the efforts of national-networked laboratories, there are still a high number of undiagnosed cases. The massive parallel sequencing allows the screening of a large number of genes, exomes or the entire genome (Di Resta et al., 2020). With the advent of next generation sequencing (NGS), the panorama of genetic testing dramatically changed. NGS approach was first adopted in research laboratories, and now it is becoming widely used in clinical laboratories thanks to the reduction of cost of DNA sequencing. Many neurological phenotypes are characterized by a heterogenous genetic basis, and testing all known disease genes was not feasible. The advances and adoption of NGS in clinical molecular laboratories significantly increased the ability to identify the causative genes in diseased individuals. In parallel with the advances in NGS technology, the characterization of new causative genes responsible for many neurologic diseases was possible, providing new insight into their pathogenic mechanisms, and leading to advances in gene therapy and identification of new treatments for previously incurable diseases.

1.17 15. Educational and Training Programs in Next-Generation Sequencing Technologies

Educational and training programs in next-generation sequencing (NGS) technologies have been developed to meet the needs of researchers and medical professionals. A needs assessment at the University of Pittsburgh Medical Center investigated the data management needs of health sciences investigators using or planning to use NGS methods. The assessment revealed concerns related to managing sequencing data, specifically the volume of data, tools for accessing data, availability of trained staff, and lack of knowledge about options for data management and analysis. The assessment's findings can help other biomedical research institutions understand and plan for the impact of NGS technologies on research and pathology practice (Geskin et al., 2015). The traditional model of clinical medicine education has relied on bedside interaction with "tutor" physicians, hands-on exposure to instruments and techniques, and a gradual increase in independence. However, a paradigm shift in medical education may be necessary in the face of disruptive technologies such as NGS. As with other technologies, NGS is likely to be first adopted and then used to its full potential by forward-looking practitioners in the field (R. Gullapalli et al., 2012). Currently, there is a lack physician-friendly computational data analysis tools and structured training programs to train physicians in the use and interpretation of NGS and genomic data. At present efforts are being made to develop training programs that make data analysis interpretation NGS data routine for pathology residents and biomedical informatics graduate students.

1.18 16. Standardization and Quality Control in Next-Generation Sequencing

Next-generation sequencing (NGS)-based technologies are rapidly evolving, expanding their use beyond basic research into their utilisation for genetic testing in clinical/diagnostic laboratories. These developments have prompted the need to establish best practices for NGS applications in laboratory-developed tests (LDTs). LDTs are defined as in vitro diagnostic tests designed, developed, and validated by a single laboratory, which ensure that appropriate quality control procedures are in place to guarantee accuracy, reliability, and timeliness. The accreditation program ensures that all medical laboratories meet the minimum standards for conducting tests, which include LDTs. As NGS technologies have evolved over time, their use in medical and clinical laboratories has dramatically increased, from less than 1% in 2010 to more than 30% in 2014. While many diagnostic next-generation sequencing tests are under development, only a limited number are commercially available (Aziz et al., 2015).

Next-generation sequencing has the potential to affect health care by improving existing diagnostic tests or developing novel NGS-based tests, particularly for infectious disease or oncology applications. As the use of NGS in clinical laboratories continues to grow, it is crucial to establish best practices to ensure the safety and accuracy of testing. This article describes and summarises best practices for NGS genetic tests being performed or planned in a clinical setting. The goal is to facilitate understanding and implementation for laboratories wishing to develop NGS-based genetic tests.

1.19 17. Cost Analysis and Economic Impact of Next-Generation Sequencing Technologies

Sequencing technology has evolved at a fast pace, with declining costs-per-base sequenced. The genome sequenced for around 3 billion dollars for the HGP can now be sequenced for a few thousand dollars. This accelerating rate of decline in the cost of sequencing has been reverberating throughout domains connected with DNA sequencing, from ecology to forensics to medicine and everything in between. Several next generation sequencing (NGS) technologies have been developed since 2001, each with distinctive strengths and weaknesses. Currently, commercially available NGS technologies fall into three categories: (1) second generation massively parallel sequencing platforms that generate millions of reads in a single run, (2) bench-top second generation sequencers to accommodate clinical applications requiring less throughput, (3) third generation sequencers that produce reads longer than 10 kb (R. Gullapalli et al., 2012).

Intense competition among second generation DNA sequencing entities has driven down the cost per megabase of sequence produced by a staggering 60% a year from 2006-2010. There is a direct relationship between the throughput of a DNA sequencing platform and the resulting architecture's size and complexity. Unfortunately, size and complexity are usually at odds with user-friendliness. NGS technologies employ many unique and complex chemistry steps that require careful monitoring and control, as a precise environment is necessary for successful results. The original instrumentation developed for DNA sequencing had a research focus, resulting in devices capable of extraordinary amounts of DNA sequence output per single run. However, due to the large throughput, a single run is approximately 10 days per instrument. In response, companies introduced bench-top DNA sequencing instrumentation which can complete runs within a few hours. The sequencing throughput of benchtop instrumentation is far less compared to larger instruments, but for targeted clinical DNA sequence applications, this is unlikely to be a major impediment.

1.20 18. Regulatory Landscape and Approval Processes for Next-Generation Sequencing Tests

The broader implementation of next-generation sequencing (NGS) technology in clinical settings requires thoughtful discussion and planning by agencies responsible for the regulation of laboratory tests and devices. Presently, in the United States, the Food and Drug Administration (FDA) regulates the NGS hardware, software, and laboratory-developed tests within both commercial and academic clinical laboratories. Starting in late 2012, the next generation sequencing (NGS) technology began to be used by the clinical laboratories for the testing of several genetic diseases, or the use of NGS-based in vitro diagnostic tests (IVDs), as a result of the widely acceptance of NGS technology by the academic research labs and the drop in the cost and the size of NGS equipment. A draft discussion paper was published by the FDA in July, 2014, outlining the proposed regulatory approach to the oversight of NGS tests, including the key factors influencing the testing's regulatory tier classification and the importance of biosafety and bioinformatics. This discussion paper elicited considerable feedback from medical and scientific communities, as well as the public. The FDA responded to many comments on the pre-publication version of this overview by revising it for clarity and additional insights (Aziz et al., 2015).

In Europe, the implementation of the IVD Directive is the responsibility of the national authorities. Each member state of EU regulates IVDs according to their own national laws. Several member states have guidelines to assist laboratories and manufacturers in complying with the respective IVD Directive. For example, in France, a guidance document was published to help laboratories to comply with the IVD Directive, describing necessary steps to be undertaken by clinical laboratories before considering NGS tests as laboratory IVDs. In Germany, a national law implemented the IVD Directive and mandated that NGS tests developed by a laboratory are classified as IVDs.

1.21 19. Global Initiatives and Collaborative Research Efforts in Next-Generation Sequencing Technologies

The rapidly declining cost for next-generation sequencing (NGS) technologies and associated bioinformatics is creating vital opportunities for academic researchers, physicians, clinical laboratories, and pharmaceutical companies to generate and analyze genomic data at a large scale. It is essential to establish a flexible and robust framework for data generation, analysis, interpretation, and archiving along with harmonized standard operating procedures (SOPs). This enables academic and clinical research laboratories to set up the desired NGS applications with mono or multi-plexed PCR enriched barcoded libraries and share the SOPs. In the academic arena, tumor samples from patients are usually enriched for actionable mutations by targeted NGS in clinical sequencing. Where possible, somatic mutation calls from tumor samples are applied to their matched normal samples to filter out germline variants, and only variant calls retained from the ClinVar database are assessed for pathogenicity. Genomes are sequenced to 30-fold coverage while exomes and panels are sequenced to 100 and 1000-fold coverage respectively. A single run typically generates 400 Gb of sequencing data and at present, around 4800 whole genomes or 72,000 exomes can be resequenced per year (Chiara & Pavesi, 2017). Considering the whole genome 1000 project benchmarking in the context of current workflows, this discusses various issues in NGS data quality assessment and tuning of quality control (QC) tools, and recommends relevant tools and settings for optimal performance on NGS data from Illumina platforms. To date, almost 2200 NGS platforms comprising 18 different technologies have been cataloged worldwide. The vast majority up 1693 are Illumina genome analyzers sequencing by synthesis platforms that employ reversible dye terminators as sequencing chemistry. Alternative chemistry platforms include 152 Ion Torrent semiconductor sequencers, 87 SOLiD dye ligation sequencers, 6 Pacific Biosciences SMRT single-molecule zero-mode waveguides, and 63 systems from 7 additional vendors. Facilities employing NGS and Sanger technologies are present in 62 countries. Europe and the US currently account for 52% and 27% of NGS capacity, respectively, and China recently became the third major player with 10% (R. Gullapalli et al., 2012).

1.22 20. Conclusion and Future Prospects

Next-generation sequencing (NGS) systems have rapidly evolved over the past decade. The development of NGS has made sequencing widely available and affordable. It has changed the way genomics is used in research and in clinical settings. With NGS technology, almost any "-ome" can be studied, including genome, transcriptome, epigenome, microbiome, and exome. NGS has paved the way for personalized medicine, which relies on genomic analysis. NGS technology has been successfully implemented in the investigation of many diseases, including genetic disorders, cancers, and infectious diseases (Shekhar Pareek et al., 2011). NGS systems currently

available to the research community range from high-throughput sequencing with massively parallel sequencing systems to portable, low-throughput, and low-cost sequencing with nanopore sequencing systems. Each sequencing platform has its own chemistry, design, performance, and sequencing applications. Due to the differences in sequencing chemistry, the accuracy, throughput, read length, and cost-per-base of sequencing data vary widely across platforms. A careful evaluation of each available NGS technology is essential to determine its suitability for a specific application (Feng, 2018). The continued reduction in the cost-per-base of sequencing data is favorable for large-scale studies, including metagenomics and sequencing population- or species-level phylogeny. Furthermore, combined experimental and computational approaches to improving reference genomes will be critical for de novo assembly, especially for large and complex genomes. Continued development and improvement of bioinformatic pipelines for data processing, quality control, and analysis of non-model species that have no reference genomes will be necessary. The joint development of novel sequencing systems and accompanying sample preparation methods, data analysis pipelines, and bioinformatic tools is crucial for promoting genomic research.

References:

1. Shekhar Pareek, C., Smoczynski, R., & Tretyn, A. (2011). Sequencing technologies and genome sequencing. ncbi.nlm.nih.gov
2. Shekhar Pareek, C., Smoczyński, R., & Tretyn, A. (2011). Sequencing technologies and genome sequencing. [PDF]
3. Wolf, B., Kuonen, P., Dandekar, T., & Atlan, D. (2015). DNAseq Workflow in a Diagnostic Context and an Example of a User Friendly Implementation. ncbi.nlm.nih.gov
4. Chen, P., Sun, Z., Wang, J., Liu, X., Bai, Y., Chen, J., Liu, A., Qiao, F., Chen, Y., Yuan, C., Sha, J., Zhang, J., Xu, L. Q., & Li, J. (2023). Portable nanopore-sequencing technology: Trends in development and applications. ncbi.nlm.nih.gov
5. R. Gullapalli, R., V. Desai, K., Santana-Santos, L., A. Kant, J., & J. Becich, M. (2012). Next generation sequencing in clinical medicine: Challenges and lessons for pathology and biomedical informatics. ncbi.nlm.nih.gov
6. A Pavlopoulos, G., Oulas, A., Iacucci, E., Sifrim, A., Moreau, Y., Schneider, R., Aerts, J., & Iliopoulos, I. (2013). Unraveling genomic variation from next generation sequencing data. ncbi.nlm.nih.gov
7. Di Resta, C., Battista Pipitone, G., Carrera, P., & Ferrari, M. (2020). Current scenario of the genetic testing for rare neurological disorders exploiting next generation sequencing. ncbi.nlm.nih.gov
8. Dong, L., Wang, W., Li, A., Kansal, R., Chen, Y., Chen, H., & Li, X. (2015). Clinical Next Generation Sequencing for Precision Medicine in Cancer. ncbi.nlm.nih.gov
9. Hassouneh, R. (2014). Is Next-Generation Sequencing Appropriate for the Clinic?. [PDF]
10. L. Blackburn, H., Schroeder, B., Turner, C., D. Shriver, C., L. Ellsworth, D., & E. Ellsworth, R. (2015). Management of Incidental Findings in the Era of Next-generation Sequencing. ncbi.nlm.nih.gov

11. Harismendy, O., C Ng, P., L Strausberg, R., Wang, X., B Stockwell, T., Y Beeson, K., J Schork, N., S Murray, S., J Topol, E., Levy, S., & A Frazer, K. (2009). Evaluation of next generation sequencing platforms for population targeted sequencing studies. ncbi.nlm.nih.gov
12. Cabello-Aguilar, S., A. Vendrell, J., & Solassol, J. (2023). A Bioinformatics Toolkit for Next-Generation Sequencing in Clinical Oncology. ncbi.nlm.nih.gov
13. Corbett, A. (2015). Assessment of Alignment Algorithms, Variant Discovery and Genotype Calling Strategies in Exome Sequencing Data. [PDF]
14. Vogl, I., Benet-Pagès, A., H. Eck, S., Kuhn, M., Vosberg, S., A. Greif, P., H. Metzeler, K., Biskup, S., Müller-Reible, C., & Klein, H. G. (2013). Applications and data analysis of next-generation sequencing. [PDF]
15. Wright, C., Groot, J., & Estes, S. (2016). Biotherapeutic development in the u27Omics Age: The CHO genome and beyond. [PDF]
16. Geskin, A., Legowski, E., Chakka, A., R Chandran, U., Michael Barmada, M., A. LaFramboise, W., Berg, J., & S. Jacobson, R. (2015). Needs Assessment for Research Use of High-Throughput Sequencing at a Large Academic Medical Center. ncbi.nlm.nih.gov
17. Aziz, N., Zhao, Q., Bry, L., K. Driscoll, D., Funke, B., S. Gibson, J., W. Grody, W., R. Hegde, M., A. Hoeltge, G., G. B. Leonard, D., D. Merker, J., Nagarajan, R., A. Palicki, L., S. Robetorye, R., Schrijver, I., E. Weck, K., & V. Voelkerding, K. (2015). College of American Pathologistsu27 Laboratory Standards for Next-Generation Sequencing Clinical Tests. [PDF]
18. Chiara, M. & Pavesi, G. (2017). Evaluation of quality assessment protocols for high throughput genome resequencing data. [PDF]
19. Feng, K. (2018). Comprehensive Sequencing with Surface Tagmentation Based Technology. [PDF]