



Current Status and Challenges of DNA Sequencing

5

Indra Mani

Abstract

This chapter analyzes the DNA sequencing as its core technology impact to establish a relatively new scientific discipline as genomics to metagenomics and transcriptomics to metatranscriptomics. Due to the development of technology from conventional Sanger method to next-generation sequencing (NGS) techniques, it has transformed the genomics. NGS has improved the expense, scale of characterization of genome, and has significantly expanded the possibility of an investigation. Enormously comparable sequencing has demonstrated innovative bioinformatics tools, changing the assumption of genomics to address life issues at a genome- level. Emerging and established evidence of various genomic databases have demonstrated that the NGS has played a great role in the establishment of the whole-genome sequence (WGS) and whole exome sequence (WES). Currently, sequencing allows disease investigation like prognosis (mutational status), diagnosis (DNA and RNA-based), treatment (therapeutic identification), and molecular phylogeny. Preferably, sequencing should be precise, rapid, and cover long read. The elucidation of genomes from NGS platforms needs to be programmed and completely interconnected. In addition, the curation of data is very much necessary and helpful in an accurate annotation of the genome. Though retaining uniformity and precision in genome annotation is a challenging issue. Without annotation of the genome, sequencing data is meaningless. Consequently, appropriate reasonable investigation requires a constant and broad opinion of biological data.

Keywords

Next-generation sequencing (NGS) · Sanger sequencing · DNA · Genome · PCR

I. Mani (✉)

Department of Microbiology, Gargi College, University of Delhi, New Delhi, India

5.1 Introduction

The invention of DNA sequencing has been about 42 years, and the technologies are continuing to advance in the upcoming eras. Due to the continuous development of DNA sequencing technology and understanding of the genome, it has established comparatively a novel scientific discipline as genomics. For studies of genomics, purification of DNA is an essential step to further manipulation, including sequencing, restriction digestion, ligation, transformation, mutagenesis, and construction of probes (for hybridization). The utilization of different cloning vectors (M13mp or pUC) in *Escherichia coli* (*E. coli*) to increase the copy number (amplification) of DNA is very useful (Saiki et al. 1988). In addition, bacteriophage such as M13 is also very potential for these purposes (Messing 2001). A technique for the semi-automation of DNA sequence evaluation has been established, during an initial improvement of DNA sequencing. For each of the reactions particular for the bases A, C, G, and T, a different colored fluorophore has been used (Smith et al. 1986).

The whole-genome random sequencing and assembly have been utilized to determine the entire genome sequence (580,070 base pairs) of the *Mycoplasma genitalium*. It was the minimum known genome of any free-living organism. After comparative evaluation of this genome with *Haemophilus influenzae*, it has been suggested that the variances in genome content reflected as great differences in metabolic capacity and physiology of these two microbes (Fraser et al. 1995; Fleischmann et al. 1995). Enormously comparable sequencing has been established as novel and changing the hypothesis of genomics helps to understand the biology of organisms at a genome level. Due to the rapid development of genomics, it has revolutionized the biomedical research and clinical medicine (Shendure et al. 2017). Remarkably, DNA sequencing has been utilized in like prognosis (mutational status), diagnosis (DNA and RNA-based), treatment (therapeutic identification), and molecular phylogeny (Marian 2011; Koboldt et al. 2013; Smith 2017; Hu et al. 2019; Won et al. 2019).

Recently, for bacteria, algae, fungi, and protozoa, a DNA barcode has been developed. Remarkably, it was associated with error-free and rapid species identification that supported in understanding the microbial species which involve in a particular disease and microbial diversity (Chakraborty et al. 2014). Similarly, next-generation sequencing has facilitated the clinical metagenomic (culture-independent approach) and plays a vital role in research laboratories to clinical applications. This developing technology helps to vary diagnosis and treatments of communicable and noncommunicable (like cancer, diabetes, heart attack, etc.) diseases (Chiu and Miller 2019). In the oncology discipline, methylated-DNA sequencing tools are yielding great amounts of methylome records from cancer samples, from which cancer-associated differentially methylated CpG sites (cDMCs) have constantly recognized and filed. The addition of as numerous cDMCs as likely helps advance the precision of cancer examination and occasionally identify cancer subtypes. Nevertheless, the absence of a well-known technique for the analysis of 100 s of cDMCs normally hinders their vigorous practice in treatment (Jeon et al.

2017). Moreover, high-throughput sequencing (HTS) is progressively essential in defining cancer diagnoses, with subsequent prognostic and therapeutic implications (Guillermin et al. 2018). Presently, NGS plays a significant role to understand the gene expression in normal and cancer cells (Craig et al. 2016). It was possible because of the development of NGS technologies.

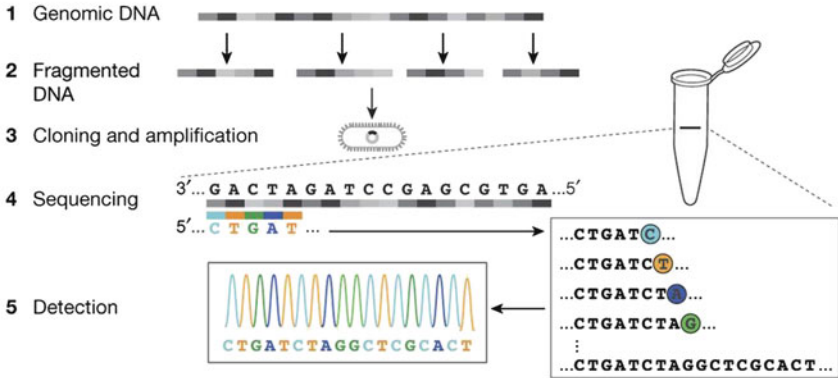
Recently, Oxford Nanopore sequence technology, MinION, has been developed, and it has improved the other technical requirements and is portable for onsite sequencing. This technology has been applicable in various fields like clinical medicine, environmental sphere, and biosecurity (Runtuwene et al. 2019). In contrast, for quantitative gene expression analyses, the RNA-Seq technology has been used as a widespread methodology. Further, it has also helped in the annotation of the transcriptome in gene expression study under various conditions (Blow 2009; Roberts et al. 2011; Garalde et al. 2016). Nevertheless, precise gene expression estimation needs exact genome information. Due to the development of technology from conventional Sanger method (Fig. 5.1a) to next-generation sequencing (NGS) techniques, it has significantly helped to understand genomics, transcriptomics, metagenomics, and metatranscriptomics Fig. 5.1.

5.2 Next-Generation Sequencing

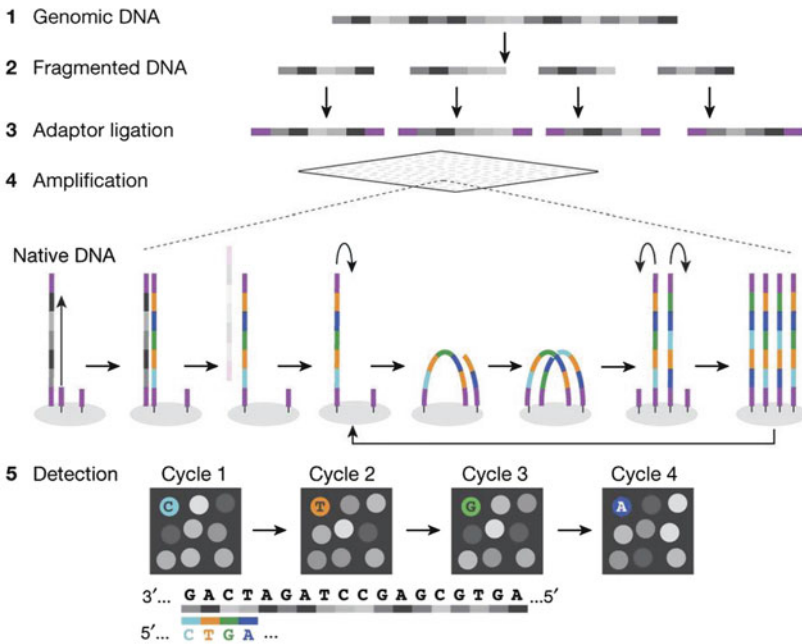
Modern DNA sequencing technology is commonly known with different names, such as next-generation sequencing (NGS), deep sequencing, enormously comparable, second generation, or third generation (Fig. 5.1b, c). However, the previous Sanger sequencing technology is considered as the conventional or the first generation sequencing, which has taken a decade to complete the final draft of the human genome. Presently, NGS is mostly utilized in the research work (Behjati and Tarpey 2013). An alternate method of polymerase chain reaction (PCR) colonies, or polonies, has been used to amplify for a single template molecule. At a time, it can amplify the millions of clones (Mitra and Church 1999).

Recently, the proliferation of genome sequencing projects has encouraged a search for alternate approaches to decrease time and expense. Toward accomplishing about 100-fold expansion in throughput over Sanger sequencing technology, it has established an emulsion method for DNA amplification and an instrument for sequencing by synthesis using a pyrosequencing procedure enhanced for solid support and picoliter-scale capacities. The pyrosequencing technique has demonstrated the efficiency, accuracy, throughput, and robustness of the system through de novo assembly and shotgun sequencing of the *Mycoplasma genitalium* genome with 96% coverage at 99.96% accuracy (Margulies et al. 2005). Currently, the virome analysis relies on deep sequencing, NGS data, and nucleic acid databases like the National Center for Biotechnology Information (NCBI), DNA Data Bank of Japan (DDBJ), and European Molecular Biology Laboratory Report (EMBL) databases. There are two frequently used NGS platforms: Illumina and Ion Torrent, which advocate maximum fragment lengths (300 to 400 nucleotides) for assessment separately. Recently, NGS technology has been enhanced with real-time sequencing

A First generation sequencing (Sanger)



B Second generation sequencing (massively parallel)



C Third generation sequencing (Real-time, single molecule)

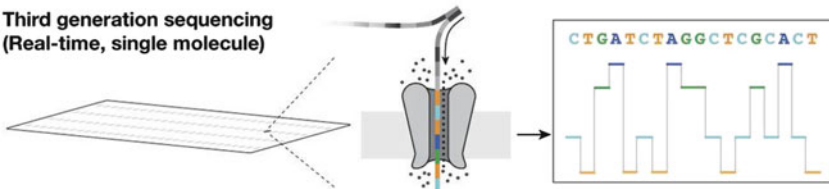


Fig. 5.1 DNA sequencing technologies. (a) Schematic examples of first, (b) second and (c) third generation sequencing are shown. Second generation sequencing is also referred to as next-generation sequencing (NGS) in the text (Shendure et al. 2017. Adopted with permission)

at single-molecule level (third generation) (Fig. 5.1c). It has indicated that the sequencing at single-molecule level will help in the advancement of various fields of biology, including virology, cancer biology, metagenomics, transcriptomics, and bioremediation (Ramamurthy et al. 2017; Schloss and Handelsman 2004; Bharagava et al. 2019; Jaswal et al. 2019; Costeira et al. 2019; Goto et al. 2019).

The advancement in DNA sequencing technology, NGS method, has supported to understand the microbiome diversity and also its characterization at a molecular level (Zaura 2012; Cao et al. 2017). The software developed for conventional sequencing technologies is becomes frequently incompetent to agree with nature of NGS technologies, which produce small and massively parallel reads. It has shown an amplification-free approach for analysis of the nucleotide sequences of about 0.3 million individual DNA molecules concurrently (Stein 2011; Harris et al. 2008). These findings have helped to understand the comparative analysis of human genomics.

5.3 Application of NGS in Various Fields

The NGS produces an amazing vision into metagenomes, metatranscriptomes, and metabarcodes of various organisms such as viruses, archaea, bacteria, and eukaryotes (Faure and Joly 2015; Bruno et al. 2015). The NGS tools have established the competence to sequence DNA at a unique speed, thereby allowing previously thought inconceivable scientific attainments and new biological applications. However, a significant challenge for data storage, analyses, and management solutions has been arising due to enormous data generated by the NGS (Zhang et al. 2011). Innovative bioinformatics tools are important for the effective management of NGS data. Through exploiting NGS approaches, investigators have identified and analyzed the important genes (Bai et al. 2012).

Multiplex ligation-dependent probe amplification and Sanger sequencing have been utilized for the genetic analysis of BREast CANcer genes 1 and 2 (BRCA1/2) which include the assessment for single nucleotide variants and insertion/deletion and for larger copy number variations (CNVs). Due to the introduction of NGS, it has become possible to specify the CNV information and sequence data (Schmidt et al. 2017). On the other hand, the NGS technique has been utilized to understand the molecular mechanism involved in gene regulation in hypertension. A remarkable study has identified the several genes' loci associated with different cardiovascular diseases (Costa and Franco 2017). The influence of NGS technologies on genomics will be very useful in the understanding of such type of complex diseases.

With the introduction of NGS tools, genome sequencing has become reasonable for regular genetic analyses. It has much facilitated the understanding of pathways involved in disease progression and analyses of rare genetic information of complex traits from the large datasets (Weissenkampen et al. 2019). The NGS has significantly extended our knowledge and skill to identify and characterize the gene and genetic composition from the microbial communities. Moreover, it has facilitated the analysis of microbiome with the help of metagenomics (culture-independent)

approaches (Song et al. 2013). In the metagenomics, total genetic material is extracted from total communities and processed and sequenced simultaneously.

Interestingly, NGS has revealed the metabolic interaction between gut microbiome and host. Genome-scale modeling is an evolving method which has established the various relations (microbe–microbe, host–microbe) under biological environments (Sen and Orešič 2019). However, the mechanisms of these interactions are still unclear. Gradual advancement of NGS tools has helped to understand the transcriptomics and metatranscriptomics (Tarkkonen et al. 2017). Furthermore, RNA-Seq data allow the analysis of transcriptome in the absence of a reference genome (de novo assembly) (Saggese et al. 2018). Due to large datasets generated from NGS tools, sophisticated *in silico* tools and skilled person are required.

Soil microbial communities are directly affected through natural environmental conditions, and functions are also fluctuating (Barboza et al. 2018). Taxonomic and functional profiles of soil samples can be analyzed by NGS approaches. On the other hand, bioremediation is usually observed as one of the effective methods to clean the environment with the help of microorganisms, instead of conventional physical and chemical methods. Due to an emerging concept of metagenomics, it can be utilized to understand the active microbial species, beneficial genes, enzymes, and bioactive molecules from the particular environmental sample. Such microbial species or specific gene can be utilized for effective bioremediation of the particular biohazardous compounds (Marco 2008; Ju and Zhang 2015; Czaplicki and Gunsch 2016; Conrads and Abdelbary 2019).

In addition, the series of omics, like metagenomics, metatranscriptomics, metaproteomics, metabolomics, and fluxomics, is also being utilized in the characterization, identification, and selection of particular strain of microbes (Schloss and Handelsman 2004; Bharagava et al. 2019). Therefore, a multi-omics approach, like metagenomics, metatranscriptomics, metaproteomics, and metabolomics, provides an excellent way to understand the metabolic pathway and microbes, which are involved in the bioremediation of particular contaminants in the environmental sites. Utilizing these approaches in the establishment of microbial consortium may be useful to provide specific microbial strain for the degradation of a particular contaminant from the environments.

5.4 Challenges of DNA Sequencing

Advancement in DNA sequencing technology like NGS and availability of international nucleotide sequence database collaboration (INSDC), and *in silico* tools (Softwares) significantly help to rapidly generate the genome sequences and understand the functional genomics of any organism. However, the analysis of NGS data requires sophisticated *in silico* tools, expensive infrastructure, and skilled person (Iacoangeli et al. 2019). The interpretation and characteristic of the NGS data have been challenged at several steps; such as sequencing errors, storage, algorithm, and statistical analyses (Zaura 2012).

In the recent past, rapid technological developments directed by academic institutions and companies are continuous to extend NGS methods from basic research to the clinical applications. However, the NGS implementation offers various process such as sequence analysis, storage, and quality control (Xuan et al. 2013; Vallenet et al. 2017). Genome-wide association studies (GWAS) have developed an applied technique to identify the genetic loci associated with disease by examining numerous markers throughout the genome. Ultimately, NGS has gained increasing popularity in the current years through its ability to analyze a much larger number of markers throughout the genome. Though NGS platforms have accomplished examining a higher number of single nucleotide polymorphisms (SNPs) associated with GWA studies (GWAS) (Alonso et al. 2015). In the case of GWAS, trivial effects started large sample sizes, usually made possible through meta-analysis by exchanging summary statistics throughout consortia. While NGS studies groupwise test for the association of multiple potentially causal alleles by every gene. Therefore, they have developed MetaSeq, a procedure for meta-analysis of genome-wide sequencing data, and it is publicly available as open source (Singh et al. 2013). The results gained by the NGS need thorough analysis, as their biological relationship is not well understood.

Numerous emerging biological applications, such as targeted exome sequencing, chromatin immunoprecipitation sequencing (ChIP-Seq), and whole transcriptome shotgun sequencing technology or RNA-Seq, have been established to accomplish various biological determinations. Exome sequencing (Mamanova et al. 2010) affects the disadvantage of the high expense of sequencing the whole genome without intronic regions, and selectively sequencing the exonic regions, which might be of further direct interest. ChIP-Seq (Johnson et al. 2007) is used to study protein–DNA/RNA interactions while RNA-Seq (Mortazavi et al. 2008) is used to exploit the NGS technologies to sequence cDNAs. The gathering of NGS reads leaves over a puzzling assignment. It is particularly true for the assembly of environmental samples that originate through metagenomics approaches possibly containing huge microbial diversity (Warnke-Sommer and Ali 2016). Nevertheless, due to lack of availability of the reference genome, it creates problems while executing the conclusion of analyzed data (Hiendleder et al. 2005). Moreover, NGS has facilitated the understanding of microbiomes associated with infection, environmental, bioremediation, and diversity of agriculturally important microorganisms. Recently, solid-state nanopore-based NGS has been developed. It has demonstrated that advancement in the technology, such as the processability, the robustness, and the large-scale integratability (Goto et al. 2019) and other sequencing platforms, is under development.

5.5 Concluding Remarks

This chapter analyzes an innovation in the DNA sequencing technology, like second and third generation sequencing techniques, which has demonstrated the efficiency, accuracy, throughput, and robustness of the system. In addition, it is

significantly helps to generate rapid and large datasets, like metagenome and metatranscriptome. Moreover, the availability of various biological databases, such as NCBI-GenBank, DDBJ, EMBL-EBI (European Bioinformatics Institute), and in silico tools (Softwares), considerably helps to understand the functional genomics of the organisms. Exploiting the NGS technology and databases in the field of microbiology, metagenomics molecular method is one of the powerful culture-independent approaches to understand the microbial diversity and mine a specific gene from any sample such as environmental, human gut, and rumen. The advantages of NGS are the high throughput, low cost, and accuracy of the data and exponentially support the NGS datasets. However, there are sequencing artifacts (low-quality and contaminated reads) that need to be tackled when using the NGS analysis. A major DNA sequencing method, with high precision, extended read length, and high throughput, would be necessary for further developments of fields.

Competing Interests There is no competing interest.

References

- Alonso N, Lucas G, Hysi P (2015) Big data challenges in bone research: genome-wide association studies and next-generation sequencing. *Bonekey Rep* 4:635
- Bai Y, Sartor M, Cavalcoli J (2012) Current status and future perspectives for sequencing livestock genomes. *J Anim Sci Biotechnol* 3(1):8
- Barboza ADM, Pylro VS, Jacques RJS, Gubiani PI, de Quadros FLF, da Trindade JK, Triplett EW, Roesch L (2018) Seasonal dynamics alter taxonomical and functional microbial profiles in Pampa biome soils under natural grasslands. *PeerJ* 13(6):e4991
- Behjati S, Tarpey PS (2013) What is next generation sequencing? *Arch Dis Child Educ Pract Ed* 98(6):236–238
- Bharagava RN, Purchase D, Saxena G, Mulla SI (2019) Applications of metagenomics in microbial bioremediation of pollutants: from genomics to environmental cleanup. In: *Microbial diversity in genomic era*. Academic Press, London, pp 459–477
- Blow N (2009) Transcriptomics: the digital generation. *Nature* 458(7235):239–242
- Bruno F, Marinella M, Santamaria M (2015) e-DNA meta-barcoding: from NGS raw data to taxonomic profiling. *Methods Mol Biol* 1269:257–278
- Cao Y, Fanning S, Proos S, Jordan K, Srikumar S (2017) A review on the applications of next generation sequencing technologies as applied to food-related microbiome studies. *Front Microbiol* 8:1829
- Chakraborty C, Doss CG, Patra BC, Bandyopadhyay S (2014) DNA barcoding to map the microbial communities: current advances and future directions. *Appl Microbiol Biotechnol* 98(8):3425–3436
- Chiu CY, Miller SA (2019) Clinical metagenomics. *Nat Rev Genet* 20(6):341–355
- Conrads G, Abdelbary MMH (2019) Challenges of next-generation sequencing targeting anaerobes. *Anaerobe* 58:47
- Costa A, Franco OL (2017) Impact and influence of “omics” technology on hyper tension studies. *Int J Cardiol* 228:1022–1034
- Costeira R, Doherty R, Allen CCR, Larkin MJ, Kulakov LA (2019) Analysis of viral and bacterial communities in groundwater associated with contaminated land. *Sci Total Environ* 656:1413–1426
- Craig DW, Nasser S, Corbett R, Chan SK, Murray L et al (2016) A somatic reference standard for cancer genome sequencing. *Sci Rep* 6:24607

- Czaplicki LM, Gunsch CK (2016) Reflection on molecular approaches influencing state-of-the-art bioremediation design: culturing to microbial community fingerprinting to omics. *J Environ Eng* 142(10):03116002. [https://doi.org/10.1061/\(ASCE\)EE.1943-7870.0001141](https://doi.org/10.1061/(ASCE)EE.1943-7870.0001141)
- Faure D, Joly D (2015) Next-generation sequencing as a powerful motor for advances in the biological and environmental sciences. *Genetica* 143(2):129–132
- Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM et al (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269(5223):496–512
- Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA (1995) The minimal gene complement of *Mycoplasma genitalium*. *Science* 270(5235):397–403
- Garalde DR, Snell EA, Jachimowicz D, Heron AJ, Bruce M et al (2016) Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods* 15(3):201–206
- Goto Y, Akahori R, Yanagi I (2019) Challenges of single-molecule DNA sequencing with solid-state nanopores. *Adv Exp Med Biol* 1129:131–142
- Guillermin Y, Lopez J, Chabane K, Hayette S, Bardel C, Salles G, Sujobert P, Huet S (2018) What does this mutation mean? The tools and pitfalls of variant interpretation in lymphoid malignancies. *Int J Mol Sci* 19(4):E1251
- Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, Braslavsky I et al (2008) Single-molecule DNA sequencing of a viral genome. *Science* 320(5872):106–109
- Hiendleder S, Bauersachs S, Boulesteix A, Blum H, Arnold GJ, Frohlich T, Wolf E (2005) Functional genomics: tools for improving farm animal health and welfare. *Rev Sci Tech* 24(1):355–377
- Hu P, Shao Y, Xu J, Wang T, Li Y et al (2019) Genome wide study on genetic diversity and phylogeny of five species in the genus *Cervus*. *BMC Genomics* 20(1):384
- Iacoangeli A, Al Khleifat A, Sproviero W, Shatunov A, Jones AR et al (2019) DNAscan: personal computer compatible NGS analysis, annotation and visualisation. *BMC Bioinf* 20(1):213
- Jaswal R, Pathak A, Edwards B III, Lewis R III, Seaman JC, Stothard P, Krivushin K, Blom J, Rupp O, Chauhan A (2019) Metagenomics-guided survey, isolation, and characterization of uranium resistant microbiota from the Savannah River Site, USA. *Genes* 10(5):E325
- Jeon K, Min B, Park JS, Kang YK (2017) Simultaneous methylation-level assessment of hundreds of CpG sites by targeted bisulfite PCR sequencing (TBPseq). *Front Genet* 8:97
- Johnson DS, Mortazavi A, Myers RM, Wold B (2007) Genome-wide mapping of in vivo protein-DNA interactions. *Science* 316(5830):1497–1502
- Ju F, Zhang T (2015) Experimental design and bioinformatics analysis for the application of Metagenomics in environmental sciences and biotechnology. *Environ Sci Technol* 49(21):12628–12640
- Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER (2013) The next-generation sequencing revolution and its impact on genomics. *Cell* 155(1):27–38
- Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J, Turner DJ (2010) Target-enrichment strategies for next-generation sequencing. *Nat Methods* 7(2):111–118
- Marco D (2008) Metagenomics and the niche concept. *Theory Biosci* 127(3):241–247
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437(7057):376–380
- Marian AJ (2011) Medical DNA sequencing. *Curr Opin Cardiol* 26(3):175–180
- Messing J (2001) The universal primers and the shotgun DNA sequencing method. In: Graham CA, Hill AJM (eds) *DNA sequencing protocols. Methods in molecular biology™*, vol 167. Humana Press, Totowa
- Mitra RD, Church GM (1999) In situ localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Res* 27(24):e34
- Mortazavi A, Williams BA, Mccue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat Methods* 5(7):621–628

- Ramamurthy M, Sankar S, Kannangai R, Nandagopal B, Sridharan G (2017) Application of viromics: a new approach to the understanding of viral infections in humans. *Virus Disease* 28(4):349–359
- Roberts A, Pimentel H, Trapnell C, Pachter L (2011) Identification of novel transcripts in annotated genomes using RNA-seq. *Bioinformatics* 27(17):2325–2329
- Runtuwene LR, Tuda JSB, Mongan AE, Suzuki Y (2019) On-site MinION sequencing. *Adv Exp Med Biol* 1129:143–150
- Saggese I, Bona E, Conway M, Favero F, Ladetto M, Liò P, Manzini G, Mignone F (2018) STABLE: a novel approach to de novo assembly of RNA-seq data and its application in a metabolic model network based metatranscriptomic workflow. *BMC Bioinf* 19(Suppl 7):184
- Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Ehrlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487–491
- Schloss PD, Handelsman J (2004) Status of the microbial census. *Microbiol Mol Biol Rev* 68:686–691
- Schmidt AY, Hansen TVO, Ahlborn LB, Jønson L, Yde CW, Nielsen FC (2017) Next-generation sequencing-based detection of germline copy number variations in BRCA1/BRCA2: validation of a one-step diagnostic workflow. *J Mol Diagn* 19(6):809–816
- Sen P, Orešič M (2019) Metabolic modeling of human gut microbiota on a genome scale: an overview. *Metabolites* 9(2):E22
- Shendure J, Balasubramanian S, Church GM, Gilbert W, Rogers J, Schloss JA, Waterston RH (2017) DNA sequencing at 40: past, present and future. *Nature* 550(7676):345–353
- Singh AP, Zafer S, Pe'er I (2013) MetaSeq: privacy preserving meta-analysis of sequencing-based association studies. *Pac Symp Biocomput*:356–367. https://doi.org/10.1142/9789814447973_0035
- Smith M (2017) DNA sequence analysis in clinical medicine, proceeding cautiously. *Front Mol Biosci* 4:24
- Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, Hood LE (1986) Fluorescence detection in automated DNA sequence analysis. *Nature* 321(6071):674–679
- Song S, Jarvie T, Hattori M (2013) Our second genome-human metagenome: how next-generation sequencer changes our life through microbiology. *Adv Microb Physiol* 62:119–144
- Stein LD (2011) An introduction to the informatics of “next-generation” sequencing. *Curr Protoc Bioinformatics* 36(1):11.1.1–11.1.9
- Tarkkonen K, Hieta R, Kytölä V, Nykter M, Kiviranta R (2017) Comparative analysis of osteoblast gene expression profiles and Runx2 genomic occupancy of mouse and human osteoblasts in vitro. *Gene* 626:119–131
- Vallenet D, Calteau A, Cruveiller S, Gachet M, Lajus A et al (2017) MicroScope in 2017: an expanding and evolving integrated resource for community expertise of microbial genomes. *Nucleic Acids Res* 45(D1):D517–D528
- Warnke-Sommer J, Ali H (2016) Graph mining for next generation sequencing: leveraging the assembly graph for biological insights. *BMC Genomics* 17:340
- Weissenkampen JD, Jiang Y, Eckert S, Jiang B, Li B, Liu DJ (2019) Methods for the analysis and interpretation for rare variants associated with complex traits. *Curr Protoc Hum Genet* 101(1):e83
- Won YJ, Kang LH, Lee SG, Park SW, Han JI, Paik SY (2019) Molecular genomic characterization of severe fever with thrombocytopenia syndrome virus isolates from South Korea. *J Microbiol* 57:927. <https://doi.org/10.1007/s12275-019-9174-8>
- Xuan J, Yu Y, Qing T, Guo L, Shi L (2013) Next-generation sequencing in the clinic: promises and challenges. *Cancer Lett* 340(2):284–295
- Zaura E (2012) Next-generation sequencing approaches to understanding the oral microbiome. *Adv Dent Res* 24(2):81–85
- Zhang J, Chiodini R, Badr A, Zhang G (2011) The impact of next-generation sequencing on genomics. *J Genet Genomics* 38(3):95–109